INVESTIGATION OF PLASMA IMPLANTATION AND GATE OXIDE CHARGING DURING PLASMA PROCESSING

by

Barry Paul Linder

Memorandum No. UCB/ERL M99/8

18 February 1999

COVER

INVESTIGATION OF PLASMA IMPLANTATION AND GATE OXIDE CHARGING DURING PLASMA PROCESSING

by

Barry Paul Linder

Memorandum No. UCB/ERL M99/8

18 February 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Investigation of Plasma Implantation and Gate Oxide Charging during Plasma Processing

by

Barry Paul Linder

B.S. (Pennsylvania State University) 1993
M.S. (University of California at Berkeley) 1996

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

Doctor of Philosophy

in

Engineering - Electrical Engineering
and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor Nathan W. Cheung, Chair
Professor Chenming Hu
Professor Michael A. Lieberman
Professor Timothy D. Sands

Spring 1999

The dissertation of Barry Paul Linder is approved by

Nathan Cheung Feb. 18, 1999

Chair Date

Feb. 18, 1999

Feb 18, 1999

Feb 18, 1999

University of California, Berkeley

Spring 1999

Abstract

Investigation of Plasma Implantation and Gate Oxide Charging during Plasma Processing

by

Barry Paul Linder

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California at Berkeley

Professor Nathan W. Cheung, Chair

Plasma Immersion Ion Implantation (PIII) is an alternative to conventional implantation for high dose and low cost applications. Full commercialization of PIII requires process models for understanding the effects of the implant parameters on the final implant profile.

All plasma processes, including PIII, can cause plasma induced gate oxide charging damage. A model is developed that predicts charging damage for all plasma processes, including PIII.

The basic coupled plasma model for PIII contains three separate modules, the plasma, the wafer structure, and the substrate bias. The plasma module consists of a set of physically derived equations from a Quasi-Static Child Law analysis for the sheath expansion and plasma currents. The wafer structure module accounts for the type of devices and the presence of thin gate oxides on the wafer at the time of the plasma process. The substrate bias drives the implant, and hence, the effects of its characteristics on the plasma process is studied. The coupled PIII model contains no fitting parameters, and only requires a Langmuir Probe measurement to determine the ion density, plasma potential, plasma floating potential, and electron temperature.

Implant energy profiles are generated with the coupled PIII plasma model. A pulsed PIII implant is not mono-energetic, but rather contains a significant energy spread.

Ions from the matrix sheath fundamentally limit the energy integrity of an implant. The contribution of the rise time of the implantation pulse to the energy spread is minimized by reducing the rise time as close to zero as possible. On the other hand, the fall time of the pulse simply need be less than the sheath collapse time.

The basic PIII plasma model has been extended to take into account dielectric substrates, multiple plasma ion species, and ion sheath collisions. With PIII into dielectric substrates (as in silicon-on-insulator or thin film transistor processing) substrate charging and substrate coupling corrupt the implant profile. The extended model allows optimization of the pulse width, pulse frequency, bias voltage, and plasma ion density to control substrate charging while maintaining an acceptable dose rate. The single species model may be adopted for multiple ion species plasmas by utilizing an effective mass and an effective Bohm velocity. Most practical PIII processes operate in a slightly collisional regime, with more than 10% of the ions undergoing a collision before implantation. A Monte Carlo analysis of collisions enables generation of implant energy profiles under these conditions.

As gate oxides scale from 5 nm to 3 nm and below, the issue of plasma induced charging damage has come to the forefront. The issue of gate oxide scaling and charging damage has been resolved through the development of a Universal Charging Damage Model. A load line analysis of the plasma impedance and gate conduction establishes the stress condition during the plasma process, and an oxide reliability model correlates the stress condition with oxide damage.

Assuming identical plasma conditions during processing, there are three stressing regimes observed with oxide scaling. Thicker oxides undergo constant voltage stressing, while thinner oxides undergo constant current stressing. Ultra-thin oxides, thinner than about 3 nm, also undergo constant current stressing, but the electrons tunnel by direct tunneling rather than Fowler Nordheim tunneling. The oxide damage due to plasma process-

ing peaks in the constant current regime, for an oxide thickness around 5nm. The exact thickness depends on the processing conditions and antenna ratio.

Combining both the Coupled Plasma Model and the Universal Charging Damage Model allows the prediction of charging damage for PIII. Simulations predict that the amount of damage depends on the frequency of pulsing. Device circuit structures and parameters, such as wells, channel doping, circuit antennas, and dielectric substrates affect PIII charging damage. Oxides in N-wells charge more negatively, and oxides in P-wells charge more positively, while a depletion region underneath the oxide protects the oxide. Large area antennas create single pulse charging damage by amplifying the effective charge deposition density. Simulations show that devices on dielectric or SOI substrates are generally immune to gate oxide charging damage during PIII.

The addition of the dielectric, multiple species, and collision modules to the plasma, wafer structure, and bias models forms a fairly comprehensive one dimensional PIII dose and implant simulator. The Universal Charging Damage load line analysis forms the framework for analyzing plasma induced charging damage for all plasma processes.

Nathan Cheung

Professor Nathan W. Cheung
Chair, Dissertation Committee

Investigation of Plasma Implantation and Gate Oxide Charging during Plasma Processing

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank all of the professors and colleagues who have helped me throughout my graduate career. I thank my advisor, Professor Cheung, for continuously guiding my research, providing innumerable suggestions while teaching me to critically analyze my research. I thank Professor Chenming Hu for discussions on oxide reliability and charging damage. I thank Professor Michael A. Lieberman for teaching me about plasma sheath dynamics, plasma diagnostics, and all other aspects of plasma processing. I would like to thank Professors Nathan W. Cheung, Michael A. Lieberman, Chenming Hu, and Timothy Sands for serving on my thesis committee.

I am indebted to William En for mentoring me during my first years of graduate research. None of this work would have been possible without his initial guidance. Dong-gun Park has provided invaluable advice concerning plasma charging damage and measurement techniques. Erin Jones has always been my eyes and ears to the outside world, and taught me to critically review journal articles. Without Adam Wengrow's exceptional machining techniques, none of the recent modifications to the PIII machine would have been possible. Sundar Kumar Iyer has always taught me to strive for excellence, and his editing of this thesis is greatly appreciated.

I would like to thank all the inhabitants of 144MA, Kaustav Banerjee, James Chan, Yonah Cho, Bill En, Erin Jones, Xiang Lu, Adam Wengrow, William Wong, Bing-Ling Yang, Ching-Fa Yeh, and Changhan Yun for their support, companionship, and technical advice.

My life at Berkeley has been far more rewarding because of the friendships I have formed. I met Nelson Chow in my first weeks at Berkeley, and he has been a great friend ever since. He introduced me to the wonderful world of bicycling, snow shoeing, and the art of whirlwind trips around California and the world. All my other friends at Berkeley have enriched my Berkeley experience. I am would also like to thank all of the EECS

Bombers, whom I have played softball with every Monday for 5 years. Finally, I am grateful for my family, especially my mother, for their love and support.

1 Introduction

1.1 Introduction

Plasma processing is an integral IC fabrication technology enabling sub-micron scaling. Plasma processes are generally dry, single wafer, and cluster tool compatible. Ion milling, Reactive Ion Etching (RIE), Plasma Enhanced Chemical Vapor Deposition (PECVD), and sputtering plasma processes are mainstays of integrated circuit fabrication. A new plasma based process, Plasma Immersion Ion Implantation (PIII), is emerging as an economical alternative to conventional implantation for high dose applications. A full understanding of the trade-offs between the implant parameters is needed before the commercialization of the process.

All plasma processes, including PIII, are susceptible to plasma induced gate oxide charging damage. Therefore, a generalized damage model is developed to understand and predict charging damage for all plasma processes, including PIII.

1.2 Plasma Immersion Ion Implantation

Ion Implantation is one of the crucial steps in a semiconductor process flow. In current CMOS technology, there are numerous fabrication implantation steps. Namely, well formation, channel stop, threshold adjust, punchthrough protection, dual poly implant, and source/drain implant. Thin film transistor technology contains additional implantations such as hydrogenation and poly-Si grain size control. Conventional beam-line implantation excels at dose control and uniformity, but is limited in implant current especially at low implant energies. The shift to larger wafer sizes for IC fabrication and

the production of large arrays of thin film transistors for flat panel displays exacerbates the dose rate requirement. For instance, a $10^{16}/cm^2$ doping of a 500 mm x 500 mm substrate with a 10 mA conventional implanter requires over 6 minutes per substrate, yielding a woefully low throughput. PIII is a promising alternative for high dose, high throughput doping, requiring less than 10 seconds for the same implant.

PIII is a novel implantation technique in which the substrate is immersed in a plasma containing the implant ion species (Figure 1-1). Applying a high voltage negative bias to the substrate accelerates and implants the plasma ions. If wafer charging is a concern, the bias can be pulsed. The pulse off time following each implant pulse allows the plasma electrons to neutralize the deposited positive charge.

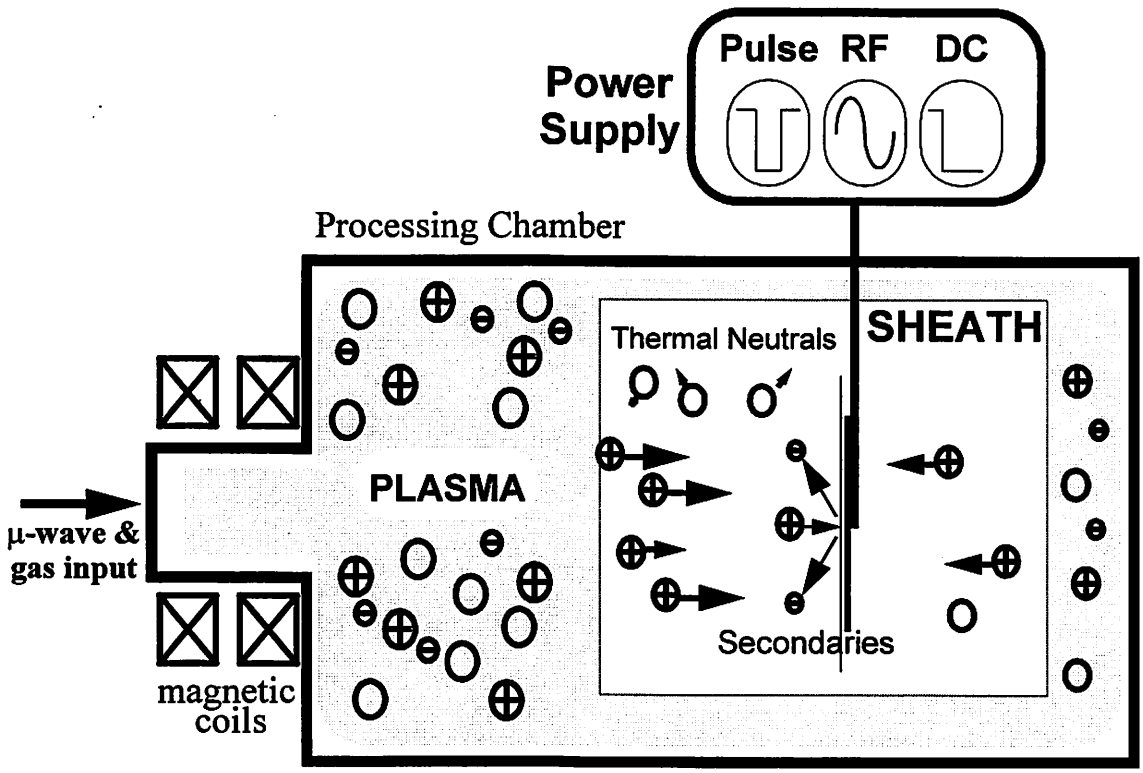

Figure 1-1 Plasma Immersion Ion Implantation Concept

Diagram of Plasma Immersion Ion Implantation. A plasma source generates a plasma containing the ion implant species, which flows from the source into the main chamber enveloping the entire wafer. The substrate bias extracts, accelerates, and implants the ion species.

PIII's main advantage is the high attainable dose rate. Since the plasma surrounds the entire wafer, the whole wafer is implanted simultaneously, yielding an implantation time independent of wafer size, i.e. the implant time for a 300 mm (12 inch) diameter wafer is the same as a 200 mm (8 inch) wafer. This contrasts sharply with conventional implantation where the implant time scales with wafer area (Figure 1-2). For comparison, typical PIII dose rates can exceed 1 mA/cm^2 over the entire wafer, while a state of the art high beam implanter current achieves 100 mA. With a 300 mm substrate, the current density is only 0.14 mA per cm^2 [1-1, 1-2]. To the first order, PIII would be faster for wafers larger than 4.5 inches, if wafer handling time is negligible.

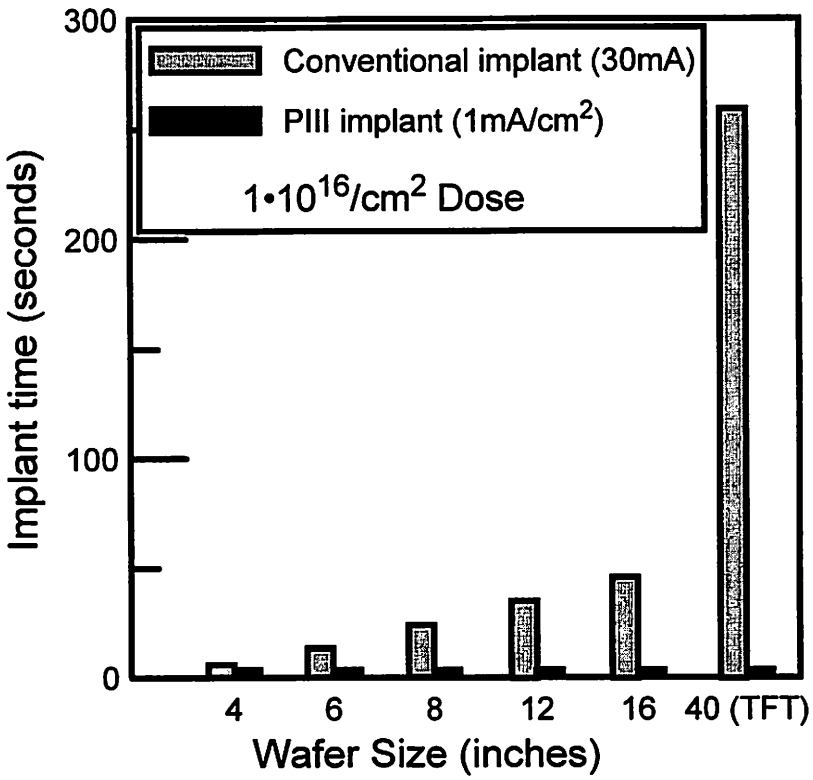

Figure 1-2 Implant Time Comparison

Since PIII implants the entire wafer simultaneously, the implant time remains constant regardless of the wafer size. In contrast, the conventional beamline implantation time scales with the square of the wafer radius. The difference between the two becomes significant at larger wafer sizes.

Due to beam optics, conventional implanter currents deteriorate sharply at implant energies below 10 keV. The ions are usually extracted at tens of kilovolts and then decelerated to the required implant energy [1-3]. Besides introducing implant energy spread [1-4], the deceleration reduces beam currents significantly. This limitation has led semiconductor manufacturers to search for alternatives for low energy implantation.

Beside dose rate, PIII has other advantages. Since the entire wafer is implanted simultaneously, a PIII machine does not require beam scanning mechanisms. As shown in Figure 1-1, the machine contains no mass separation unit or a long acceleration tube, simplifying the machine design and maintenance. The machine is flexible, fully scalable, and cluster tool compatible. Because the PIII machine is not specific to just implantation, it could also be used as an etch or a low temperature CVD tool. In principle, processes that include a pre or post implantation etch or deposition are all possible in one machine without breaking vacuum.

The applications currently under development are shallow junction formation [1-5 - 1-7], SIMOX fabrication [1-8 - 1-20], SOI wafer fabrication by the Ion Cut process [1-18, 1-19], trench doping [1-21, 1-22], hydrogenation of TFT's [1-23], palladium doping for copper plating [1-24, 1-25], and metallurgical hardening by nitrogen implants [1-26 - 1-28].

A number of issues need to be investigated before full PIII implementation. Some of the main concerns are the relationship between substrate bias voltage, implant energy and dose rate, the extent of the implant energy spread and implant profile, and the mechanisms of gate oxide charging. Optimization of the plasma with respect to the ion density, electron temperature, floating potential, plasma potential, and the substrate bias variables of pulse width and pulse frequency will be required.

To accomplish these goals, a PIII model has been developed that solves physical equations to predict the implant current and voltage. The model predicts the ion implant energy spread for varying implant and substrate conditions, and estimates theoretical limits for the energy spread with PIII. Coupling the PIII model with a thin oxide tunneling current model, allows calculation of plasma charging damage. Finally, the simulation then helps optimize the implant conditions for minimal oxide damage.

Chapter 2 describes the Berkeley PIII tool, Chapter 3 details the PIII Coupled Plasma Model, Chapter 4 investigates the sources of PIII implant energy spread, Chapters 5, 6, and 7 adapt the coupled plasma model for dielectric substrates, multiple plasma ion species, and ion sheath collisions, while Chapter 11 examines charging damage during PIII.

1.3 Plasma Induced Gate Oxide Charging Damage

An undesirable by-product of all plasma processes, including PIII, is gate oxide plasma charging damage. Concern over plasma induced charging damage has mounted in the IC fabrication industry as designers continue scaling gate oxides towards 3 nm and below. Since thinner oxides breakdown at lower voltages, it has been generally assumed that thinner oxides will be more susceptible to charging damage. On the other hand, electrical stress data suggests that thinner oxides may be more robust [1-29]. It is, thus, important to fully understand plasma charging damage in contemporary high plasma density tools, especially for the PIII process, and to predict the future role of plasma charging damage.

1.4 Origin of Plasma Damage

Plasma damage, in its most general definition, is any inadvertent degradation of the MOS system from plasma processing. The three main damage mechanisms are electrical stressing of the oxide during plasma processing, radiation damage from high energy

photons emitted by the plasma, and damage to the gate dielectric from physical bombardment of the oxide by the plasma ions. With typical plasma processing conditions, the dominant damage mechanism for CMOS processing is electrical stressing.

During the plasma process, both ions and electrons bombard the surface of the wafer (Figure 1-3). If for whatever reason, the ion and electron currents do not balance, charge will buildup on the surface of the wafer. The interconnects conduct this charge from the wafer surface down to the transistor gate, electrically stressing the gate oxide. With a large enough electrical field, significant tunneling currents flow. These currents may break bonds in the gate dielectric, degrading the bulk and interface properties of the oxide.

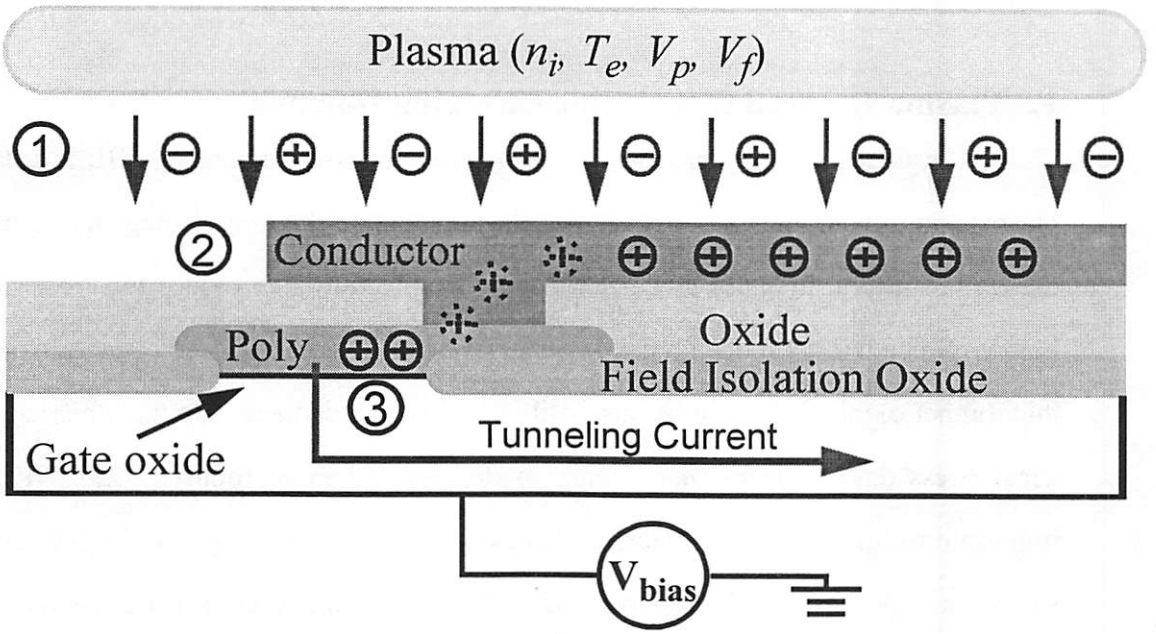

Figure 1-3 Origin of Plasma Induced Charging Damage

During plasma processing, ions and electrons bombard the wafer surface (1). Often, the fluxes are not completely balanced, and a net charge builds-up on the surface. The surface conductor transfers the charge from the surface to the poly layer (2). If the electric field generated by this conducted charge is great enough, tunneling currents flow through the gate oxide (3), resulting in electrical stress, which is called plasma induced charging damage.

The key element initiating electrical plasma damage is the imbalance between the ion and electron fluxes. Traditionally, plasma non-uniformities across the wafer generated the flux imbalance [1-30]. More recently, electron shading has been implicated as a cause of damage in sub-micron geometries [1-31]. Figure 1-5 depicts plasma damage from spatial plasma non-uniformities. At equilibrium, a uniform plasma will charge the substrate to the plasma floating potential. A non-uniform plasma contains a space-varying floating potential; a conducting substrate, however, must have a single substrate potential. The equilibrium substrate voltage will be a spatial average of the plasma floating potential. Locally, the plasma floating potential is not equal to the substrate potential. Currents may flow from the high plasma potential region in the plasma, through a gate oxide, into the substrate, and finally back to the plasma. In the case of a plasma source with a higher ion density in the center of the reactor, current flows from the center to the edge of the wafer, damaging the gate dielectrics.

Electron shading is the second dominant electrical stressing mechanism. Figure 1-4 depicts a high aspect ratio metal etch process. With a substrate bias, the ions bombard the wafer nearly anistropically, creating the desired sharp sidewalls. The electron are not significantly accelerated by the substrate bias, and therefore impinge upon the wafer with nearly thermal energies. Negative charge builds up on the photoresist, repelling electrons from the surface. Overall, the isotropic nature of the electrons and the negative charge on the photoresist, significantly reduces the electron current density at the bottom of the trench. This charges the metal line and the MOS gate positively. A large enough electric field allows significant tunneling current which stress the gate oxide. The return path to the plasma is through any open area on the wafer.

Historically, plasma non-uniformities were the dominant damage mechanism. Plasma uniformities have improved greatly over the past decade, minimizing this damage mechanism. On the other hand, sub-micron scaling requires high aspect ratios magnifying the role of electron shading.

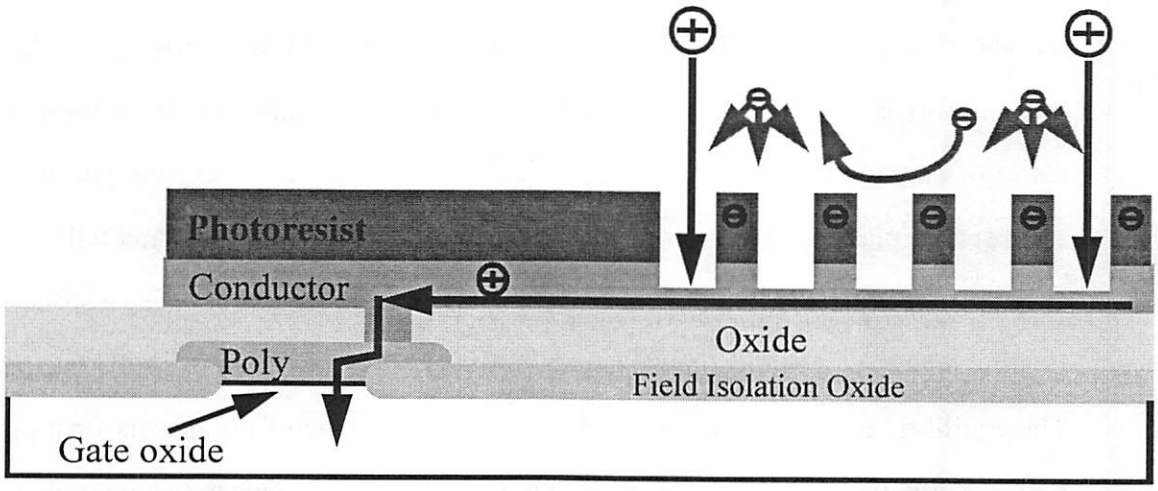

Figure 1-4 Electron Shading

The directional ions travel to the bottom of the etched trench, while the electrons are blocked by the photoresist. Positive charge builds-up on the conductor, and electrically stresses the gate oxide.

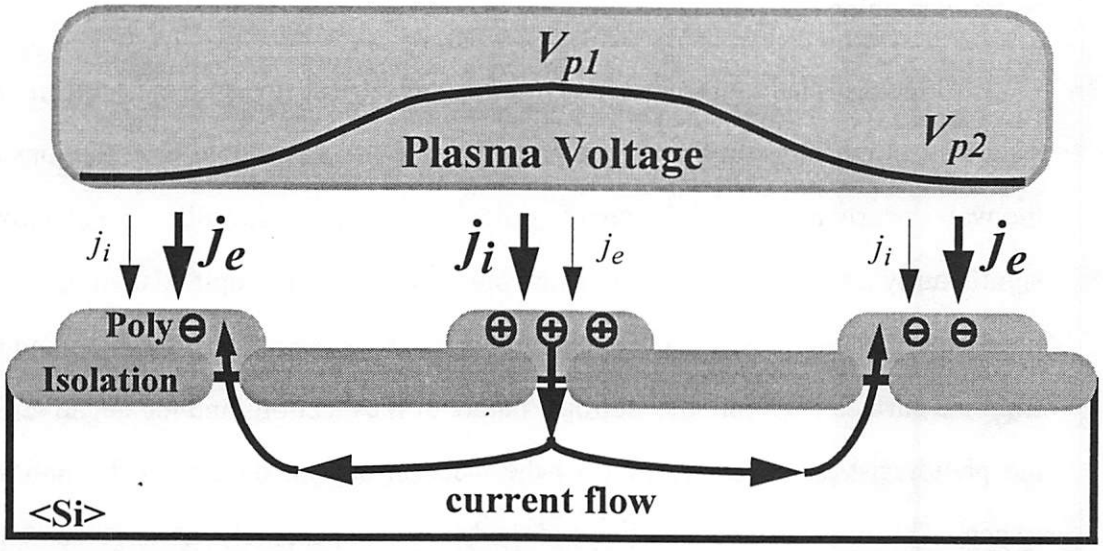

Figure 1-5 Charging Damage from Spatial Plasma Non-uniformities

Plasma non-uniformities create varying ion and electron currents across the wafer. With a higher plasma potential in the center of the wafer, current flows from the center of the plasma, through the wafer, and back to the plasma near the edges of the wafer.

1.5 Effect of Plasma Charging Damage

The plasma damage mechanism studied in this work is the electrical stressing of the gate oxide during plasma processing. The effects of electrical plasma damage are similar to those of bench electrical stressing tests. Damage manifests itself in three main ways: catastrophic oxide damage, reduced oxide lifetime, and transistor performance degradation.

In extreme cases, plasma damage breaks down the oxide, creating a short circuit. In general, however, the damage manifests itself in a more subtle manner, such as reliability and performance degradation of the oxide. Oxide damage is a cumulative process. Therefore, even though the oxide may appear intact after the plasma process, the plasma stressing does subtract from the total useful lifetime of the oxide, possibly leading to early and unexpected operational failures.

The plasma damage also degrades the transistor performance. A stressed oxide contains a higher density of bulk and interface traps. These increase the depletion capacitance of the MOS system, trap charge, and reduce the inversion charge mobility. The altered capacitance and trapped charge shift the transistor threshold voltage. The increased depletion capacitance increases the sub-threshold slope. The interface traps reduce the inversion mobility and transistor transconductance. All these products of plasma damage combine to reduce transistor performance. Overall, plasma damage degrades the performance while increasing the variance of the transistor parameters, ultimately leading to yield loss.

1.6 CMOS Scaling and Plasma Damage

The expected scaling of CMOS into the nanometer regime will affect the extent of plasma damage. The two processing trends that will have an affect plasma damage are the evolution of plasma processing tools, and the thinning of the gate oxide.

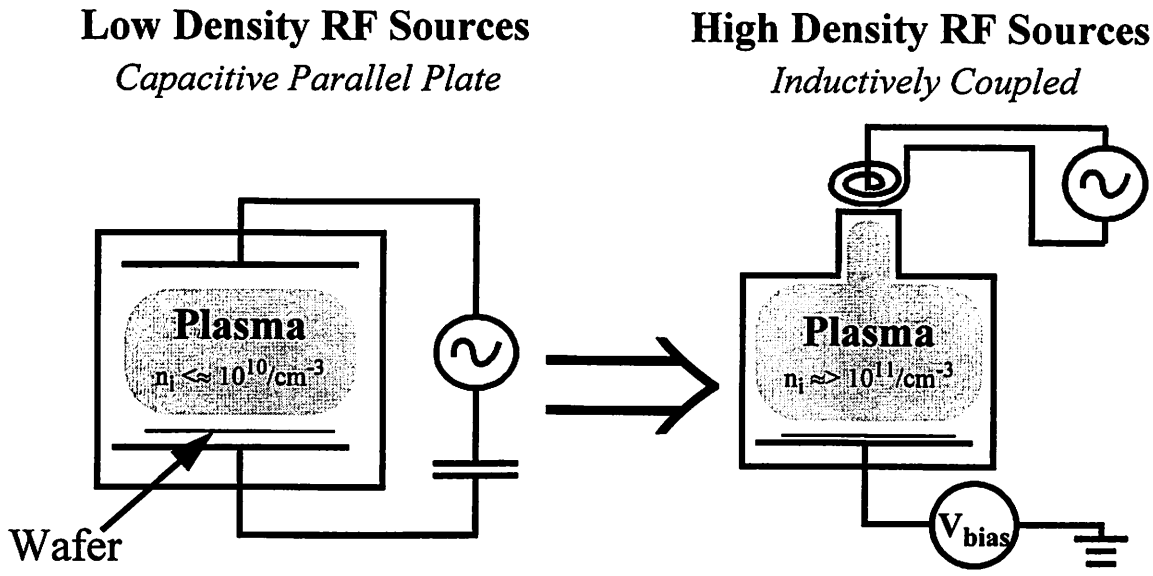

Figure 1-6 Comparison between Capacitive and Inductive Plasmas

Over the past decade plasma sources have evolved from capacitive coupling to inductive coupling. Besides providing independent control of wafer and plasma source biases, inductively coupled plasmas create higher ion densities, allowing for better linewidth and throughput. These higher ion densities, however, may increase plasma damage on the wafer.

Over the past decade the plasma processing tools have evolved from low density capacitively coupled systems to high density inductively coupled tools (Figure 1-6). Capacitively coupled plasmas utilize a single RF bias that doubles as the plasma generation source and the wafer bias. Inductively coupled plasmas have separate RF sources for plasma generation and for wafer bias. This allows independent control of plasma density and ion impingement energy. This permits the newer tools to increase the ion densities from below $10^{10}cm^{-3}$ to $10^{11}cm^{-3}$ and above. The independent voltage control and higher ion densities allow for better linewidth control and throughput. The higher ion densities and the corresponding lower electron temperature from advanced ion sources alter the plasma impedance affecting the level of plasma charging damage.

The second major processing trend that influences plasma damage is the scaling of the gate oxide thickness. Figure 1-7 shows the SIA roadmap for oxide thickness into the next century. Gate oxide thicknesses have been scaled extremely aggressively over the

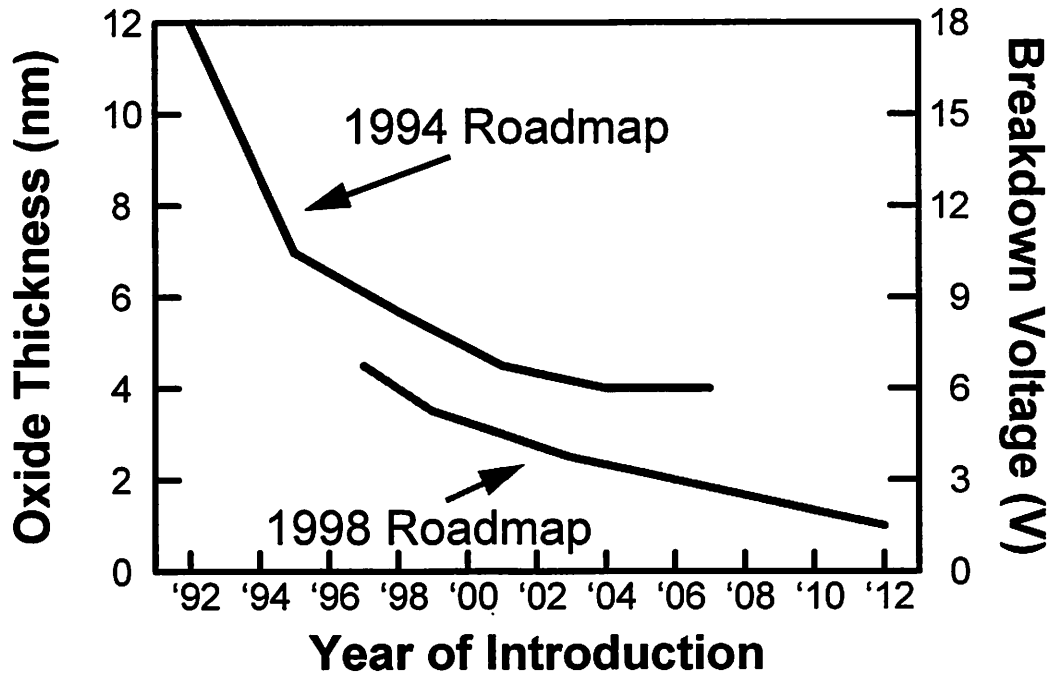

Figure 1-7 SIA Roadmap for Gate Dielectric Scaling

Gate oxides are being scaled very aggressively, with the 1998 SIA roadmap 2 nm thinner than previous predictions. Gate dielectric scaling will most definitely affect plasma charging damage [1-32].

past 5 years, with the original roadmap being pushed forward by more than 2 nm in the current roadmap. The latest papers report SiO_2 gate dielectrics as thin as 1 nm [1-33]. Intuitively, thinner oxides should be more susceptible to plasma damage, because of their lower breakdown voltage. But, the plasma is not a perfect voltage source, and this complicates the situation. Plasma damage as a function of gate oxide thickness is not a monotonic relationship, but is affected by the complex interaction of the plasma and the gate dielectric.

1.7 Universal Charging Damage Model

To date, there is no standardized model for predicting gate oxide charging damage. To this end, this work formulates a universal plasma charging damage model applicable to a wide range of plasma processes. The model developed predicts plasma charging dam-

age as a function of gate oxide thickness, plasma density, plasma electron temperature, and device antenna ratio. Chapter 8 develops a charging damage model derived from a loadline analysis of the plasma and gate oxide impedances and discusses the ramifications of the model's predictions. Chapter 9 examines the various methods for detecting charging damage. Chapter 10 presents experimental verification of the model's predictions as a function of gate oxide thickness, plasma charge density, and device antenna ratio. In Chapter 11, the model has been combined with the PIII process model to predict and minimize plasma charging damage in the emerging process.

References

[1-1] G. Ryding, T.H. Smick, M. Farley, et. al., "The IBIS 1000 SIMOX Production Implanter," *Proceedings of the Eleventh International Conference on Ion Implantation Technology*, pp. 436-439, 1996.

[1-2] K. Tokiguchi, T. Seki, K. Amemiya, and Y. Yamashita, "Advanced Microwave Ion Source for 100mA-class SIMOX Ion Implantation," *Proceedings of the Eleventh International Conference on Ion Implantation Technology*, pp. 287-290, 1996.

[1-3] S. Moffett, "Life Below 1keV," *Proceedings of the Eleventh International Conference on Ion Implantation Technology*, pp. 5-8, 1996.

[1-4] L. M. Rubin, "Effect of Beam Energy Purity on Junction Depths in Sub-micron Devices," *Proceedings of the Eleventh International Conference on Ion Implantation Technology*, pp. 96-99, 1996.

[1-5] E. C. Jones and N. W. Cheung, "Characteristics of Sub-100 nm p^+/n Junctions Fabricated by Plasma Immersion Ion Implantation," *IEEE Electron Device Letters*, Vol. 14, no. 9, pp. 444-446, 1993.

[1-6] E. C. Jones, S. Im and N. W. Cheung, "Ultra-shallow P^+/N Junctions Formed by SiF_4 Preamorphization and BF_3 Implantation using Plasma Immersion Ion Implantation," *Beam Solid Interactions: Fundamentals and Applications Symposium*, pp. 255-259, 1993.

[1-7] C. A. Pico, X. Y. Qian, E. C. Jones, M. A. Lieberman, and N. W. Cheung, "Shallow P^+-N Junction Fabrication by Plasma Immersion Ion Implantation," *Low Energy Ion Beam and Plasma Modification of Materials, MRS Symposium,* vol. 223, pp. 115-119, 1991.

[1-8] J. Min, P. K. Chu, Y. C. Cheng, S. Im, S. Iyer, and N. W. Cheung, "Fabrication and Characterization of SIMOX wafers by Plasma Immersion Ion Implantation." *Conference Digest International Conference on Plasma Science and Technology*, Chengdu China. p. 153, 1994.

[1-9] S. S. K. Iyer, X. Lu, N. W. Cheung, and C. Hu, "SOI MOSFET on Low Cost SPIMOX Substrate," *Proceedings of the 1998 International Electron Device Meeting*, pp. 1001-1004, 1998.

[1-10] J. Min, P. K. Chu, Y. C. Cheng, J. B. Liu, S. Im, S. Iyer, and N. W. Cheung, "Buried Oxide Formation by Plasma Immersion Ion Implantation." *Materials Chemistry and Physics*, vol. 40, pp. 219-222, 1995.

[1-11] J. B. Liu, S. S. K. Iyer, J. Min, P. Chu, R. Gronsky, C. Hu, and N.W. Cheung, "Formation of Silicon on Insulator with Separation by Plasma Implantation of Oxygen (SPIMOX)." *Beam-Solid Interactions for Materials Synthesis and Characterization Symposium*, pp. 117-122, 1995.

[1-12] J. Min, P. K. Chu, Y. C. Cheng, J. B. Liu, S. Iyer, and N.W. Cheung, "Synthesis and Characterization of SIMOX by PIII." *1995 International Conference on Materials and Process Characterization for VLSI (ICMPC 94)* p. 238, 1994.

[1-13] J. Liu, S. S. Iyer, C. Hu, N. W. Cheung, J. Min and P. K. Chu, "High Dose-Rate SIMOX Synthesis by Plasma Implantation." *The Second International Workshop on Plasma Based Ion Implantation*, Sydney, Australia, p. 1010, 1995.

[1-14] J. Min, P. K. Chu, Y. C. Cheng, J. B. Liu, S. S. K. Iyer, and N. W. Cheung, "An Economical Fabrication Technique for SIMOX using Plasma Immersion Ion Implantation." *International Conference on Solid-State and Integrated Circuits Technology 95*, Beijing. pp. 253-256, 1995.

[1-15] J. B. Liu, S. S. K. Iyer, J. Min, P. Chu, R. Gronsky, C. Hu and N.W. Cheung, "Competitive Oxidation during Buried Oxide Formation using Separation by Plasma Implantation of Oxygen (SPIMOX)." *Symposium Proceedings from the 1995 MRS Spring Meeting*. pp. 385-91, 1995.

[1-16] J. Min, P. K. Chu, Y. C. Cheng, J. B. Liu, S. Iyer, and N. W. Cheung, "A New Development of PIII for SIMOX Structure." *Chinese Journal of Semiconductors*, 16 (B), p. 636, 1995.

[1-17] J. B. Liu, S. S. K. Iyer, C. Hu, N. W. Cheung, R. Gronsky, J. Min, and P. Chu, "Formation of Buried Oxide in Silicon using Separation by Plasma Implantation of Oxygen." *Applied Physics Letters*, vol. 67, no. 16, pp. 2361-2363, 1995.

[1-18] X. Lu, S. S. K. Iyer, J. Lee, B. Doyle, Z. Fan, P. K. Chu, C. Hu, and N. W. Cheung, "Plasma Immersion Ion Implantation for SOI synthesis: SIMOX and Ion-cut," *Journal of Electronic Materia*l, vol. 27, no. 9, pp. 1059-1066, Sept., 1998.

[1-19] P. K. Chu, S. Qin, C. Chan, N. W. Cheung, and P. K. Ko, "Instrumental and Process Considerations for the Fabrication of Silicon-on-Insulators (SOI) structures by Plasma Immersion Ion Implantation, *IEEE Transactions on Plasma Science*, vol. 26, no. 1, pp. 79-84, Feb. 1998.

[1-20] J. B. Liu, S. S. K. Iyer, J. Min, P. K. Chu, R. Gronsky, C. Hu, and N.W. Cheung. "Synthesis of Buried Oxide by Plasma Implantation with Oxygen and Water Plasma." *1995 IEEE International SOI Conference*, pp. 166-167, 1995.

[1-21] X. Y. Qian, N. W. Cheung, M. A. Lieberman, R. Brennan, and M. I. Current, "Conformal Implantation for Trench Doping with Plasma Immersion Ion Implantation," *Nuclear Instruments & Methods in Physics Research*, Section B, vol. B55, no. 1-4, pp. 898-901, 1991.

[1-22] C. Yu, and N. W. Cheung, "Trench Doping Conformality by Plasma Immersion Ion Implantation (PIII)," *IEEE Electron Device Letters*, vol. 15, no. 6, pp. 196-198, 1994.

[1-23] S. Qin, Y. Zhou, T. Nakatsugawa, I. F. Husein, C. Chan, and T. J. King, "Plasma Ion Implantation of poly-Si CMOS Thin-Film Transistors at Low Energy and High Dose Rate using an Inductively-Coupled Plasma Source," *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1324-1328, June 1998.

[1-24] M-H. Kiang, M. A. Lieberman, N. W. Cheung, and X. Y. Qian, "Pd/Si Plasma Immersion Ion Implantation for Selective Electroless Copper Plating on SiO_2," *Applied Physics Letters*, vol. 60, no. 22, pp. 2767-2769, 1992.

[1-25] M-H. Kiang, J. Tao, W. Namgoong, C. Hu, M. A. Lieberman, N. W. Cheung, H. K. Kang, and S. S. Wong, "Planarized Copper Interconnects by Selective Electroless Plating," *Advanced Metallization and Processing for Semiconductor Devices and Circuits - II.*, pp. 745-755, 1992.

[1-26] J. R. Conrad, and T. Castagna, "Plasma Source Ion Implantation for Surface Modification," *Bull. Am. Phys. Soc.*, vol. 31, no. 9, pp. 1479-1482, 1986.

[1-27] N. C. Horswill, K. Sridharan, and J. R. Conrad, "A Fretting Wear Study of a Nitrogen Implanted Titanium Alloy," *Journal of Materials Science Letters*, vol. 14, no. 19, pp. 1349-1351, 1995.

[1-28] S. M. Malik, D. E. Muller, K. Sridharan, R. P. Fetherston, et. al., "Distribution of Incident Ions and Retained Dose Analysis for a Wedge-shaped Target in Plasma Source Ion Implantation," *Journal of Applied Physics*, vol. 77, no. 3, pp. 1015-19, Feb. 1, 1995.

[1-29] K. Scheugraf and C. Hu, "Reliability of Thin $Si0_2$," *Semiconductor Science and Technology,* vol. 9, pp. 989-1004, 1994.

[1-30] S. Fang and J. McVittie, "Thin Oxide Damage from Gate Charging during Plasma Processing", *IEEE Electron Device Letters*, vol. 13, no. 5, pp. 288-290, May 1992.

[1-31] K. Hashimoto, "Charging Damage Caused by Electron Shading Effect," *Japanese Journal of Applied Physics,* Part 1, vol. 33, pp. 6013-6018, Oct. 1994.

[1-32] National Technology Roadmap for Semiconductors, published by Sematech.

[1-33] G. Timp, et. al, "Progress towards 10 nm CMOS Devices," *1998 International Electron Device Meeting*, San Francisco, pp. 615-618, Dec. 1998.

2 Plasma Immersion Ion Implantation Setup and System

2.1 Berkeley PIII Reactor

A schematic of the PIII Berkeley reactor is shown in Figure 2.1. The overall machine length is 128 cm, with a width of 50 cm. Permanent magnets line the outside of the main chamber forming a magnetic bucket that confines the plasma, thus, reducing wall losses and improving plasma uniformity. A standard dual system of a mechanical pump and turbo pump with automatic crossover attains a base pressure near 1 microtorr. The Leybold turbo has a pumping speed of 1500 liters/second. A 1500 watt ASTEX 2.45 GHz microwave source supplies the power for ECR plasma generation. The microwaves are guided from the source to a 3-stub tuner, are coupled into the machine through a quartz window. Two electro-magnets in a mirror configuration surround the source chamber, generating the required magnetic field of 875 Gauss for electron cyclotron resonance (ECR) at 2.45 GHz. The main wafer holder handles up to 12 inch diameter wafers, and slides from anywhere between 20 cm and 45 cm away from the source chamber. A second 6" wafer holder adds de-ionized water cooling. The back door provides access for loading and unloading the wafers. Mass flow controllers regulate gas flow from 0 - 100 sccm in 0.1 sccm increments. This flow range allows pressures from below 50 μtorr to above 2 mtorr, with arbitrary gas mixtures. For higher pressure applications, a bypass leak valve control pressures up to 10 mtorr. Theoretically, any gas can be used as an implant source,

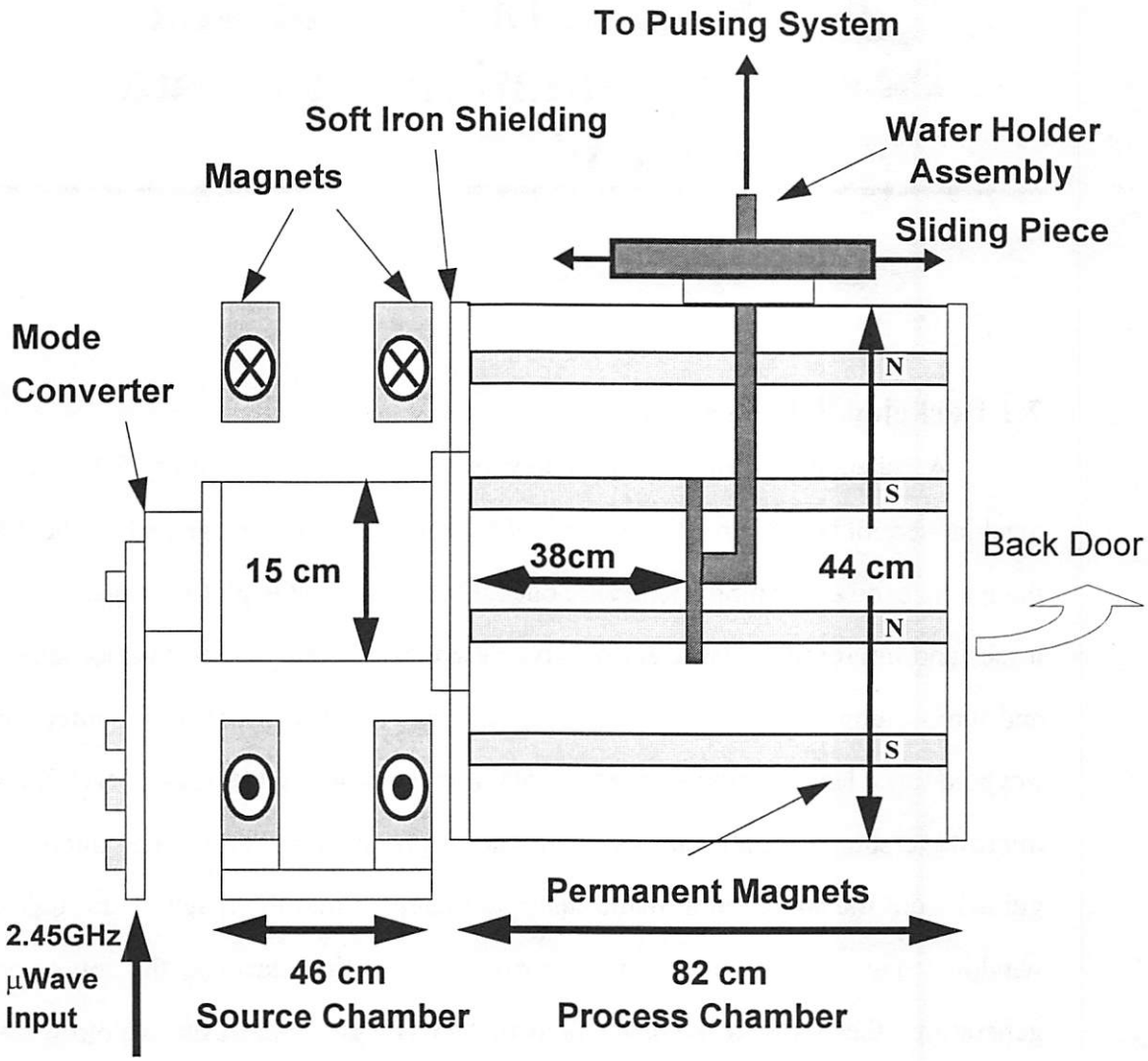

Figure 2-1 Schematic of Berkeley PIII machine

The Berkeley PIII reactor uses ECR remote plasma generation in the source chamber. The plasma diffuses from the source to the main chamber and immerses the wafer. The wafer assembly slides fore-and-aft, controlling uniformity and ion density. Permanent magnets confine the plasma improving uniformity.

but currently BF_3, He, Ne, SF_6, N_2, H_2, SiF_4, CF_4, O_2, H_2O, and Ar are available on the system

2.2 Diagnostic Tools

A variety of diagnostic tools were used in the study of the PIII system and process. The more important ones are discussed below.

2.2.1 Pressure Measurement

Three different tools monitor chamber pressure. A baratron measures pressures accurately from 0.1 to 50 mtorr, while an ion gauge measures from 0.1 to 1000 microtorr. BF_3 pressure measurement with the baratron proves inconsistent, necessitating the use of the thermocouple gauge reading of the turbopump back pressure. Figure 2-2 shows thermocouple calibration curves for Ar and BF_3. Argon was calibrated against the baratron, while BF_3 was calibrated with an ion gauge.

2.2.2 Mass Spectrometry

A mass spectrometer maps the mass and energy of the ions. The mass spectrometer's main function is to compare relative ion density ratios in multiple ion plasmas. For example, Figure 2-3 shows the percentage of H^+, H_2^+, and H_3^+ in a Hydrogen plasma as a function of pressure with the current in the ECR magnetic coils set at 220 Amps. This type of phase space aids in obtaining plasmas with one dominant species. H_3^+ is maximized at higher pressures, while H_2^+ dominates at lower pressures.

2.2.3 Langmuir Probe

Langmuir probe measurements determine the electron temperature (T_e), ion density (n_i), electron density (n_e), plasma floating potential (V_f), and the plasma potential (V_p) of the plasma. Figure 2-4 illustrates the Langmuir probe system. The probe tip is a very thin (~.07 mm in diameter) titanium or platinum wire. Sweeping the bias of the probe inside the plasma obtains a Langmuir Plasma *I-V* curve (Figure 2-5). The data are

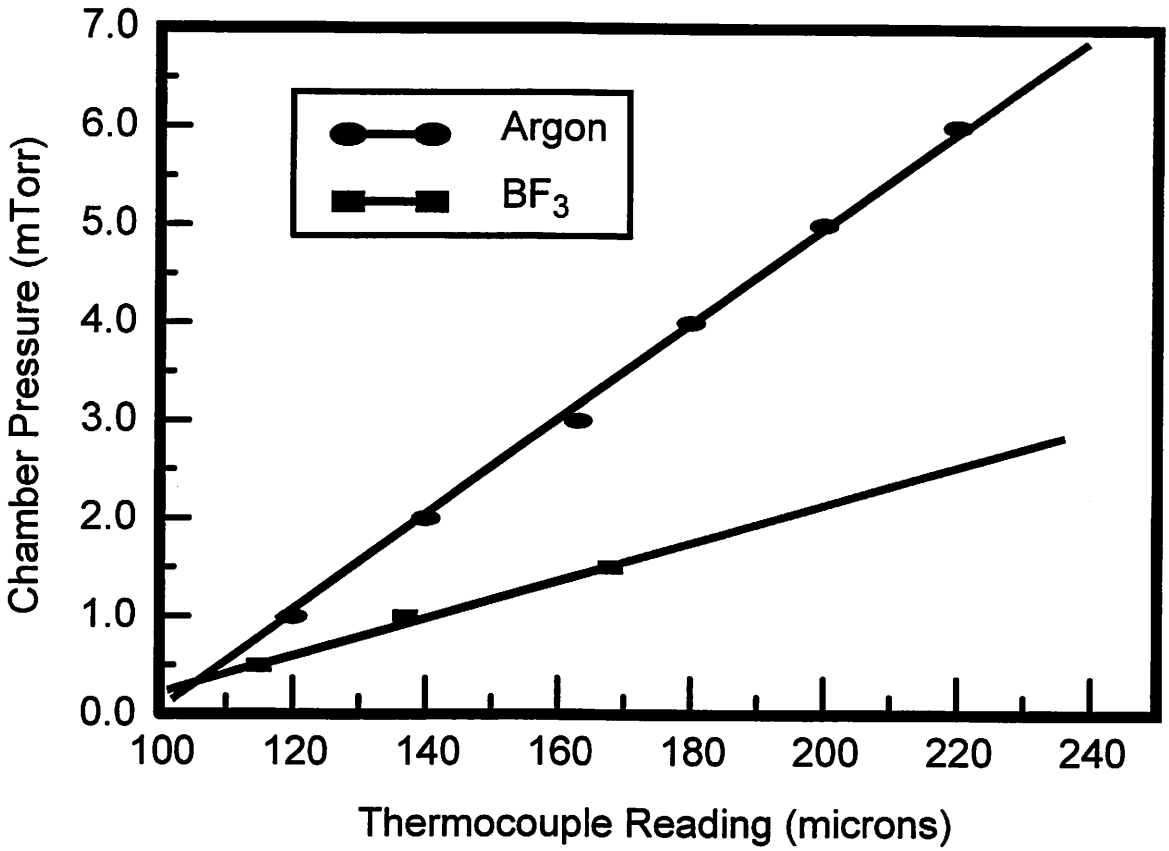

Figure 2-2 Thermocouple Pressure Calibration Curve

The thermocouple reading from the back of the turbopump correlates to the chamber pressure, and is calibrated for Argon and BF_3. The good fit validates interpolation.

acquired and analyzed in real time with the software package, Labview. Because there are no RF sources in the PIII system, RF filters and double probe techniques are unnecessary. The theory for extracting the plasma parameters from the Langmuir probe trace is found in [2-1].

The following assumptions simplify the analysis: the sheath width is much larger than the probe diameter; the mean free path is much greater than the sheath width (collisionless sheath), and the electron energy follows a single temperature Maxwellian distribution.

The flow chart shown in Figure 2-6 illustrates the process of extracting the plasma parameters from the Langmuir trace. First, V_f is the zero current point in the Langmuir

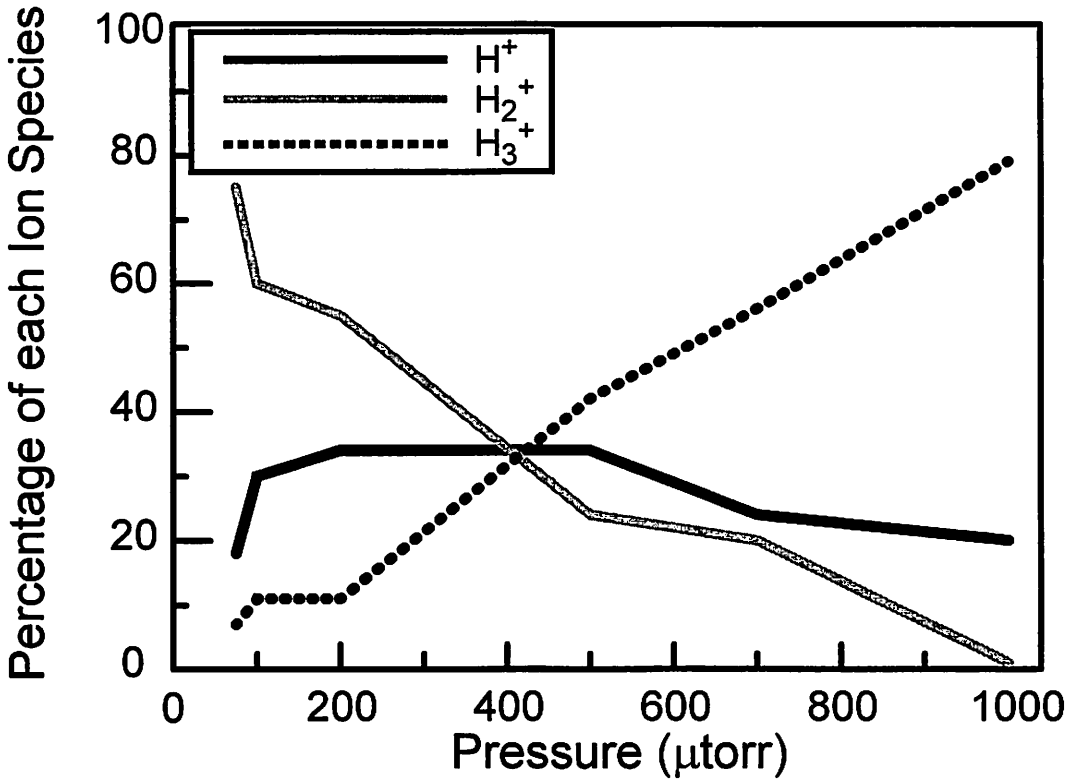

Figure 2-3 Mass Spectrometry of Hydrogen as a Function of Pressure

As pressure is increased, H_2^+ percentage decreases, while H_3^+ increases. Mass spectrometry helps identify optimal plasma conditions for PIII.

trace. The plasma potential is the positive voltage knee point in Langmuir trace. The exact inflection point is determined by either the maximum of the 1st derivative, or the zero crossing of the 2nd derivative. The 1st derivative method is preferred because it is less susceptible to noise. To obtain a noiseless 1st derivative, the Langmuir trace is fitted with a high order polynomial. Fits on the order of 10 perform acceptably. The current at V_p is the electron saturation current (I_{esat}), which is utilized for determining the electron density.

With V_f and V_p known, and assuming single temperature Maxwellain electrons, T_e is determinable. With probe biases approximately between V_f and V_p, the electron current is exponentially dependent on probe voltage. Assuming the electron current is much

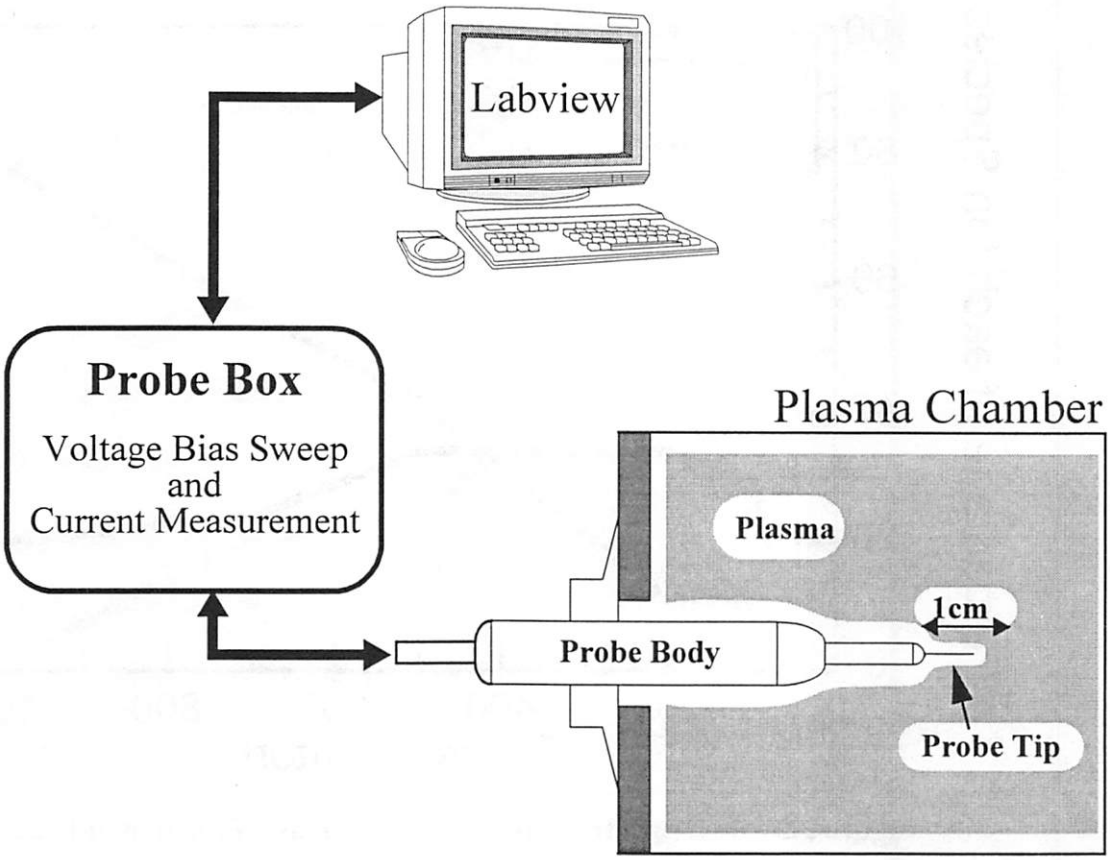

Figure 2-4 Langmuir Probe System

The Langmuir Probe contains a thin (0.07 mm in diameter) cylindrical wire probe. The probe box sweeps the probe bias, measuring the current. The computer, through a DAQ board, acquires the data, while the software package, Labview, analyzes the data in real time.

greater than the ion current, exponentially fitting the probe trace yields T_e. To ensure $I_{electron} >> I_{ion}$, the fit should only include voltages a couple of volts greater than V_f.

For cylindrical probes,

$$n_i = \sqrt{\frac{\frac{\mathrm{d}}{\mathrm{d}V}(I^2)}{(2 \cdot q \cdot a \cdot d)^2 \cdot \frac{2q}{m}}} \tag{2-1}$$

where a, d, m, are the probe radius, the probe length, and electron mass, respectively. The $\mathrm{d}I/\mathrm{d}V$ term should only be taken with $V < V_f$, to ensure negligible electron current.

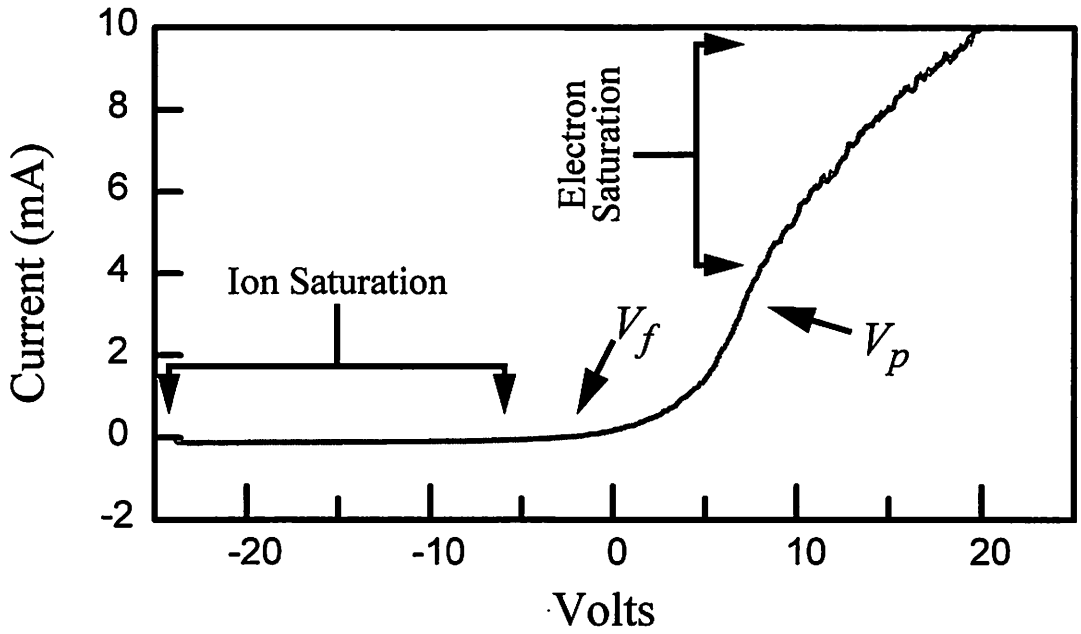

Figure 2-5 Typical Langmuir Probe *I-V*

I-V from a Langmuir Probe with a tip length of 1cm and a probe diameter of 0.07 mm in a 1 mtorr, 900 W Argon plasma. At negative voltages, the probe is biased in ion saturation, while with significantly positive voltages, the probe is biased in electron saturation. The x-axis crossing of the curve is V_f, and the knee in the positive current portion of the curve is V_p.

With multiple ion plasmas, it is necessary to calculate the electron density rather than the ion density. Moreover, for single component plasmas in which n_i is equal to n_e, the extracted value for n_e is often more accurate than that of n_i [2-2]. Electron density for cylindrical probes is:

$$n_e = \frac{I_{esat} \cdot 4}{(2 \cdot \pi \cdot q \cdot a \cdot d \cdot \mathrm{v}_e)} \tag{2-2}$$

$$\mathrm{v}_e = \sqrt{\frac{8(q \cdot T_e)}{\pi m}} \tag{2-3}$$

where I_{esat}, v_e are the electron saturation current, and the electron velocity, respectively. For this work all ion density measurements are actually obtained from the electron part of the trace. Even so, the values of the ion density should be interpreted as approximate density measurements.

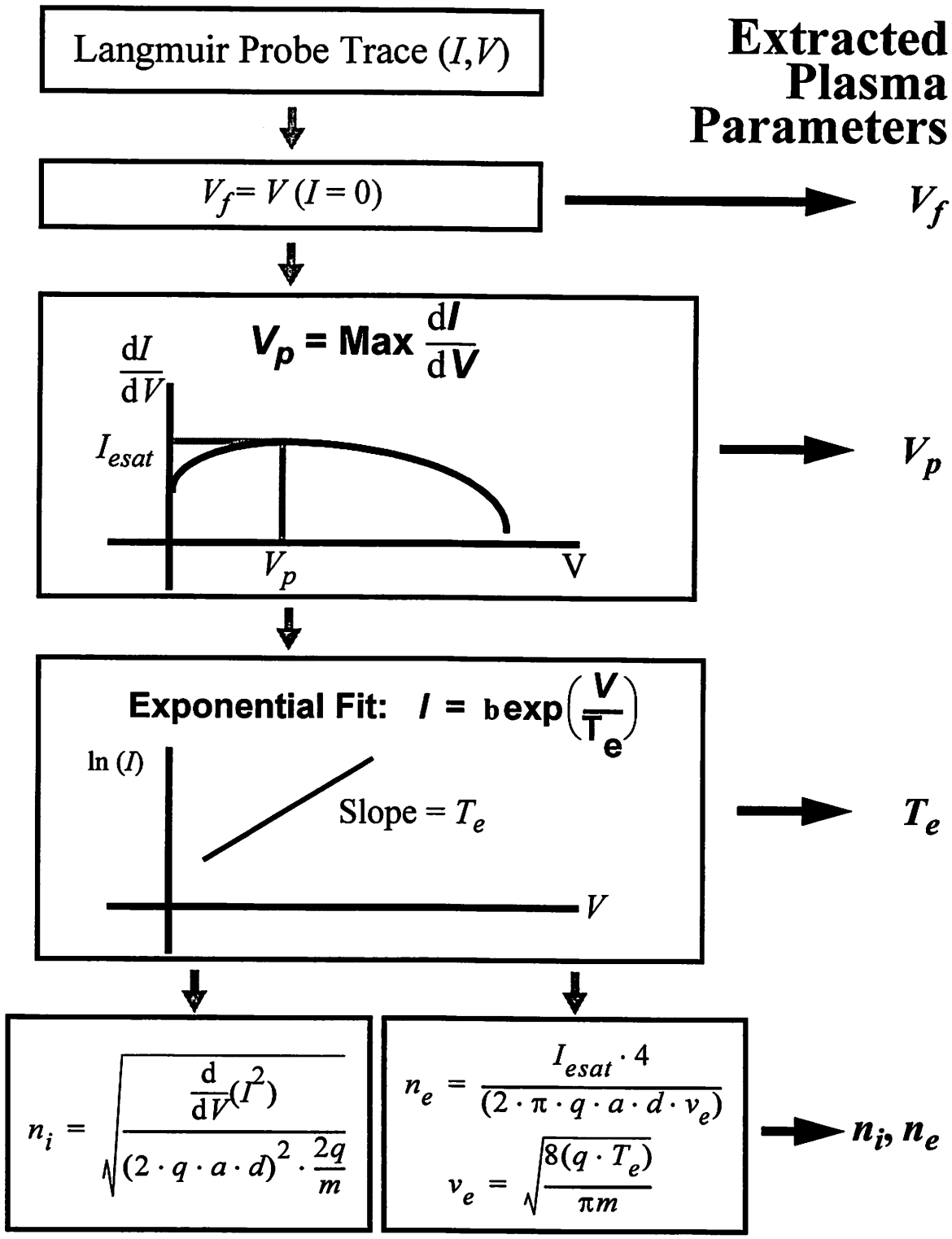

Figure 2-6 Flow Chart for Plasma Parameter Extraction

From a Langmuir Probe trace all plasma parameters, n_i, n_e, V_f, V_p, and T_e, may be extracted.

The most suspect assumption in this analysis is that the electrons are a single temperature Maxwellian. With many processes, the electrons follow a two temperature, or other non-Maxwellian distributions [2-2]. In these situations it is useful to calculate an electron energy distribution function (EEDF) from the second derivative of the probe trace. Integrating the EEDF also yields the electron density. This method is more accurate, but requires significantly more computation [2-3].

2.2.4 Fluoroptic® Probe Temperature Measurements

The substrate temperature during implantation is an important parameter of the PIII process. Excessive temperatures may result in significant diffusion of the implant species, such as hydrogen. For applications which require the formation of an amorphous layer, such as shallow junctions that amorphize the surface to prevent channeling [2-4], the substrate temperature must be kept below the silicon solid phase epitaxy temperature of 450 °C. Other applications, such as silicon on insulator formation (SPIMOX) [2-5] require a temperature around 600 °C for optimal buried oxide formation.

Conventional temperature measurements with a thermocouple prove difficult in PIII, since the implant bias is applied directly to the wafer. The electrical noise from the wafer bias swamps the thermocouple signal, yielding inaccurate results. Pyrometers are inconvenient, and require extensive calibration for temperatures below ~700 °C.

An alternative is the Fluoroptic® probe temperature measurement system provided by Luxtron Corporation. This electrically isolated, in-situ measurement system is illustrated in Figure 2-7. Fluorescent phosphor on the tip of the optical fiber acts as the temperature sensitive element. A xenon flashlamp excites the manganese-activated magnesium fluorogermanate phosphor. After the flash, the phosphor fluoresces at ~670 nm, with a decay time well correlated with temperature. The decay time at the measurement limits are approximately 5 ms at -200 °C, and 0.5 ms at 450 °C. A look-up table correlates the measured decay time with the phosphor temperature. The Fluoroptic®

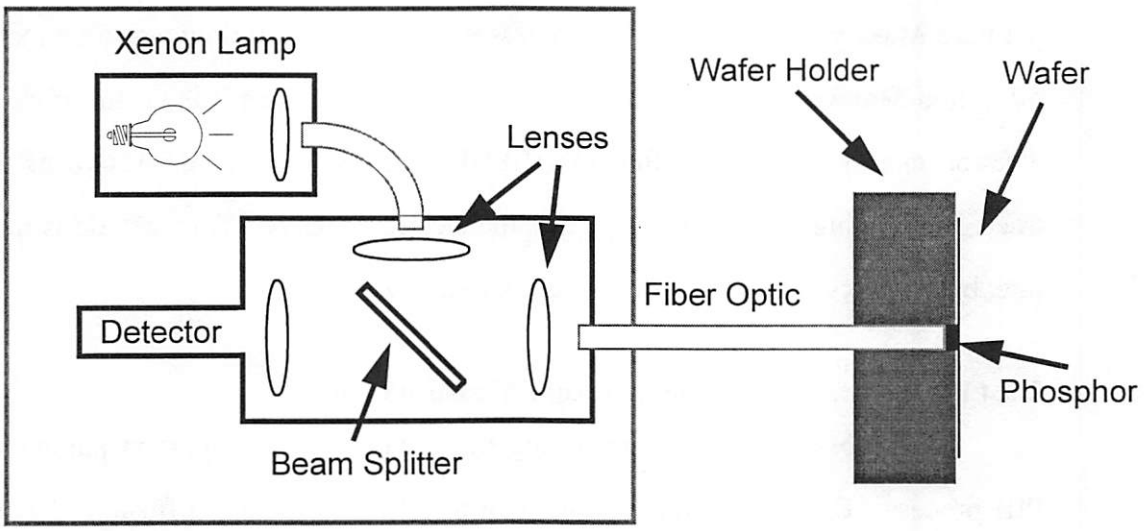

Figure 2-7 Fluoroptic® Probe Temperature Measurement System

The Fluoroptic® probe temperature technique provides in-situ, non-electrical real time wafer temperature measurements. A xenon lamp illuminates the phosphor in contact with the wafer, and a detector calculates the fluorescent decay time of the phosphor. A look-up table correlates decay time with temperature.

technique exhibits high accuracy and repeatability. Un-calibrated probes are accurate within ±2 °C, with calibrated probes improving this to ±0.1 °C. With 20 sample averaging, repeatability is ±0.1 °C, with increased averaging reducing this value. Employing the maximum sample rate of 10 flashes/second, successive measurements are 2 seconds apart.

The Fluoroptic® technique relies on the phosphor being in intimate contact with the wafer. In a vacuum system this contact requirement becomes more stringent. The primary measurement method achieves contact by pressing an optical fibre tipped with phosphor against the backside of the wafer (through a hole in the wafer holder). The phosphor is mounted on a transparent elastomer, allowing the tip to be pressed firmly without damage to the fiber. The maximum operating temperature for the elastomer is 250 °C. Higher temperature measurements require an alternative non-contact method. In this remote technique, phosphor is painted directly on to the back of the wafer. A bare fiber optic probe placed within 0.25 inches from the phosphor illuminates the phosphor and collects the flu-

orescence. This technique enables temperature measurement up to 450 °C. The main drawbacks of the remote phosphor technique is the difficult and expensive phosphor application process, and low signal amplitude. For these reasons remote phosphor is employed only if temperatures exceed 250 °C.

2.3 Wafer Bias

In PIII the wafer itself is biased to the implant voltage. The wafer bias creates a high voltage plasma sheath that accelerates and implants the plasma ions. The three typical bias modes are DC, pulsed, and RF. DC bias achieves the highest dose rate, but may only be used with conducting substrates. The main applications of DC bias are nitridation of metals, and SOI wafer formation. Pulsed or RF biases are preferred for insulating substrates, such that implant charge must is periodically neutralized. RF bias requires tremendous amount of power to drive the displacement currents, and therefore mainly applies to lower voltage implants. One example of an RF bias application is shallow junction formation, which requires low implant voltages [2-6]. Pulsed bias is chosen for all applications in which DC and RF are not feasible. Pulsed bias achieves implant voltages up to and exceeding 100 keV. In addition, pulsed bias is the most flexible with independent control over pulse frequency and duty ratio. With the Berkeley PIII reactor, pulsed and DC biases are the preferred modes of operation.

2.3.1 DC Bias

DC bias PIII is significantly simpler and achieves higher voltages for a lower cost than pulsed PIII. The Berkeley PIII system contains a 100 kV, 1 Amp DC power supply. Non-contact DC current measurements are achieved by measuring the DC magnetic field produced by the current in the high voltage line. With ECR plasma generation in the mirror magnetic field configuration, the secondary electrons from the implantation are focused by the source magnets onto the quartz microwave feed through window. Excessive heating and melting of the quartz by the secondary electrons may be suppressed by

either reducing the implant current or altering the magnets from the mirror configuration. Placing a partially drilled out shutter (quartz or aluminum) in between the source and main chambers attenuates the implant current. Optimizing the location, number, and size of the shutter holes, reduces the current without affecting the plasma uniformity significantly. An alternative approach is to alter the magnets from mirror to the cusp configuration or to use only one of the two coils. This significantly reduces the plasma density, and prevents electron focusing, but at the cost of more difficult plasma generation. We have found that for implant voltages less than 20 kV, the shutter alone reduces the secondary electrons sufficiently. For 20 kV to 40 kV applications, the magnets must be changed to cusp mode. Implants greater than 40 kV require both single coil configuration and shutter.

2.3.2 Pulsed Bias

Pulsed PIII eliminates the charge buildup that is observed during DC implants into insulating substrates. Pulsed bias also provides the flexibility of independent pulse voltage, pulse frequency, and duty factor. Either pulsed or RF biases are necessary for most semiconductor applications that contain charge sensitive gate oxides. Pulsed and RF biases do not completely eliminate charging. Chapter 11 examines in detail gate oxide charging during pulsed PIII.

Figure 2-8 depicts the PIII pulsing system. A 6 kV, 100 mA power supply and a pulse generator connect to the 25 kHz, 6 kV modulator. The signal travels across a transmission line containing various matching elements, terminating at the wafer holder. A 1000:1 high voltage probe connected to the wafer holder monitors the implant voltage, while a Rogowski loop around the signal line measures the AC current. Various circuits shunt the transmission line to ground for matching and protection purposes. A reverse-biased diode circuit prevents the line from going positive, while capacitor/resistor circuits control the signal bounce. The modulator performs best with a 50Ω terminating impedance. Since the impedance of the system depends on the plasma impedance, the matching

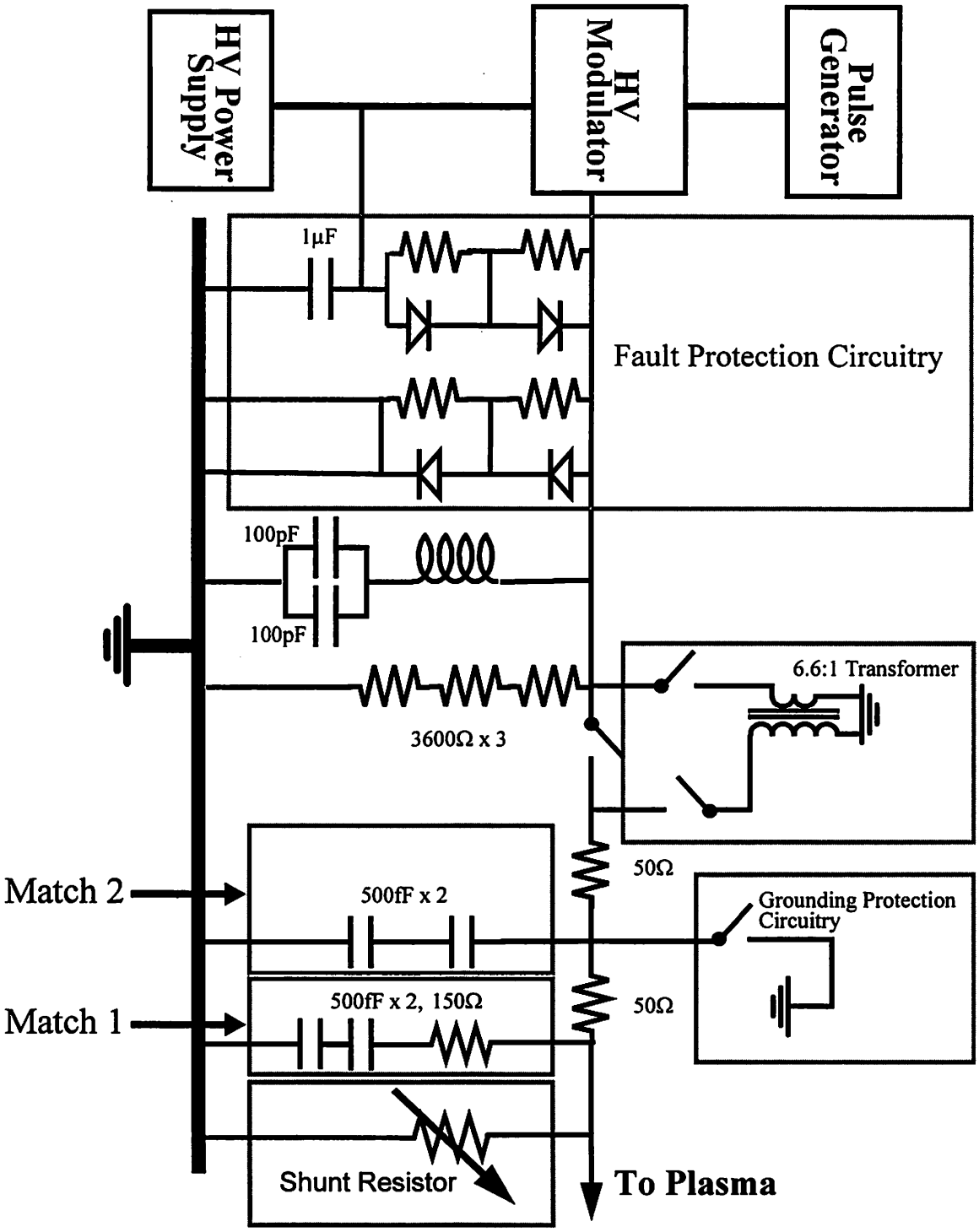

Figure 2-8 Pulsing Network

The PIII pulsing network including the matching network and fault protection circuits. The switches route the signal through the 6.6:1 transformer. The fault protection circuitry prevents positive voltages on the line or voltages more negative than the HV Power Supply. The variable resistor controls the fall time of the pulse, with lower resistances shortening the fall time.

network can not be optimized for all conditions. Typically, for the matching network shown, the fall time (defined as the time until the bias voltage decays to 500 V) is ~16.2 μs. When the modulator shuts off, it turns into an open, not a short to ground, forcing the capacitors to discharge through the 10 kΩ resistor (3 x 3.6 kΩ) or the plasma. Adding a shunt resistor decreases the discharge time significantly. Table 2-1 and Figure 2-9 show that the shunt resistor effectively reduces the fall time from over 10 μs to below 1 μs.

Although the fall times are reduced, the shunt resistor draws extra current that becomes restrictive at high pulsing frequencies.

$$I_{sh} = \frac{V_{pulse}}{R_{sh}} t_w f_p \tag{2-4}$$

where I_{sh}, R_{sh}, t_w, and f_p are the extra current drawn by the shunt resistor, the shunt resistance, the pulse width, and the pulsing frequency. Equation (2-4) assumes the current during the fall time may be neglected. In light of the 100 mA limit of the power supply, a 270 Ω resistor wastes 50% of the total current capacity of the power supply for a 6 kV implant at a duty factor of just 2.25% or 2.25 kHz with 1μs pulses. Therefore, a short fall time must be balanced with maximum pulse frequency. A 50Ω shunt resistor obtains the shortest fall time, but consumes excessive amounts of power. The 270Ω resistor, without either matching network, performs next best but suffers from signal bounce. The best compromise is using the 270 Ω resistor, with both matching networks, achieving a fall time less than 1μs.

Table 2-1 Shunt Resistor Effect on Pulse Fall Times

Shunt Resistor (Ohms)	Match 1	Match 2	Fall Time (μs)
none	yes	yes	16.22
50	yes	yes	0.12
270	yes	yes	0.88
270	no	no	0.36

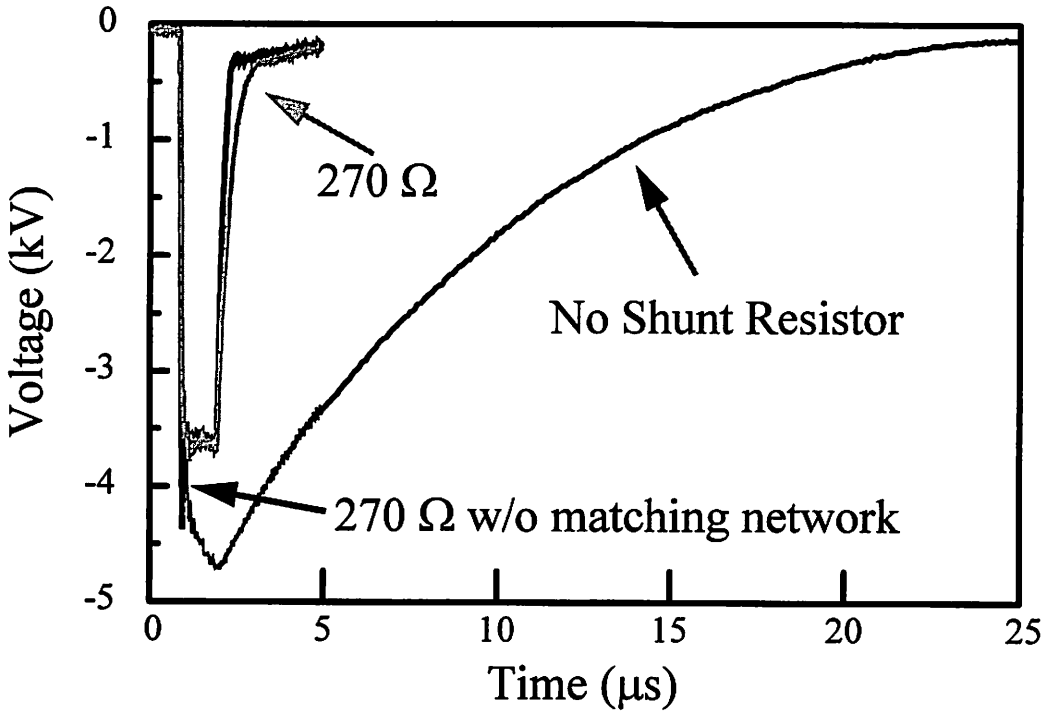

Figure 2-9 Effect of Shunt Resistor on Fall Time

Without an added shunt resistor, the pulses suffer from excessive fall time. Adding a 270 Ω shunt resistor reduces the fall time to below 1 μs. Removing some of the matching network, further reduces the fall time, but at the cost of signal bounce.

2.4 Wafer Cooling

For a number of applications it is imperative that the wafer temperature does not rise significantly during the implant. Gate oxide charging damage is a strong function of temperature, with damage doubling with only a 50 °C increase in temperature [2-7]. High temperatures are also detrimental during hydrogen implantation, since hydrogen diffuses rapidly at temperature above 300 °C.

The original PIII wafer holder mounted the substrate with clamping and did not have water cooling capabilities. Clamping provides good electrical contact, but does not provide substantial thermal contact in a vacuum. Figure 2-10 shows the temperature of the wafer with clamping for a variety of implant voltages. A small hole in the wafer holder allowed access to the back of the wafer for substrate temperature monitoring. With simple exposure to an Argon plasma (900 W), the temperature of the wafer rises to ~175 °C, while for 6 kV, 5 kHz implant, the temperature rises above the measurement limit of 250 °C within 50 seconds. Energy conservation explains these high temperature. Figure 2-11 displays the amount of energy radiated by a 4" wafer. The emissivity constant for the aluminum wafer holder is assumed to be 0.15. The emissivity of the silicion wafer changes with the carrier density, which itself depends on the temperature. Silicon emissivity is less than 0.2 for temperatures less than 400 °C, and equal to 0.72 for temperatures above 700 °C. Since the wafer is thermally floating, radiation is the dominant energy loss mechanism. But radiation power is not significant until the wafer temperature exceeds 400 °C.

With a simple plasma exposure, the plasma deposits ~2 W of power onto a 4" substrate (900 W, 1 mtorr, Argon plasma), translating into a temperature slightly below 200 °C. A 6 kV, 5 khz pulsing deposits ~55 W, producing a temperature in excess of 450 °C. In addition to allowing the wafer to reach unacceptable temperatures above 250 °C, simple clamping results in run to run temperature variations. There is always some random thermal contact between the wafer and the holder which provides a path for heat loss. Typi-

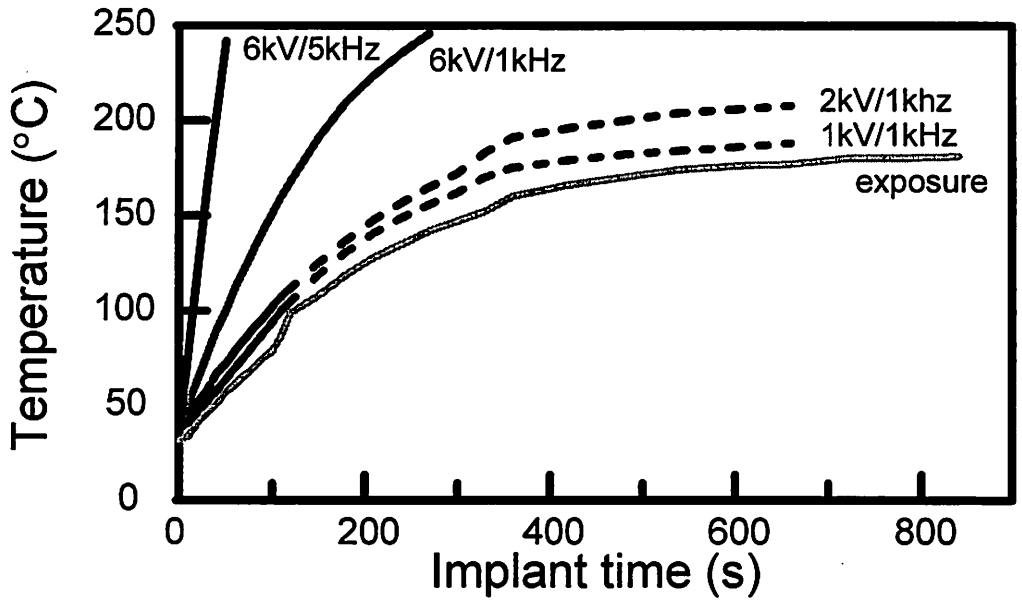

Figure 2-10 Wafer Temperature during Implantation

The wafer temperature with simple clamping. There is very little thermal contact with simple clamping, allowing high wafer temperatures. The implant voltages range from 0 volts (exposure only) to 6 kv, with pulsing frequencies of 1 kHz, and 5 kHz, and a 1μs pulse width.

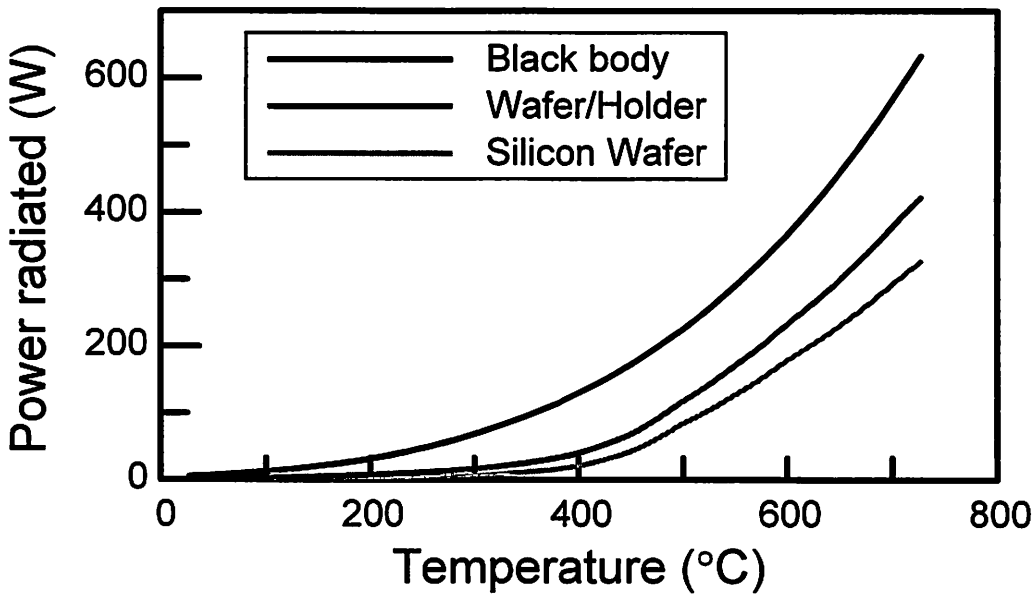

Figure 2-11 Radiation Losses as a Function of Temperature

Radiation emission for a silicon 4" wafer, a 4" black body, and a wafer bonded to an aluminum wafer holder combined. Radiation does not emit significant power until the temperature exceeds 400 °C.

cally this is small, but varies with the clamping pressure. For the 6 kV, 1 kHz case, strong clamping pressure reduced the temperature rise to 100 °C after 900 s, but the temperature was still rising. Simple clamping is thus inadequate for implants that require temperature control.

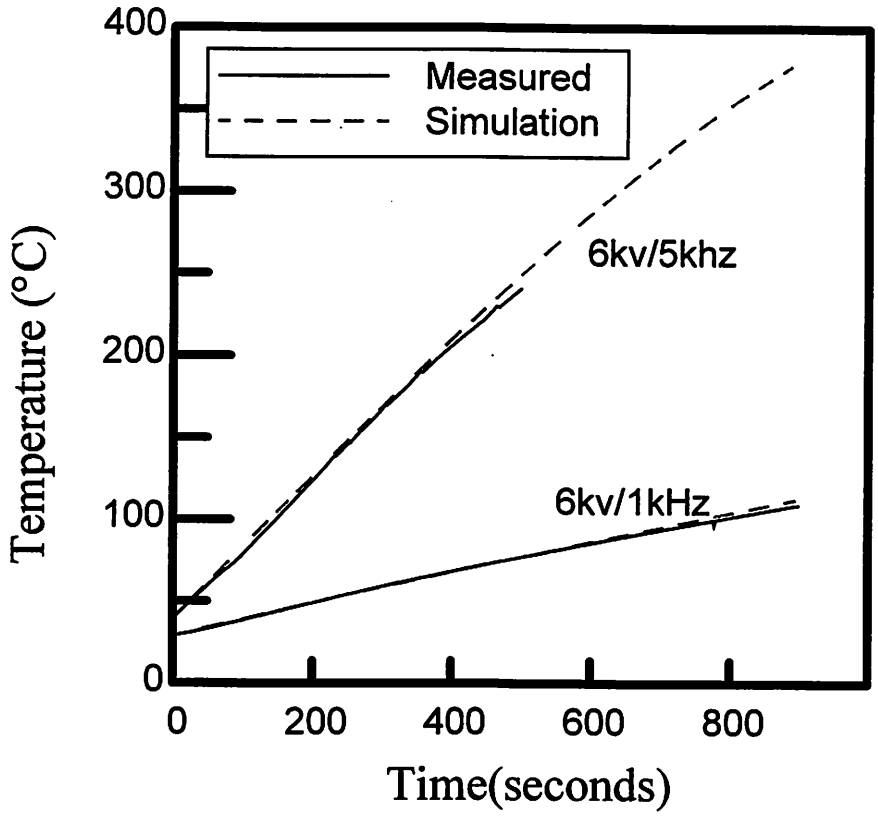

Figure 2-12 Temperature Rise with Silver Paste

With silver paste holding the wafer to the wafer holder, the rate of temperature rise is reduced significantly.

2.4.1 Clamping Alternatives

The basic problem is to identify a wafer holding method that provides good thermal and electrical contact. There are three basic alternatives to simple clamping that provide good thermal contact. The most complicated is helium backside cooling. With this technique, the holder only contacts the wafer edges, while Helium gas passes along the backside of wafer. This is the most effective technique, but usually requires an electro-

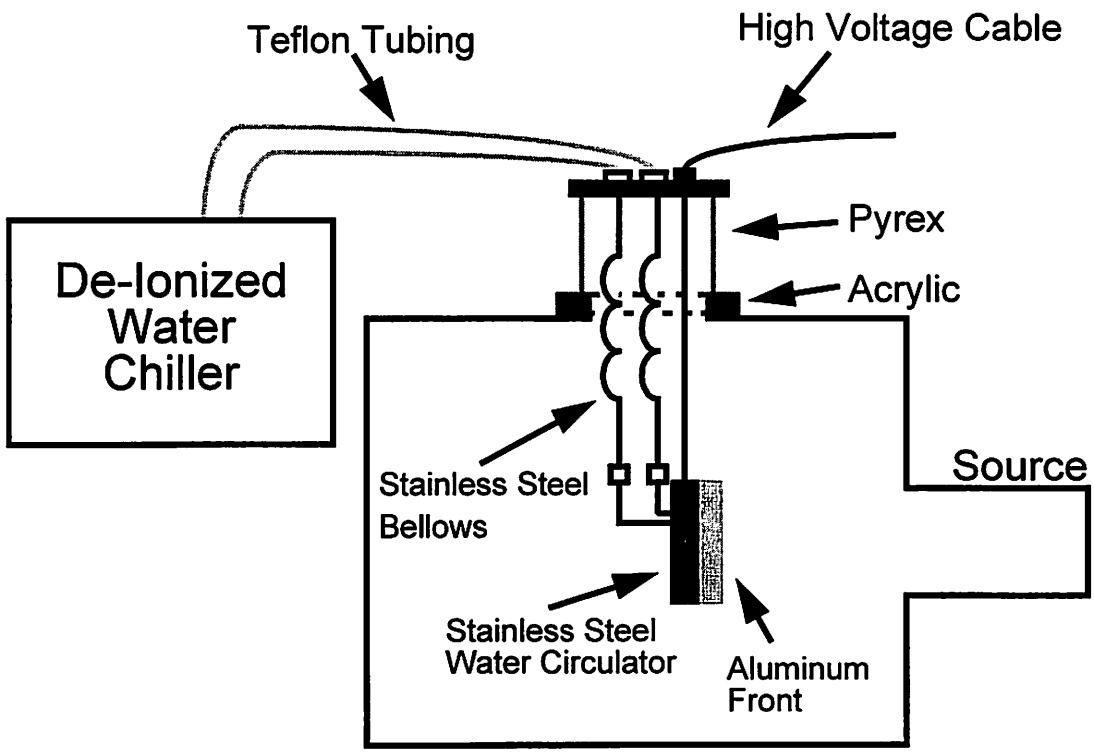

Figure 2-13 De-ionized Water Chiller System

De-ionized chilled water travels through Teflon tubing and stainless steel bellows to the wafer holder. The stainless steel backing of the holder contains the circular pattern water circulator, while the Aluminum front minimizes contamination. The high voltage and water lines are isolated from the grounded chamber by Pyrex and Acrylic feedthroughs. The bellows allow flexibility for loading/unloading the wafer holder.

static chuck, and is best suited for use with a load-lock. A second alternative is to place a compressible membrane between the wafer and the holder. Clamping presses the wafer against the membrane and the wafer holder. This membrane provides good thermal contact between the wafer and the holder. However, this technique requires a custom made membrane dependent on the exact clamping pressures. The disadvantage for these techniques is that both backside cooling and thermal membrane are unacceptable for implanting odd size silicon pieces. An alternative is to use a metal bond to physically attach the wafer to the holder. The metal bond material must not outgas in the plasma and must easily dissolve in a solvent after processing. Silver paste meets these needs. Silver paste mounting requires more steps than simple clamping, and this takes up to 2 hours per sam-

ple. First, the paste is spread on the backside of the wafer and the wafer holder. The wafer is pressed and twisted against the holder, ensuring good contact. The silver paste solvent is dried with a 150 °C anneal. Before loading the wafer into the chamber the holder must be cooled back to room temperature. An acetone ultrasonic bath dissolves the silver paste effectively after implantation. Figure 2-12 depicts the temperature rise with silver paste, comparing that to simulation predictions. The model assumes that radiation is the only loss mechanism, and the wafer and holder and in perfect thermal contact. Appendix C.1 lists the simulation source code in Matlab format. Silver paste doesn't eliminate heating but simply slows the temperature rise. Instead of the power being absorbed by only the wafer, it is now absorbed by the wafer and wafer holder. The thermal mass of the wafer/ wafer holder system is nearly 30 times greater than the wafer only. This reduces the rate of temperature increase, but not the final equilibrium temperature. The rate of radiation loss does not increase much with silver paste, since the surface area of the holder is only approximately 50% greater than the wafer area. The emissivity of aluminum (0.06 - 0.2) is also much less than silicon (0.1 - 0.7). This analysis is supported by the simulation matching the measured temperature rise as shown in Figure 2-12. Since the wafer heats to over 200 °C, good thermal contact alone is not enough to solve the temperature issue, but requires active wafer holder cooling.

Water cooling a PIII wafer holder requires special care, since the wafer holder is at high voltage (up to 100 kV for the Berkeley system). Figure 2-13 illustrates the Berkeley PIII water cooling setup. The cooling system must not provide an electrical path between the wafer holder and the water chiller. This requires de-ionized water and non-conducting tubing outside the chamber. To ensure electrical isolation up to 100 kV, the chiller de-ionizes the water to above 10 MW/cm, and teflon tubing is used. To minimize outgassing, only stainless steel tubing is used inside the chamber with a VCR fitting between the teflon and the stainless steel tubing. The stainless tubing connects to the back of the wafer holder with VCO fittings. Water circulates within the wafer holder in a spiral pattern. The

DI chiller controls the water temperature from 0 - 60 °C, with less than 0.1 °C variance. The chiller provides flows greater than 3 gallons per minute at 50 psi, providing up to 1.7 kW of cooling capacity. Because of the high capacity of the chiller, typical implants of a couple hundred watts or less can be executed without significantly increasing the wafer temperature above the water coolant temperature. Fluoroptic® probe measurements detect no increase in the back-side substrate temperature for all performed implant conditions with wafer holder water cooling.

References

[2-1] M. A. Lieberman, and A. J. Lichtenberg, "Principles of Plasma Discharges and Materials Processing." John Wiley and Sons, Inc., New York, pp. 171 - 188, 1994.

[2-2] V.A. Godyak, R. B. Piejak, and B. M. Alexandrovich, "Probe Diagnostics of non-Maxwellian plasmas," *Journal of Applied Physics*, vol. 73, no. 8, pp. 3657-3663, April, 1993.

[2-3] M. A. Lieberman, and A. J. Lichtenberg, "Principles of Plasma Discharges and Materials Processing." John Wiley and Sons, Inc., New York, pp. 181, 1994.

[2-4] E. C. Jones, S. Im and N. W. Cheung, "Ultra-shallow P^+/N Junctions Formed by SiF_4 Preamorphization and BF_3 Implantation using Plasma Immersion Ion Implantation," *Beam Solid Interactions: Fundamentals and Applications Symposium,* pp. 255-259, 1993.

[2-5] S. S. K. Iyer, X. Lu, J. Lie, J. Min, Z. Fan, P. K. Chu, C. Hu, and N. W. Cheung, "Separation by Plasma Implantation of Oxygen (SPIMOX) Operational Phase Space," *IEEE Transactions on Plasma Science*, vol. 25, no. 5, pp. 1128-1135, Oct. 1997.

[2-6] B. Mizuno, M. Takase, I. Nakayama, and M. Ogura, "Plasma doping of boron for fabricating the surface channel sub-quarter micron PMOSFET," 1996 Symposium on VLSI Technology, pp. 66-67, 1996.

[2-7] K. Scheugraf and C. Hu, "Effects of Temperature and Defects on Breakdown Lifetime of Thin $Si0_2$ at Very Low Voltages," *IEEE Transactions on Electron Devices*, Vol. 41, no. 7, pp. 1227-1232, July, 1994.

[2-8] P. J. Timans, "Emissivity of Silicon at Elevated Temperatures," *Journal of Applied Physics*, vol. 74, no. 10, pp. 6353-6364, November 15, 1993.

3 PIII Coupled Plasma Model

3.1 Introduction

A PIII model which describes the implant current-applied voltage relationship can further the understanding and applications of PIII processing. This model must address the relationship between applied bias, plasma generation, and implant dose. In addition the model can predict gate oxide charging damage, and the energy profile of PIII implants. Finally, the model must be concise enough to aid process engineers in optimizing PIII implants. The Coupled PIII Model (*CPM*) developed in this chapter addresses all of these issues. This *CPM* is modular in nature, with the core modules accounting for 1st order effects. Modules accounting for 2nd order effects such as multiple ion species and collisions in the sheath are developed in later chapters.

Figure 3-1 illustrates the core modules of the *CPM*. To properly model PIII, the plasma, the IC structures, and the substrate bias must all be taken into account (Figure 3-1). The plasma model determines the plasma ion current and electron current to the wafer surface, and the ion impinging energies. The IC structure models calculate all the voltages and currents in the wafer device structures, especially the gate oxide voltage and tunneling currents. Finally the substrate bias model drives the implant. Solving all three models simultaneously, and allowing them to interact, forms a complete picture of the PIII system.

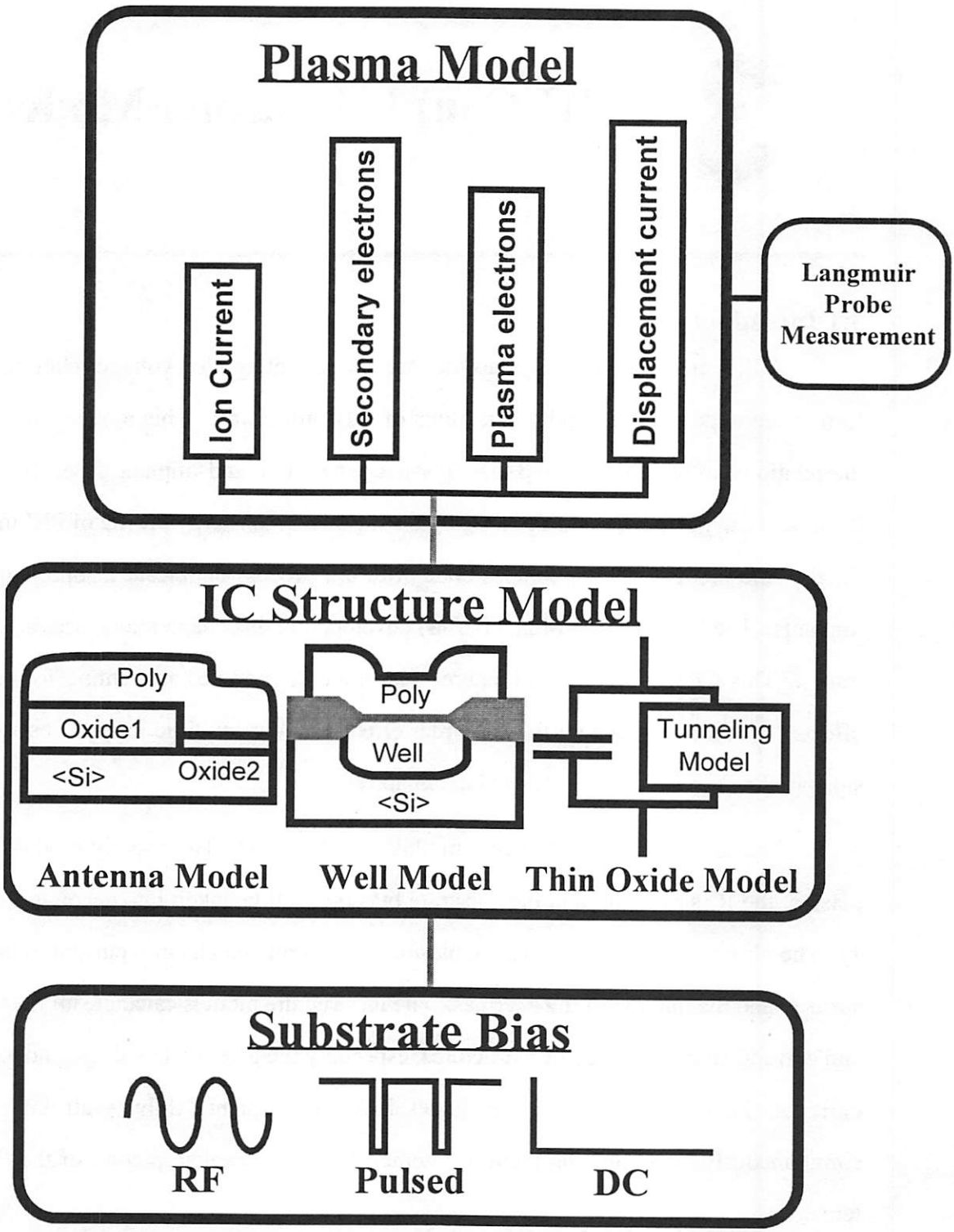

Figure 3-1 The PIII Model

The three sections of the PIII model: plasma model, IC structure model, and the substrate bias. A Langmuir probe trace provides all the parameters for the plasma model. The PIII model is inherently modular, and can accommodate more complicated structures simply.

3.2 Plasma Model

The plasma model calculates the time dependent plasma currents consisting of four main elements.

$$I_{total} = I_i + I_{se} + I_e + I_{disp} \tag{3-1}$$

where I_{total}, I_i, I_{se}, I_e, and I_{disp}, are the total plasma current, the plasma ion current, the secondary electron current, the plasma electron current, and the plasma displacement current, respectively.

The ion current calculations are based on a Quasi-Static Child law sheath. Chester [3-2] first determined the flux of ions from a moving sheath region, while Scheur et al. [3-3] and Lieberman and Stewart [3-4, 3-5] extended the model to PIII. Several simplifying assumptions make the calculations tractable. Since the sheath widths are typically less than the ion mean free path, a collisionless sheath is assumed. With a nominal implant pressure of 1 mtorr, the low energy ion mean free path for Argon is 3 cm, while the mean free path increases to 14 cm with 10 keV ions [3-8]. With an ion density of 10^{10} cm^{-3}, the maximum sheath width for a 10 kV, 1μs pulse is 2.4 cm. Therefore, typically the sheath width is less than the mean free path.

The ion transit time across the sheath (~100 ns) is much less than the nominal pulse width (1μs). This allows the accelerating field to be assumed frozen during ion transit. Combining, the two previous assumptions, we can assume that the ions bombard the surface with the instantaneous bias potential. Since the Quasi-Static Child Law sheath forms in tens of nanoseconds, it is assumed to exist instantaneously and for all time.

By applying these assumptions, the governing equations for the plasma ion current can be derived [3-4] and are summarized here. Since, the plasma ion current density (J_i) satisfies the Child Law for all time,

$$J_i = \frac{4}{9}\varepsilon_o\left(\frac{2q}{M}\right)^{1/2}\frac{(V_o)^{3/2}}{s^2} \tag{3-2}$$

where q, V_o, M, s are the electronic charge, applied voltage, ion mass, and sheath width respectively. The flux crossing the sheath boundary also defines the ion current,

$$J_i = qn_i\left(\frac{ds}{dt} + v_s\right) \tag{3-3}$$

where n_i is the ion density, and v_s is the distributed sheath velocity for ECR plasmas. v_s is related to the Bohm velocity as

$$v_s = K \cdot u_B = K \cdot \sqrt{\frac{qT_e}{M}} \tag{3-4}$$

where, K is a machine dependent parameter. For the Berkeley PIII tool K equals 1.1.

Combining Equation (3-2) and Equation (3-3) results in a differential equation for the sheath width.

$$qn_i\left(\frac{ds}{dt} + v_s\right) = \frac{4}{9}\varepsilon_o\left(\frac{2q}{M}\right)^{1/2}\frac{(V_o)^{3/2}}{s^2} \tag{3-5}$$

Solving Equation (3-5) for the sheath width, and substituting this result into Equation (3-2) determines the plasma ion currents.

Implanting ions with high voltage ejects secondary electrons. The large sheath potential accelerates these electrons away from the wafer surface, amplifying the total positive wafer current,

$$J_+ = J_i(1 + \gamma(V_i)) \tag{3-6}$$

where J_+ is the total positive current density, and $\gamma(V_i)$ is the secondary electron yield as a function of ion implant energy. The secondary electron yield for Aluminum has been determined [3-9] as

$$\gamma \approx k\sqrt{V_i} \tag{3-7}$$

where k is an empirical fit parameter. For Al, $k \approx 0.0696$. Secondary electron yields for other substrate materials is given in [3-10]. For Al, secondary electron current exceeds the ion current for voltages greater than 200 V. For a 10 keV ion, γ is nearly 7. Therefore, secondary electrons dominate positive charge deposition, and must be included when calculating gate oxide charging.

To simplify the plasma electron current modeling, the electrons are assumed to follow a single temperature Boltzmann relationship. With Boltzmann electrons, the plasma electron flux to a surface can be expressed as

$$J_e = \frac{1}{4} q n_i v_e e^{\frac{-(V_p - V_s)}{T_e}} \tag{3-8}$$

where v_e, V_p, and V_s are the electron velocity, the plasma potential, and the surface potential, respectively. If the wafer surface voltage is more positive than the plasma potential, Equation (3-8) is no longer valid, and J_e is capped at $\frac{1}{4} q n_i v_e$. One situation where this can occur is during dielectric implantation just after the fall time of the bias pulse (see Section 5.3.2).

Lastly, both the displacement currents due to the changing sheath potential and the changing sheath capacitance are calculated.

$$J_{disp}(t) = C_s(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t}(V_s(t)) + V_s(t) \cdot \frac{\mathrm{d}}{\mathrm{d}t} C_s(t) \tag{3-9}$$

where J_{disp}, C_s, and V_s are the displacement current density, sheath capacitance, and sheath voltage, respectively. Displacement currents exist during the large applied bias voltage swings during the rise and fall times of the pulse. Displacement currents are usually negligible, but may become important for fast pulsing frequencies and RF wafer biases. Equations (3-3) - (3-9) form the fundamental equations for the PIII Plasma Model.

To solve the above equations requires several plasma parameters: ion density (n_i), electron temperature (T_e), plasma potential (V_p) and floating potential (V_f). All of these values can be extracted from a single Langmuir probe measurement. Previous experiments demonstrate that this plasma model accurately determines the plasma currents [3-1, 3-6].

3.2.1 Wafer Structure Models

Modeling gate oxide charging is achieved by solving the plasma equations and the mathematical descriptions of the device structures simultaneously [3-1, 3-6, 3-7]. Most structure models are built from simple components of resistors, capacitors, inductors, diodes, and transistors. The thin gate oxide model consists of a capacitor in parallel with the models for Fowler-Nordheim and Direct tunneling current (Figure 3-2), which are known to be the main cause of charging damage for thin gate oxides [3-11]. A buried oxide layer or dielectric substrate is modeled as a capacitor in series with the substrate bias. The well model consists of a diode in parallel with a capacitor. This capacitor includes both junction and transit time capacitances.

3.2.2 Substrate Bias

The substrate bias is included by specifying voltage or current sources. Non-ideal source effects can be included by adding in transmission lines, internal source resistors, and transformers. All the simulations in this work utilize an ideal voltage source.

3.3 Coupled and De-Coupled Models

Combining all three modules - the plasma model, the wafer structures model, and the substrate bias - and solving them simultaneously yields a fully coupled plasma damage model. Under the full implementation, all the elements of the model interact and are solved simultaneously (Figure 3-3). In the *CPM*, the only interaction between the wafer structures and substrate bias with the plasma is through the surface voltage of the wafer.

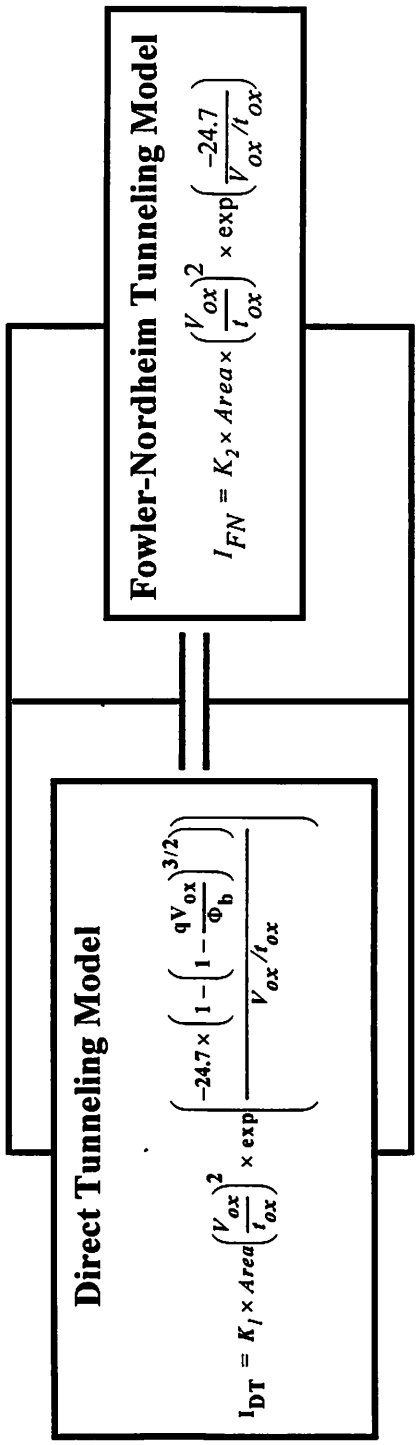

Figure 3-2 Gate Oxide Model

The gate oxide model includes a capacitor in parallel with both the Direct tunneling and Fowler-Nordheim tunneling models. If $V_{ox} < 3.2V$, direct tunneling applies, if $V_{ox} > 3.2V$ Fowler Nordheim tunneling dominates. K_1 and K_2 are constants.

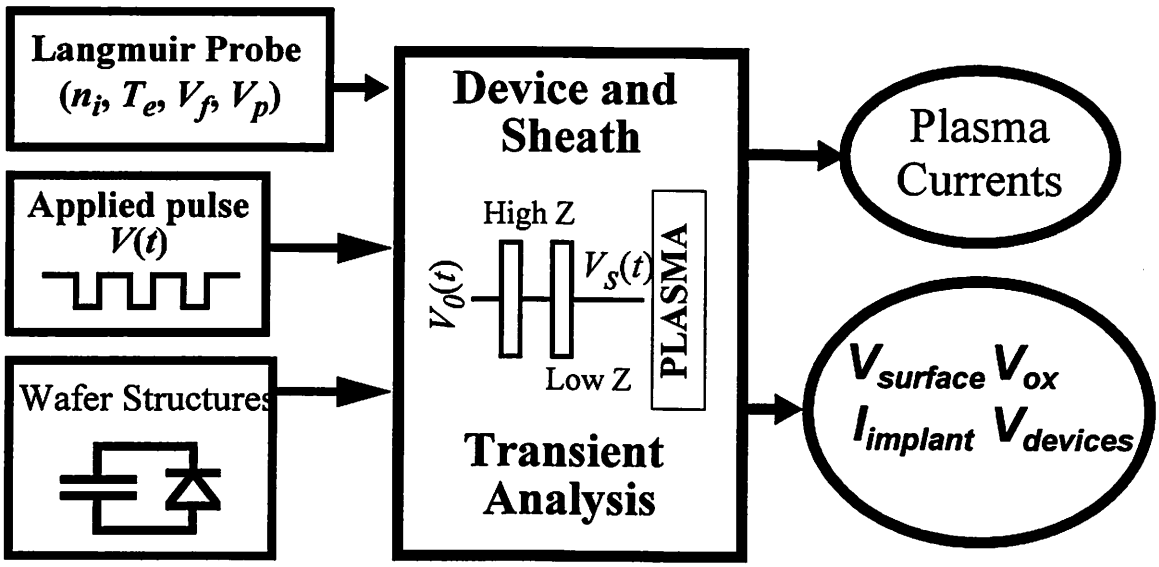

Figure 3-3 Fully Coupled PIII Model

The Fully Coupled PIII Model solves the sheath, plasma currents, surface currents, and surface voltages simultaneously. This is imperative with high impedance (Z) substrates.

The plasma currents, except for I_e, are fairly insensitive to small variations in the surface voltage. Therefore, if V_s is nearly equal to V_o, then the model may be split into two parts, the sheath transient analysis (*STA*) and the device transient analysis (*DTA*) (Figure 3-4).

The *STA* calculates J_i, s, J_{disp}, and J_{se}, with V_s equal to V_o, ignoring the effect of the wafer structures. Then, the *DTA* solves for J_e and the voltages and currents in the device structures by allowing V_s to vary from V_o. This approach is valid as along as V_s in the *DTA* does not vary much from V_o (less than tens of volts). If V_s varies significantly from V_o, than the fully coupled model must be solved. For example, this later situation applies when the substrate contains thick dielectrics, typically greater then 5 μm, which occur with buried oxide layers or dielectric substrates.

With the de-coupled model, the plasma currents, except for J_e, are independent of the wafer structures, allowing for the creation of a library of solutions for different plasma

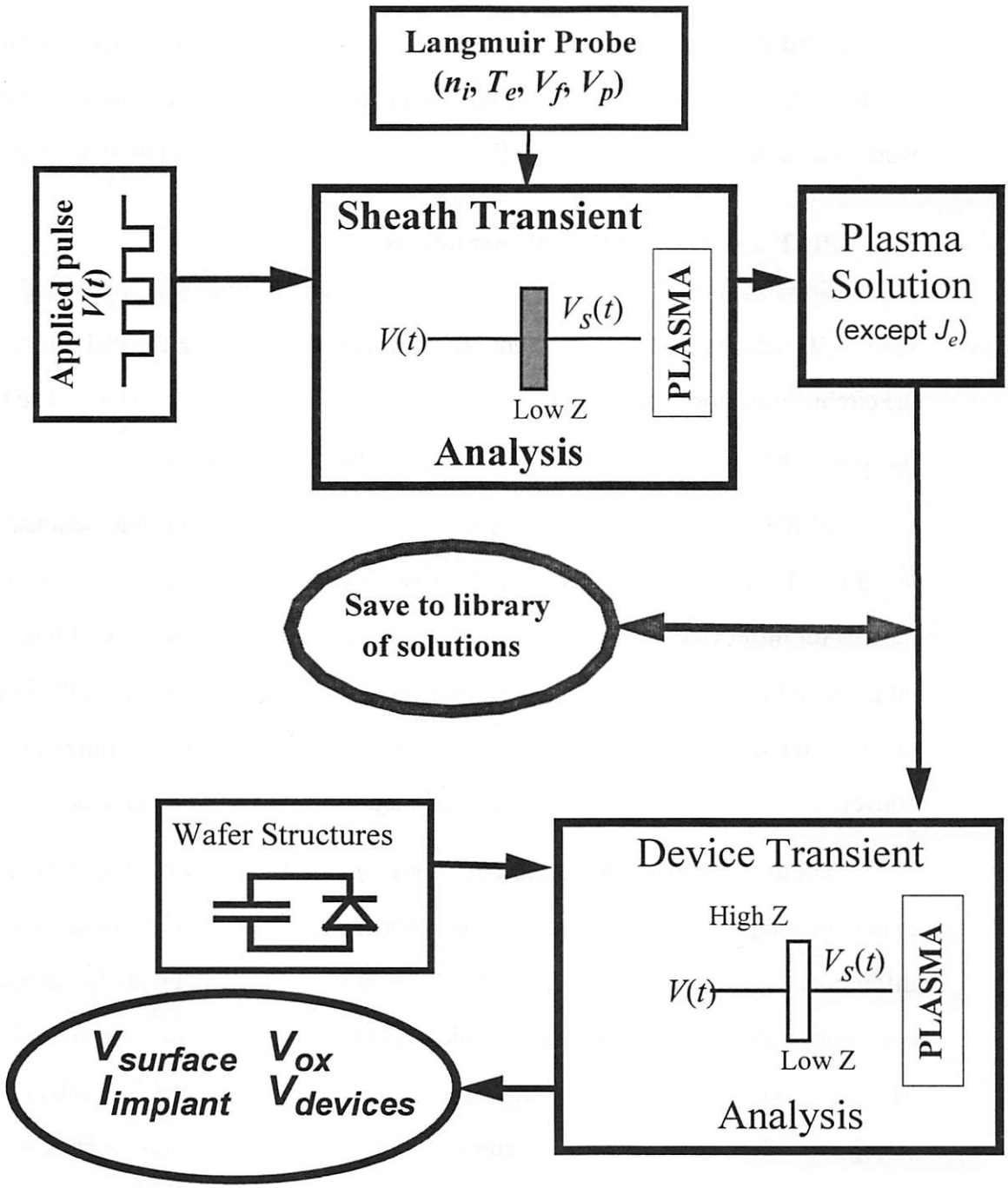

Figure 3-4 De-Coupled Modular PIII Model

In the de-coupled approach, the sheath thickness and the plasma currents, except J_e, are solved independently of the wafer structures, increasing computational speed by an order of magnitude. The de-coupled method applies when the surface voltage is nearly equal to the applied bias, implying a conducting substrate. The presence of a capacitive substrate precludes the use of the de-coupled model. The sheath and plasma currents solutions are stored in a library, and thus avoiding computing them more than once. Then, for each different wafer structure set-up, the library is accessed for the plasma currents.

conditions and substrate biases. Then, for each wafer structure setup, the library is accessed for the plasma currents. The storage of plasma solutions and the de-coupling of the differential equations, allows for up to a magnitude increase in computational speed, while maintaining accuracy. Appendix B elaborates on this concept with an example.

3.3.1 SPICE and MATLAB Implementations

Solving all the modules simultaneously requires a numerical differential equation solver. We have chosen two different programs for implementing the PIII plasma model, the circuit simulator SPICE and the general purpose matrix solver MATLAB. The best performing program depends on the exact set of simulation conditions.

SPICE has clear advantages for calculating plasma-device interactions. Besides being a full circuit simulator with a well known interface, SPICE contains built-in circuit models for the device structures and substrate-structure models. Because of this, it is trivial to extend the model and include additional devices. On the other hand, the implementation of the differential equation in SPICE is cumbersome, and sometimes experiences convergence difficulties in the fully coupled mode with dielectric substrates.

On the other hand, MATLAB contains an extensive library of simple to use differential equation solvers. If one method does not converge, it is trivial to switch to another differential equation solving method. MATLAB, however, does not include electrical models. They must be programmed in, which becomes laborious for complicated circuits. MATLAB's advantage is with the de-coupled mode, where its extensive collection of file storage functions makes the construction of the plasma solution library seamless.

In view of the above mentioned facts, SPICE solves the coupled model best, and allows easy introduction of complicated wafer surface structures and non-ideal sources, while MATLAB excels with the de-coupled model, and the construction of plasma solution libraries.

3.4 Conclusions

The complete PIII model consists of the plasma, wafer structures, and substrate bias modules. The plasma model is constructed with only physical equations, with no fitting parameters. A single Langmuir probe measurement supplies all the necessary physical parameters. The plasma model computes the ion current, plasma electron current, secondary electron current, and displacement currents. For typical implant energies, secondary electron ejection dominates the positive charge deposition and gate oxide charging. A fully coupled model, where all currents and voltages are solved simultaneously, is necessary with high impedance substrates, while a de-coupled approach suffices for conducting substrates. The SPICE platform excels at solving the fully coupled model and at incorporating complicated surface structures, while the MATLAB platform performs best in the de-coupled mode. In all, the *CPM* fully characterizes PIII implants, predicting implant energies, dose, surface currents, and gate oxide charging.

References

[3-1] W. En and N. W. Cheung, "Modeling of Charging Effects in Plasma Immersion Ion Implantation", *Nucl. Instr. and Meth. in Phys. Res.* B96, pp. 435-439, 1995.

[3-2] J. K. Chester, "Effects of Expanding Sheaths on the Operation of Mercury-Arc Rectifiers", *Journal of Science and Technology*, vol. 37, no. 1, pp. 2-10, 1970.

[3-3] M. Shamim, J. T. Scheuer and J. R. Conrad, "Measurements of Spatial and Temporal Sheath Evolution for Spherical and Cylindrical Geometries in Plasma Source Ion Implantation", *Journal of Applied Physics*, vol. 69, pp. 2904-2908, 1991.

[3-4] M. A. Lieberman, "Model of Plasma Immersion Ion Implantation," *Journal of Applied Physics*, vol. 66, no. 7, pp. 2926-2929, 1989.

[3-5] R. A. Stewart and M. A. Lieberman, "Model of Plasma Immersion Ion Implantation for Voltage Pulses with Finite Rise and Fall Times," *Journal of Applied Physics*, vol. 70, no. 7, pp. 3481- 3487, 1991.

[3-6] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *Journal of Vacuum Science and Technology B*, vol. 12, no. 2, pp. 833-837, 1994.

[3-7] W. En, M.A. Lieberman, N. W. Cheung, "Comparison of Experimental Target Currents with Analytical Model Results for Plasma Immersion Ion Implantation," *IEEE Trans. on Plasma Science,* Vol. 23, No. 3, pp. 415-421, 1995.

[3-8] A.V. Phelps, "Cross Sections and Swarm Coefficients for Nitrogen Ions and Neutrals in N_2 and Argon Ions and Neutrals in Ar for Energies from 0.1ev to 10keV." *J. Phys. Chem. Ref. Data*, Vol. 20, No. 3, pp. 557-573., 1991.

[3-9] B. Szapiro and J. J. Rocca, "Electron Emission from Glow-discharge Cathode Materials due to Neon and Argon Ion Bombardment," *Journal of Applied Physics*, vol. 65, no. 9, pp. 3713-3716, 1989.

[3-10] W. En and N. W. Cheung, "A Novel Method for Determining the Secondary Electron Yield Dependence on Ion Energy for Plasma Exposed Surfaces" *IEEE Transactions on Plasma Science*, vol. 24, no. 3, pp.1184-1887, June 1996.

[3-11] X. Y. Li, J. T. Hsu, P. Aum, V. Bissessur, D. Chan, C. R. Viswanathan, "Reliability Study of Plasma Etching Damage in ULSI Process", *Materials Reliability in Microelectronics IV, Pittsburgh, PA, USA, Material Research Society*, pp. 51-56, 1994.

[3-12] G. A. Emmert, "Model for Expanding Sheaths and Surface Charging at Dielectric Surfaces during Plasma Source Ion Implantation", *Journal of Vacuum. Science and Technology B*, vol. 12, pp. 880-883, 1994.

4 PIII Implant Energy Distribution

4.1 Introduction

Simulation of the implantation process to predict the implant depth and profile is necessary for users of implantation systems. With PIII, the unavoidable energy spread of the implanted ions complicates profile simulation. In contrast, conventional implantation has a minuscule energy spread. There are four main sources of energy spread in a PIII implant: the pulsing of the wafer bias, ion collisions in the sheath, multiple implant species, and voltage buildup on a dielectric substrate surface. This chapter describes the sources of energy spread, identifies which sources dominate under different conditions, and outlines some methods for estimating the implant energy distribution.

4.2 Pulsing Effect on Energy Spread

Pulsing the wafer bias always introduces a spread in the implant energy distribution. The main culprits are the finite rise time, matrix sheath implantation, finite fall time, and the expansion and contraction of the ion sheath. Depending on the implant conditions any one of the four sources may dominate the energy spread.

The typical pulse is shown in Figure 4-1. Rise times are generally fast, less than 50 ns for many pulsers. The rise time is defined as the time for the applied voltage to rise from -50 V to within 50 V of the full pulse voltage. The fall time is usually considerably larger, and may range from less than a microsecond to tens of microseconds, depending on

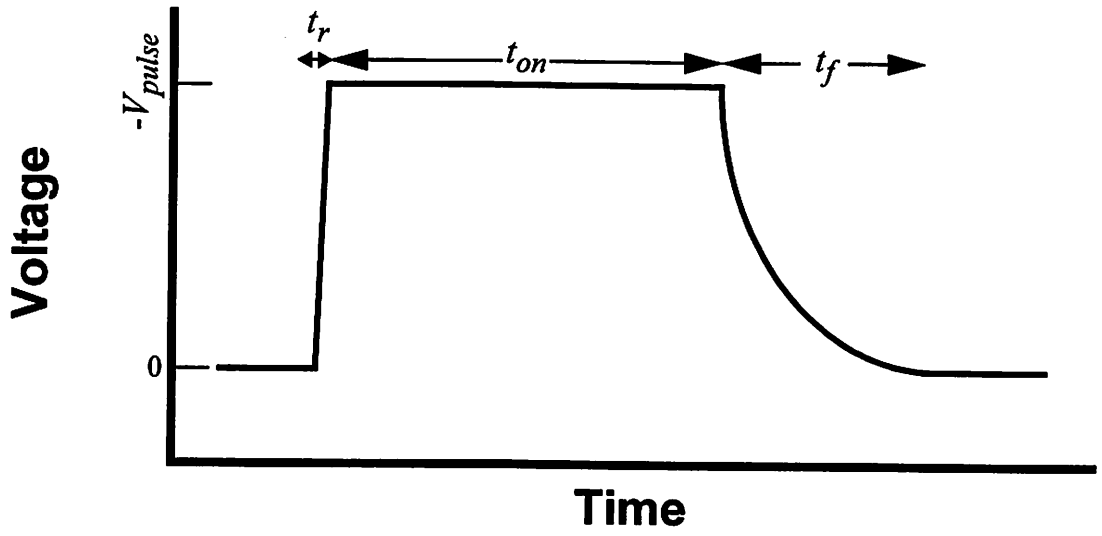

Figure 4-1 Definitions of Pulse Parameters

A typical voltage pulse depicting the rise time (t_r), the on time (t_{on}), and the fall time (t_f). The maximum voltage is defined as V_{pulse}. For most pulsing systems the rise time is much shorter than the fall time.

the pulsing network. For this work, fall time is defined as the time for the applied voltage to drop from the full pulse voltage to -50 V. The on-time for typical implants ranges from a microsecond to tens of microseconds. All the simulations in this chapter use an Argon plasma.

4.3 Matrix Sheath Implantation

When a pulse is coupled to the plasma, a sheath, named the matrix sheath, develops on the time scale of the reciprocal of the electron plasma frequency (usually greater than 1 GHz). Integrating Poisson's equation twice, assuming uniform space charge over the region, yields the matrix sheath width

$$s_m = \sqrt{\frac{2\varepsilon_0 V_{pulse}}{q n_i}} \tag{4-1}$$

where s_m is the matrix sheath thickness, which can be determined for a known V_{pulse} and n_i. None of the ions uncovered by the matrix sheath implant with the full energy (defined

as the implant energy of ions with a DC applied substrate bias of V_{pulse} in steady state). They implant with the energy determined by the voltage distribution of the matrix sheath. This is calculated by substituting x for s_m, where x is the distance from the substrate, and $V(x)$ for V_{pulse} in Equation (4-1). For example, an ion that happens to be half way between the edge of the matrix sheath and the substrate, will implant with $\frac{q \cdot V_{pulse}}{4}$, rather than an energy of qV_{pulse}.

During the pulse, the sheath expands from the initial matrix width to the steady state full Child law value (Equation (3-2), Figure 4-2). During this stage, the ion current consists of two components, those ions that diffuse across the sheath boundary and those ions that are uncovered by the expanding sheath. In contrast to the matrix sheath ions, all of these ions implant with the full pulse voltage, since they traverse the entire sheath (assuming the transit time is short compared to the sheath expansion rate).

To calculate the percentage of ions that implant with less than the peak energy from the matrix sheath effect, the amount of ions in the matrix sheath is compared to the ions uncovered by the expanding sheath and the ions that diffuse across the sheath boundary:

$$P_{low} = \frac{n_{matrix}}{n_{matrix} + n_{expand} + n_{diff}} \tag{4-2}$$

$$n_{matrix} = n_i \cdot s_m \tag{4-3}$$

$$n_{expand} = n_i \cdot s(t) \tag{4-4}$$

$$n_{diff} = n_i \cdot u_b \cdot t \tag{4-5}$$

$$P_{max} = 1 - P_{low} \tag{4-6}$$

where P_{low}, n_{matrix}, n_{expand}, $s(t)$, n_{diff}, and P_{max} are the percentage of ions that implant with less than the peak energy due to the matrix effect, the ions that are uncovered by the sheath matrix, the ions that are uncovered by the expanding sheath, the sheath width as a

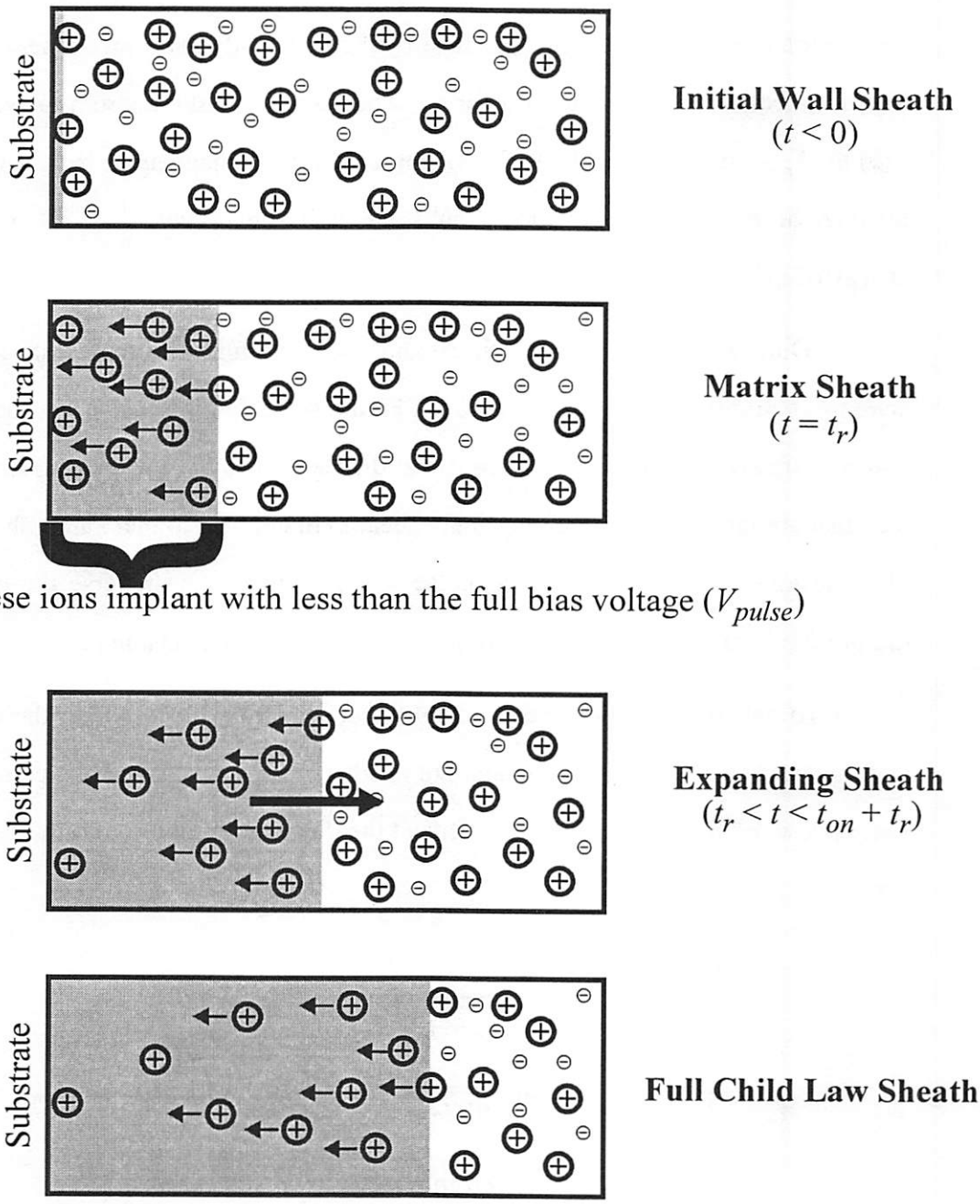

Figure 4-2 Stages of PIII: Matrix, Expanding, and Child Law Sheaths

Before the bias pulse, a small wall sheath exists. After the rise time, a matrix sheath forms nearly instantaneously, enveloping enough ions to support the pulse voltage. At this point the ion density is assumed to be constant everywhere. All the ions uncovered by the matrix sheath, implant with less energy less than V_{pulse}. During the on time, the sheath expands, uncovering more ions. Because these ions traverse the entire sheath (assuming a small transit time), they implant with the full bias voltage (V_{pulse}). The ion density in the sheath transforms from the constant density of the matrix sheath to the $x^{-2/3}$ relationship of the steady-state Child Sheath. If the pulse is held on long enough, the sheath expands out to the full Child Law Sheath thickness.

function of time, the ions that diffuse across the sheath boundary, and the fraction of ions that implant with the full pulse energy, respectively.

To calculate n_{expand} from Equation (4-4) requires $s(t)$. Lieberman [4-1] calculated the expansion rate for the sheath during PIII

$$arctanh\left(\frac{s(t)}{s_c}\right)-\left(\frac{s(t)}{s_c}\right) = \frac{u_b t}{s_c} + arctanh\left(\frac{s_m}{s_c}\right)-\frac{s_m}{s_c} \tag{4-7}$$

where s_c is the full Child law sheath thickness. Equation (4-7) tracks the sheath expansion from matrix sheath formation out to the full Child law sheath. The rise time is assumed instantaneous, since the typical rise time is less than 50 ns, much less than normal pulse widths. A finite rise time would simply delay the sheath expansion slightly, but the effect is small.

At the end of the pulse on-time, the ions that are still in the sheath will not implant with the peak energy, since the bias voltage will drop before they implant. These ions should be subtracted from n_{expand}. Assuming a Quasi-Static Child Law sheath, the number of ions in the sheath at the end of the pulse on time is:

$$n_{smax} = q\int_0^{s_{max}} \frac{4}{9}\frac{\varepsilon_o}{q}\frac{V_{pulse}}{(s_{max})^2}\left(\frac{x}{s_{max}}\right)^{-2/3} dx = \frac{4}{3}\frac{\varepsilon_0}{q}\frac{V_{pulse}}{s_{max}} \tag{4-8}$$

where n_{smax} and s_{max} are the ions in the sheath at the end of the on time and the maximum sheath thickness, respectively. Except for short pulse width biases with large peak voltages, the sheath width reaches its maximum value at the onset of the fall time. For the examples considered in this work, the actual value of s_{max} is computed, but is often nearly identical to the sheath width at the commencement of the fall time.

Figure 4-3 plots P_{max}, as a function of implant voltage and time for an ion density of 10^{10}cm^{-3}. As t_{on} increases, n_{expand} and n_{diff} increase, while n_{matrix} remains constant.

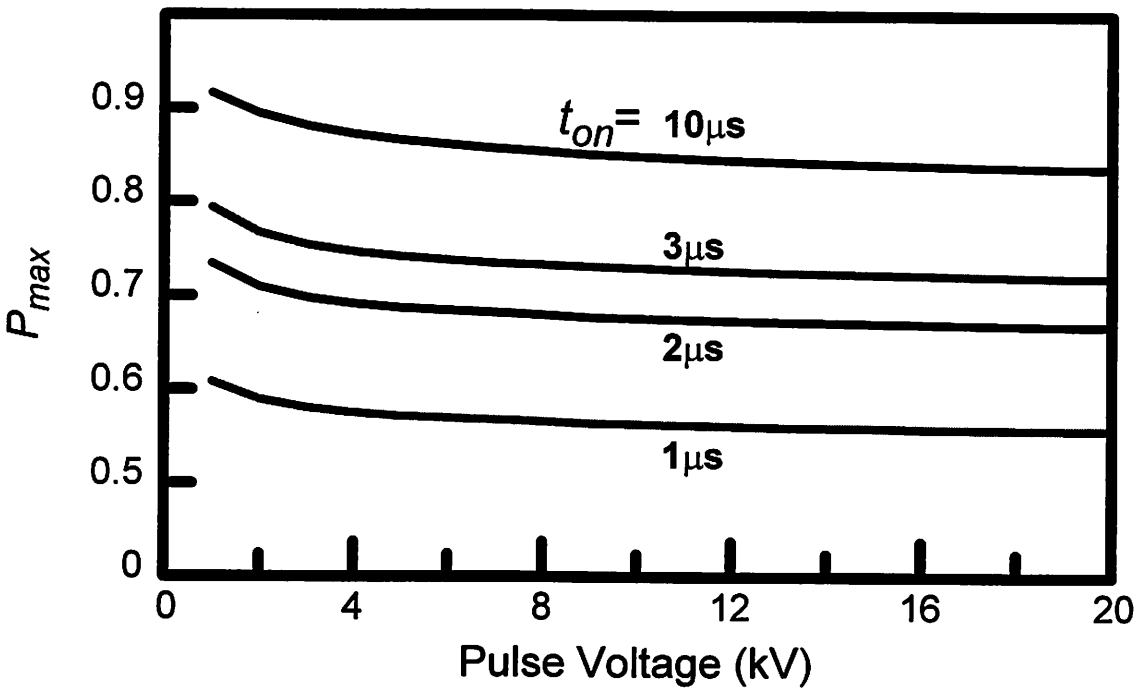

Figure 4-3 Low Energy Implantation from Matrix Sheath

The y-axis is P_{max}, which is the fraction of ions that implant with an energy corresponding to the full applied bias, V_{pulse}. A matrix sheath forms nearly instantaneously after the application of a voltage pulse. All the ions in the matrix sheath implant with less than the peak energy (V_{pulse}). Longer pulse widths dilute the matrix contribution to the implant dose, increasing $1\text{-}P_{low}$. Higher implant voltages form thicker matrix sheaths, increasing the low energy implant component. The ion density for this calculation is $10^{10}cm^{-3}$.

Therefore, as the pulse widths lengthen, P_{max} increases. As the implant voltage increases, n_{matrix} increases while n_{diff} is constant, resulting in a lower P_{max}.

Figure 4-4 shows how P_{max} changes with ion density. n_{matrix} is proportional to $(n_i)^{-1/2}$, but n_{diff} is proportional to n_i; so as the ion density increases, P_{low} decreases (Equation (4-2)). The effect is more apparent with longer pulse widths, since n_{diff} also scales with time, while n_{matrix} is independent of time.

Reduced implant energies from the matrix sheath are significant for short pulses and lower ion densities. Implants with longer pulse widths or higher ion densities diminish the matrix sheath contribution to the implant dose, and hence to the energy spread of a

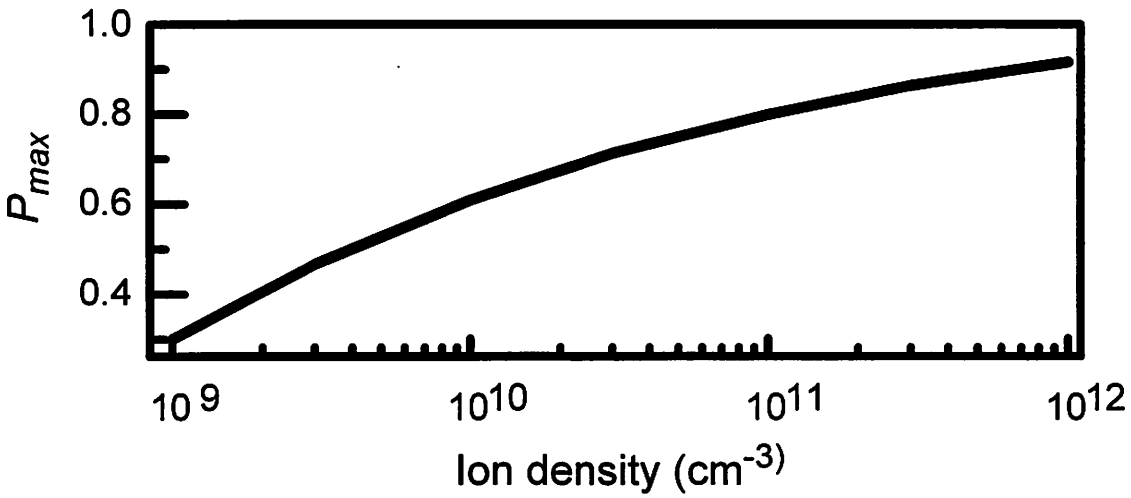

Figure 4-4 Energy Spread as a Function of Plasma Ion density

Increasing the ion density significantly reduces the low energy implantation from the matrix sheath. The graph is for a 1μs / 1 kV ideal pulse. The tend is similar for longer pulse widths.

PIII implant. Overall, the matrix sheath fundamentally limits the implant energy uniformity during a PIII implant pulse.

4.4 Fall time implantation

The second source of low energy ions, are those that implant during the fall time. Obviously, any ions that implant while the bias voltage is less than the peak voltage will implant with less than V_{pulse}. Since fall times can be comparable or longer than the on time, its contribution to low energy implantation can be significant. The charge implanted during the fall time equals the number of ions that diffuse across the sheath boundary, plus the number of ions in the sheath at the onset of the fall time (n_{smax}). The theoretical maximum sheath collapse rate is determined by the ion diffusion velocity (the Bohm velocity), yielding a minimum collapse time of $\frac{s_{max}}{u_B}$ [4-2, 4-3]. If the sheath collapses at its maximum rate, no ions will cross the sheath boundary, and the only ions implanted during the fall time will be those already in the sheath. Therefore, to minimize implantation during the fall time, t_f must be less than $\frac{s_{max}}{u_B}$. Shorter fall times have no effect on the dose or the

energy distribution. (An infrequent exception occurs if the voltage pulse falls slowly to begin with and then decreases quickly near the end of the fall time. This is an uncommon situation, since most fall times follow an exponential relationship common to RC delays.) Figure 4-5 graphs $\frac{s_{max}}{u_B}$ for a variety of implant times and voltages. For a 10 kV, 1 μs implant, the fall time simply needs to be less than 6.5 μs to minimize the fall time effect on energy spread.

For infinite pulse widths the maximum allowable fall time is simply $\frac{s_{max}}{u_B}$, since the sheath stops expanding at the full steady-state Child law thickness. Increasing the plasma ion density results in thinner sheaths, reducing the maximum fall time significantly. Therefore, higher ion densities require shorter fall times.

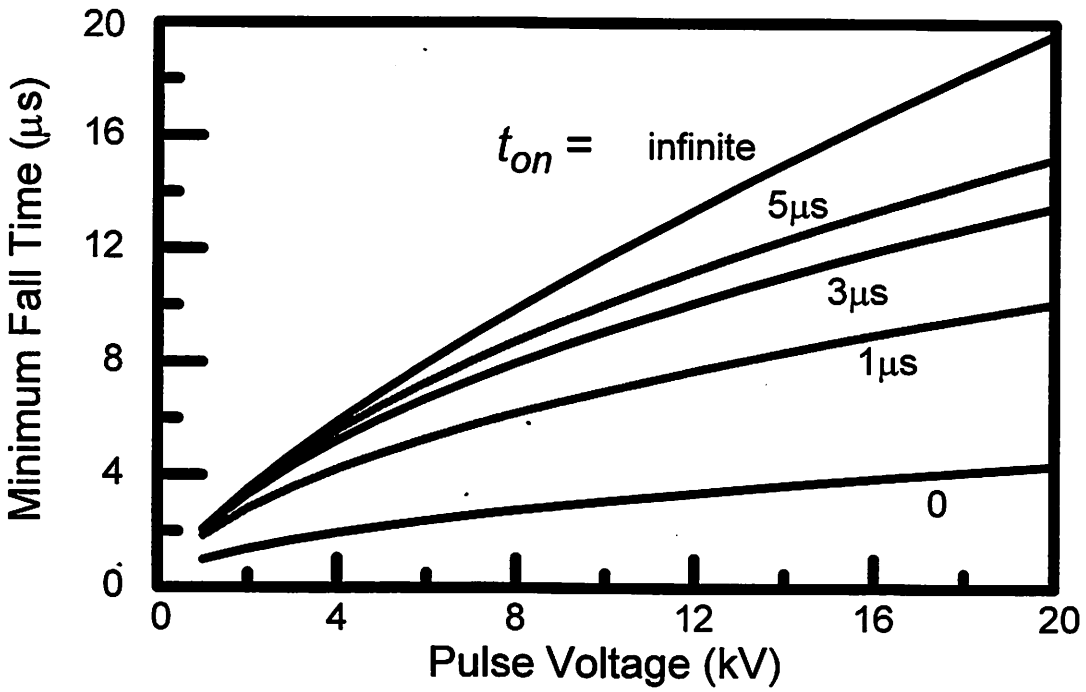

Figure 4-5 Sheath Collapse Time for Different Pulses

The minimum sheath collapse time for different pulse widths and pulse voltages. A fall time less than the minimum allowable fall time minimizes the implant flux during the fall. Any further reduction in the fall time has no effect. The n_i for this calculation is $10^{10}cm^{-3}$.

The energy distribution of the ions implanted during the fall time can be estimated by separating the fall times into three different time scales. If the fall time is fast, e.g. $t_f \ll s_{max}/u_b$, the implant energies for the fall time ions may be assumed to follow the Quasi-Static Child Law Relationship, i.e.,

$$V(x) = -V_o\left(\frac{x}{s_{max}}\right)^{4/3} \tag{4-9}$$

$$n(x) = \frac{4}{9}\frac{\varepsilon_o}{q}\frac{V_o}{(s_{max})^2}\left(\frac{x}{s_{max}}\right)^{-\frac{2}{3}} \tag{4-10}$$

where $V(x)$ is the implant energy for the ions the ions at point x at the beginning of the fall time. The second regime occurs with long fall times where $t_f \gg s_{max}/u_b$. In this situation, the ions in the sheath at the onset of the fall time simply implant with nearly the peak energy (V_{pulse}), while all the ions that diffuse across the sheath boundary during the fall time implant with the instantaneous bias voltage (which is necessarily less than the full voltage). The third time scale lies between the first two time periods. For this range of fall time, an analytical relationship may be solved for simple cases [4-8], but is not tractable for more complex fall times. The energy distribution may be simulated, though, by simply keeping track of the ions as they traverse the sheath and the sheath edge in the spirit of Particle in Cell (PIC) simulations. This may be simplified by noting V_{sheath} when an ion enters the sheath, and assuming the ion implants with the instantaneous V_{sheath}.

It is significantly simpler to estimate the sum total of the ions that implant with less than the peak energy, rather than calculate the exact implant energy distribution for this intermediate fall time scale. The ions that implant during the fall time are

$$n_{fdiff} = n_i u_b t_f - n_i s_{max} \qquad \left(t_f > \frac{s_{max}}{u_b}\right)$$

$$0 \qquad otherwise \tag{4-11}$$

and

$$n_f = n_{fdiff} + n_{smax} \tag{4-12}$$

where n_{fdiff}, s_{max}, n_f, and t_f are the ions that cross the sheath boundary during the fall time, the maximum sheath thickness, the total number of ions implanted during the fall time, and the fall time, respectively. Figure 4-6 shows the percentage of ions implanting with the full peak voltage for the entire pulse cycle for a 3 μs pulse on-time. This calculation includes the matrix sheath implantation described in the previous section. For a 1 kV pulse with a 3 μs on time, the maximum allowable fall time, as determined by Figure 4-5, is approximately 3 μs. Therefore all fall times less than this have identical P_{max}'s as indicated by a single curve in Figure 4-6. With a 4 μs fall time, implant voltages less than 3 kV are above their maximum allowable fall time, and therefore their P_{low}'s increase dramatically. With 10 μs fall time, all voltages below 12 kV are above their maximum fall time.

Figure 4-6 demonstrates how critical it is for fall time to be shorter than that determined by Figure 4-5. Longer fall times corrupt the implant energy significantly, with the fall time component dominating the total energy spread. For example, with 1 kV implant, increasing the fall time from 3 μs to 6 μs reduces P_{max} from over 70% to nearly 40%, a significant increase in the energy spread.

As enumerated above, PIII pulsed implants have a significant low energy contribution, reaching 80% or higher for long fall times. One possible solution is to increase the

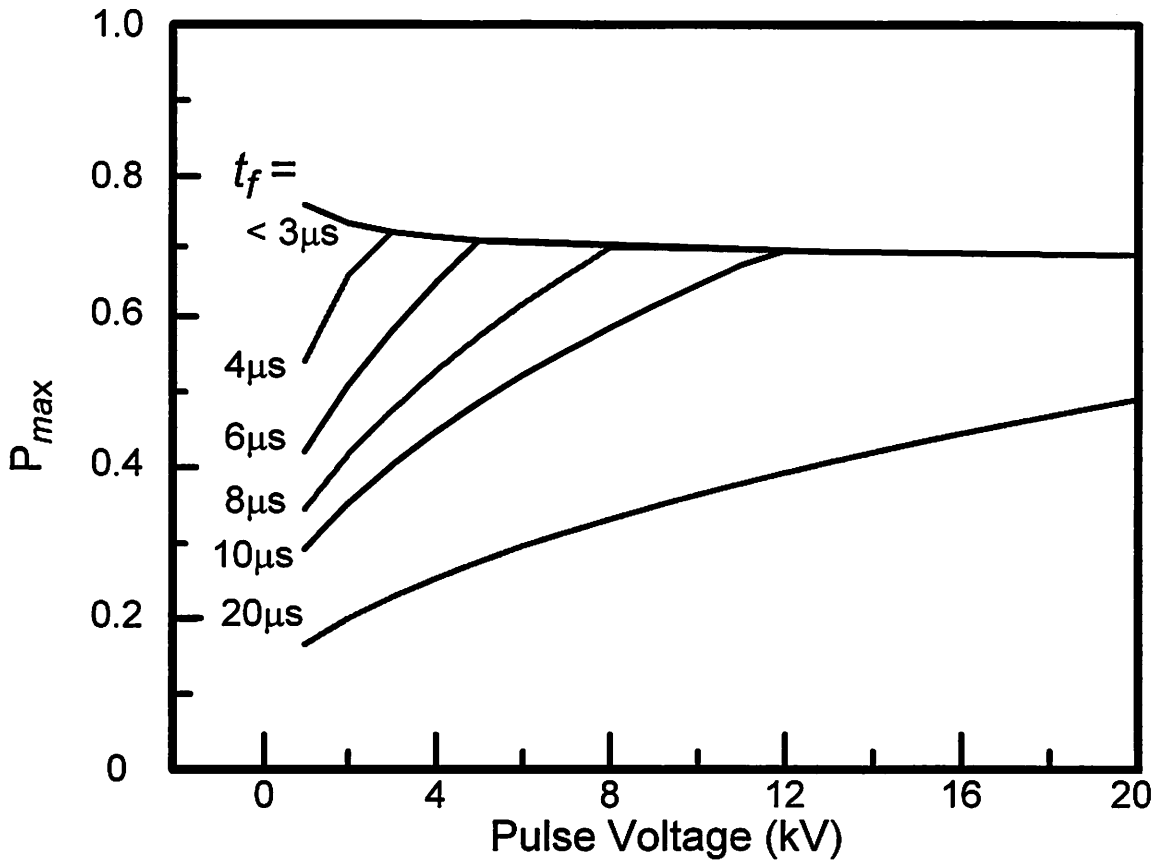

Figure 4-6 Low Energy Implantation for Entire Pulse Cycle

Percentage of full energy ions (V_{pulse}) for a complete pulse cycle for a 3μs pulse width, as a function of the fall time. All the t_f's below 3μs yield the same 1 - P_{low}, since the minimum sheath collapse time is greater than 3μs (Figure 4-5). Long fall times quickly degrade the mono-energetic quality of the implant. The ion density is $10^{10}cm^{-3}$.

ion density, which minimizes the matrix sheath contribution to energy spread. For example, an ion density of $10^{11}cm^{-3}$ instead of $10^{10}cm^{-3}$ reduces P_{low} by an impressive 20%. One caveat is that with higher ion densities, the fall time required to minimize the fall time low energy contribution is reduced. This arises, since higher ion densities yield thinner s_{max}'s, which in turn lead to faster sheath collapse times. As stated before, the fall time ought to be faster than the sheath collapse time. Optimizing the implant with respect to ion density, requires knowledge of the shortest achievable fall time and the maximum

allowable charge per pulse (as determined by gate oxide charging and cooling capacity considerations, see Chapter 5).

4.5 Implant Energy Distribution

The previous sections discussed the origins of low energy components in PIII implants, and calculated the total percentage of the implant with energies less than the full peak energy of the applied bias. It is also useful to predict the actual implant energy distribution, and subsequently the implant profile. The main caveat with distribution estimation, is that the errors in the assumptions are generally magnified, resulting in mostly qualitative distributions.

There are three dominant methods for estimating the implant energy distribution. First, and probably most accurate, is the method using Particle in Cell or similar type simulators. They solve Poisson's equation and track each ion as it traverses the sheath. These simulators make few assumptions, and therefore the results are fairly accurate. However, they suffer form long execute times, and generally provide little insight into the physical mechanisms of energy spread or the scaling of the low energy components with different plasma and implant conditions.

The second method of energy distribution estimation is to simply assume that the ion transit time across the sheath is zero. With this assumption, the implant energy is equal to the ion charge times the instantaneous applied voltage. This method provides fast distribution predictions, but suffers from an underestimation of the low energy component. For fast rise times, the actual ion transit time is longer than the rise time (Figure 4-7). This contradiction to the assumption results in an underestimation of the low energy component. A similar problem occurs with fast fall times. At the onset of the fall time, the sheath is quite wide, and the ion transit time can approach or exceed 500 ns. Because this method assumes a zero ion transit time, it does not account for the ions in the sheath at the end of the hold time. Therefore, this method will significantly underestimate the low

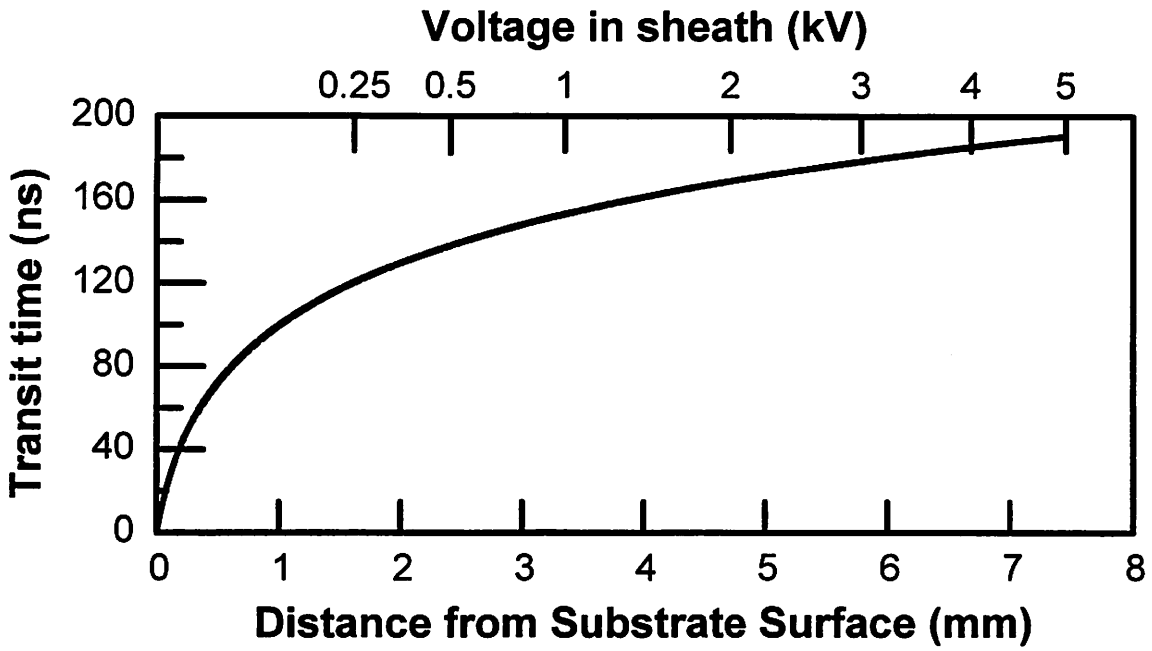

Figure 4-7 Ion Transit time for Matrix Sheath

The ion transit time for a 5 kV pulse with an ion density of $10^{10}/cm^3$. Note that the majority of the ions have a transit time above 100 ns. This contradicts the assumption of a zero transit time. The calculation assumes a frozen electric field during ion transit.

energy component for fall times on the order of the ion transit times. This technique will accurately model distributions for slow rising and slow falling voltage pulses, but this contrasts with the goal of sharp pulses.

A third method for energy distribution prediction relies on analytical equations. The obvious advantage is the physical insight afforded by analytical equations, and the simple extraction of scaling. Stewart et. al. [4-8] have attempted an analytical solution to the problem of distribution prediction. In this paper, they do assume that the ion transit time is zero, which makes the approach inaccurate for fast rise and fall times, common in current pulsing systems. By applying some of the concepts of the previous sections, a more accurate energy distribution prediction for realistic implantations with fast rise and fall times is possible. The distribution consists of 3 parts: the ions uncovered by the matrix sheath, the ions that diffuse across the sheath boundary and uncovered by the expanding Quasi-Static Child Law sheath, and the ions in the sheath at the commencement

of the fall time. The ions in the second part all implant with the maximum possible energy, with the low energy part of the implant consisting of the first and third contributions.

4.5.1 Matrix Sheath Contribution

The voltage profile in the matrix sheath is:

$$V(x) = \frac{qn_i x^2}{2\varepsilon_o} \tag{4-13}$$

where $V(x)$ is the voltage at a distance x away from the wall edge. This applies for all rise times, but it ignores ions that cross the sheath boundary during slower rise times. For practical rise times, these ions may be ignored. For example, with a 5 kV pulse, 100 ns rise time, and $10^{10}/cm^3$ ion density, the number of ions crossing the sheath during the rise time ($3x10^8/cm^2$) is less than 10% of the ions that are implanted from the matrix calculation using Equation (4-12) ($7.4x10^9/cm^2$). If the rise times become excessive, then the ions that cross the sheath should be taken into account. This may be done by assuming the transit time is near zero (which is a good assumption for long rise times) [4-8].

4.5.2 Fall Time Contribution

The second contribution to the low energy distribution are those ions that implant during the fall time. No ions cross the sheath boundary if the fall time is short. Only those ions in the sheath at the start of the fall time will implant. Previous energy distribution predictions ignored the effect of the ions in the sheath at the onset of the fall time. This significantly undercounted the low energy component of the implant for fast fall times. For fast fall times, we can assume a zero-field while these "fall time ions" implant. In other words, once the fall time begins, the ions cease acceleration, and move towards the wafer due to inertia of motion.

For a quasi-static Child Law sheath, the transit time, assuming a frozen electric field is:

$$t_c = 3s_{max}\sqrt{\frac{M}{2V_{pulse}}} \tag{4-14}$$

where t_c is the transit time across the quasi-static Child-Law Sheath. This equation underestimates the actual transit time, since it assumes a frozen electric field. With the zero field assumption, the ion transit time is:

$$t_{c2} = \frac{x}{\mathrm{v}} \tag{4-15}$$

$$\mathrm{v} = \sqrt{\frac{2V(x)}{M}} \tag{4-16}$$

$$V(x) = V_{pulse}\left(\frac{s_{max} - x}{s_{max}}\right)^{4/3} \tag{4-17}$$

where v is the ion velocity, $V(x)$ is the voltage in the sheath at position x in the sheath at the onset of the fall time, and t_{c2} is the ion transit time assuming a zero fall time. As long as t_{c2} is less than the fall time, the contribution of the "fall time ions" to the implant energy distribution can be calculated using Equation (4-9) and Equation (4-10).

4.5.3 New Implant Energy Distribution Calculation

Using the analysis in the previous two sub-sections, the implant energy distribution shown in Figure 4-8 is calculated for a 3 μs hold time, 5 kV implant pulse, with a 100 ns rise time and 1 μs fall time. The y-axis shows the distribution function split into 100 eV energy bins. Note that the fall time is faster than the ion sheath collapse time. The dashed line is the distribution prediction using the zero transit time assumption. The solid line is the prediction using the new method, which explicitly accounts for the matrix implantation. The zero transit time method predicts that 80% of the ions implant with the peak energy (V_{pulse}), while the new model estimates only 70%, a 10% difference If the

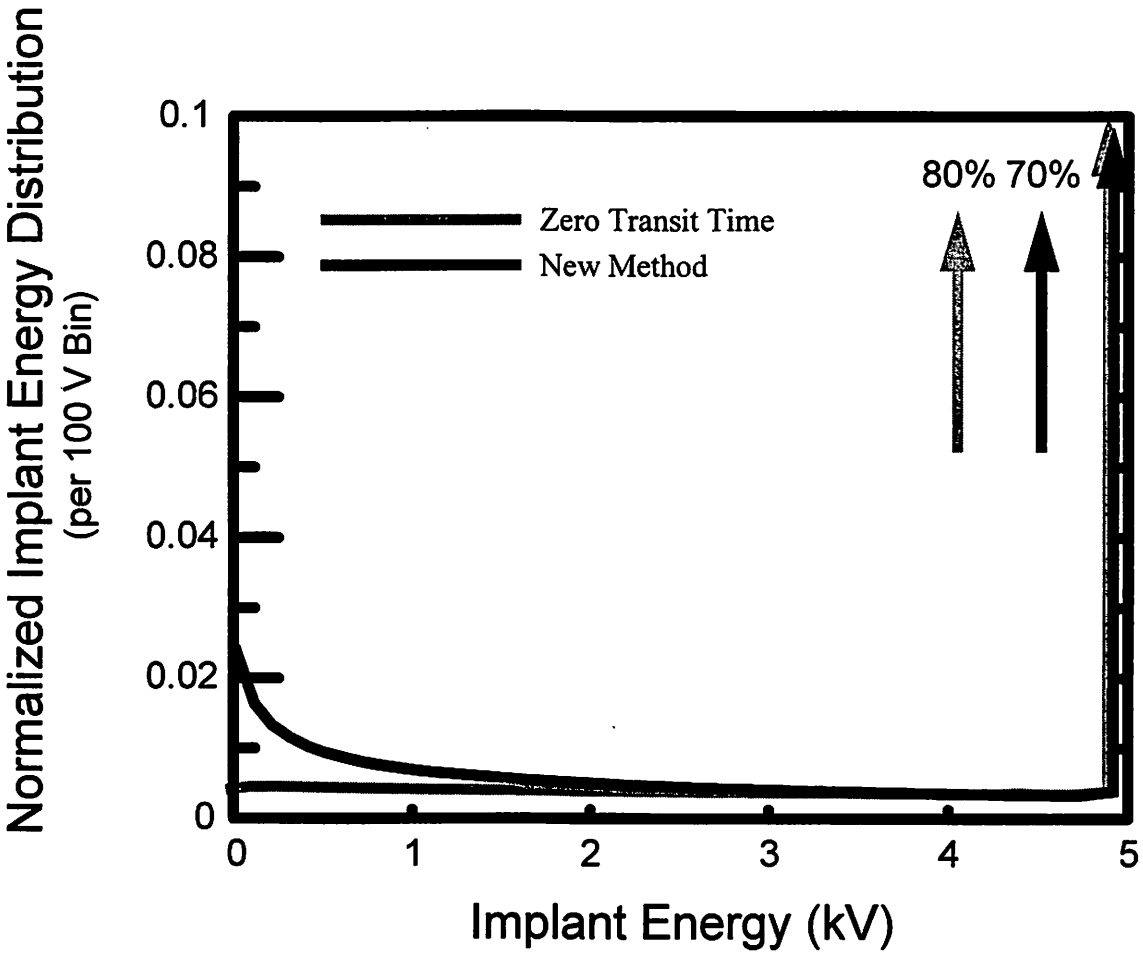

Figure 4-8 Estimated Implant Energy Distribution.

Estimated implant energy distributions for a 3μs, 5 kV pulse, with a 100 ns rise time and a 1μs fall time. The y-axis is the probability distribution, binned into 100V energy intervals. The solid line is the energy distribution accounting for the matrix implantation and the fall time implantation. The shaded line assumes a zero ion transit time with a 100 ns rise time. The zero-transit time predicts that 80% of the ions implant with the full energy, a full 10% higher than the other method. Reducing the rise time to 50 ns increases the overestimation to 15%. The zero transit time method severely underestimates the low energy component below 1 kV, which mostly results from the matrix implantation.

rise time is reduced to 50 ns, the difference between the two calculations increases to 15%. By examining the distribution, it is clear that the zero-transit time method misses many of the extreme low energy ions, which is a direct result of the zero-transit time assumption.

From these results, it is apparent that the zero-transit time method is only for pulse biases with long rise times.

Overall, for fast rise times (our pulser rise time is about 50 ns), it is imperative to consider the matrix contribution to the low energy implant. For a 50 ns rise time, neglecting the matrix contribution causes an overestimation of the high energy component of the implant by 10%. For fast fall times, it is important to consider the ions that are in the sheath at the onset of the fast fall time. Neglecting to do so will result in an overestimation of the high energy implant component by around 5%. (Note the percentages are a function of the implant time, and could change considerably for much longer or shorter pulse widths). This new analytical model is still less accurate than the full PIC simulators, especially since it still assumes a frozen electric field during sheath expansion. A non-frozen electric field would result in ions implanting at energies near the peak, but not at the peak, with the energy reduction directly dependent on the sheath expansion rate.

4.6 Conclusions

The poly-energetic nature of the PIII implant requires special attention. Implantation of ions due to the formation of the matrix sheath, and implantation of ions in the sheath at the onset of the fall time are intrinsic sources of energy spread, and must be considered. Extended rise or fall times create additional energy spread. A near zero rise time and a fall time below a critical value (a function of the maximum sheath thickness), eliminate these sources of energy spread.

Accounting for all the sources of energy spread and understanding the limitations and the scaling trends with the implant variables, allows the identification of an implant condition that yields an acceptable amount energy spread.

References

[4-1] M. A. Lieberman, and A. J. Lichtenberg, "Principles of Plasma Discharges and Materials Processing." John Wiley and Sons, Inc., New York, p. 531, 1994.

[4-2] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *Journal of Vacuum Science and Technology B*, vol. 12, no. 2, pp. 833-837, 1994.

[4-3] W. En, M.A. Lieberman, and N. W. Cheung, "Comparison of Experimental Target Currents with Analytical Model Results for Plasma Immersion Ion Implantation," *IEEE Transactions on Plasma Science*, vol. 23, no. 3, pp. 415-421, 1995.

[4-4] B. Szapiro and J. J. Rocca, "Electron Emission from Glow-discharge Cathode Materials due to Neon and Argon Ion Bombardment," *Journal of Applied Physics*, vol. 65, no. 9, pp. 3713-3716, 1989.

[4-5] W. En and N. W. Cheung, "A Novel Method for Determining the Secondary Electron Yield Dependence on Ion Energy for Plasma Exposed Surfaces," I*EEE Trans. on Plasma Science*, vol. 24, no. 3, pp. 1184-1187, June, 1996.

[4-6] Emmert, G.A., "Model for Expanding Sheaths and Surface Charging at Dielectric Surfaces during Plasma Source Ion Implantation", *Journal of Vacuum Science and Technology B*, vol. 12, pp. 880-883, 1994.

[4-7] W. En and N. W. Cheung, "Modeling of Charging Effects in Plasma Immersion Ion Implantation", *Nucl. Instr. and Meth. in Phys. Res.* B96, pp. 435-439, March, 1995.

[4-8] R. A. Stewart and M. A. Lieberman, "Model of Plasma Immersion Ion Implantation for Voltage Pulses with Finite Rise and Fall Times," *Journal of Applied Physics*, vol. 70, no. 7, pp. 3481- 3487, 1991.

5 PIII with Dielectric Substrates

5.1 Introduction

One of the most promising applications for PIII is the processing of thin film transistors (TFT) on glass or plastic substrates. The large substrates associated with TFT's (for flat panel display applications) experience extremely low throughput with the currently available conventional implant systems. Since PIII's dose rate is independent of wafer area, PIII has a tremendous advantage in TFT processing.

When implanting into dielectric substrates the issue of substrate charging must be understood and controlled. The study of the effects of substrate charging, and the optimization of the PIII process for dielectric substrates is the topic of this chapter.

5.2 Substrate Charging

Substrate charging occurs in PIII with all dielectric substrates. The substrates may either be glass or plastic substrates for TFT's or in silicon-on-insulator wafers. For TFT processing there a number of different implantation steps, such as source/drain, poly gate, hydrogenation, and poly grain size control. All of the implants require high doses with implant energies ranging from 20-100 keV. Before the commercialization of PIII for these technologies, it is imperative to understand how substrate charging affects the energy profile of these implants and the implant dose rate.

With pulsed PIII, the substrate charges by two distinct mechanisms. First, a portion of the applied bias (V_o) couples to the substrate instead of the plasma, reducing the effective peak implant voltage. Then, during the implant, positive charge deposition increases the voltage drop across the substrate at the cost of the sheath voltage. For example, a typical 10 keV implant deposits ~$3x10^{-4}$ coulombs/cm^2. The charge generates a 4 kV drop across a 0.5 mm glass substrate, reducing the implant voltage by the same amount.

5.3 Sheath Voltage During Implantation

Using the PIII model for dielectric substrates (Figure 5-1), the three stages of PIII, the rise, hold, and fall time, are re-evaluated in terms of energy profile and dose rate. Before pulsing but after plasma exposure, the surface charges to the plasma floating potential. The sheath width, a function of ion density and electron temperature, is small, usually than 1 mm for typical processing conditions

5.3.1 Rise Time

During the rise time, voltage builds-up across the substrate by two mechanisms. The applied voltage pulse capacitively couples to the sheath and to the insulating substrate. The substrate voltage subtracts from the sheath voltage, reducing the coupling efficiency to the plasma. The second mechanism is charge deposition. Initially, the sheath expands rapidly producing a large plasma ion current. The ions implanting during the rise time implant with less than the full pulse potential. The implanting ions also eject secondary electrons, amplifying the positive charge deposition. For high implant voltages, the secondary electrons dominate the surface charge deposition with a yield (γ) of 1-20 [5-1]. Previous work [5-2] has ignored the secondary electrons, vastly underestimating the surface charge. The surface charging drops a fraction of the applied voltage across the substrate, with a corresponding reduction in the sheath voltage.

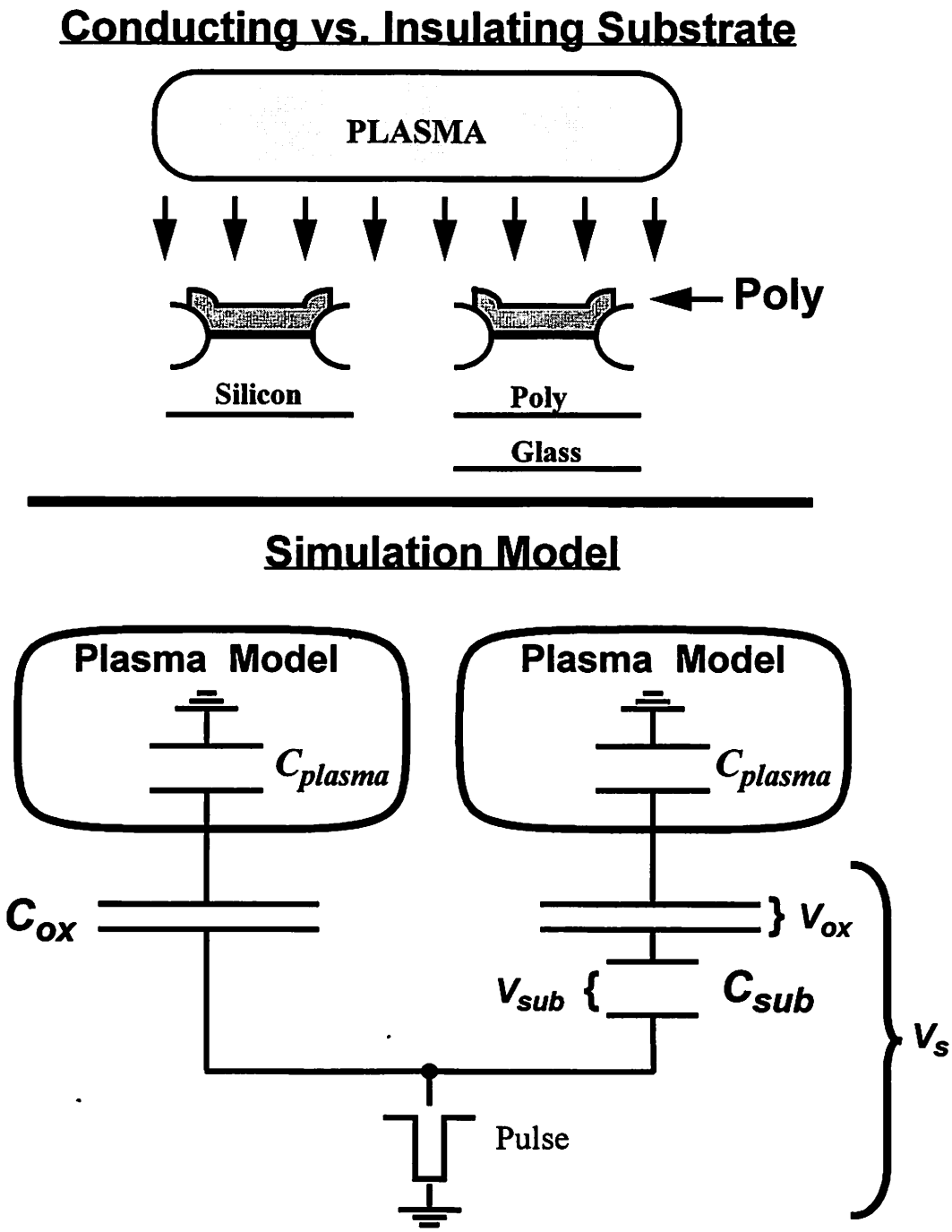

Figure 5-1 PIII Model for Dielectric Substrates

The simulation model for dielectric substrates. During implantation, the applied bias is capacitively coupled across the substrate and plasma, and therefore the capacitance of the plasma is explicitly shown. C_{sub}, C_{ox}, and C_{plasma} are the capacitances of the substrate, gate oxide, and plasma, respectively, while V_{ox}, V_{sub}, and V_s are the voltage drops across the gate oxide, the substrate, and the surface voltage, respectively. In this simulation, the dielectric substrates are assumed have zero leakage current.

For the thin substrates of interest (0.5 - 2 mm), the reduction in sheath voltage is mostly due to charge deposition rather than capacitive coupling losses. For the 0.5 mm substrate case considered in Figure 5-2, the maximum sheath voltage is 15 kV. This translates to a coupling efficiency ($\eta = \frac{V_{max}}{V_{pulse}}$) of 75%, only 8% of the total 25% reduction is attributable to capacitive coupling losses. Here, V_{max} is the maximum sheath voltage, while V_{pulse} is the maximum applied bias voltage.

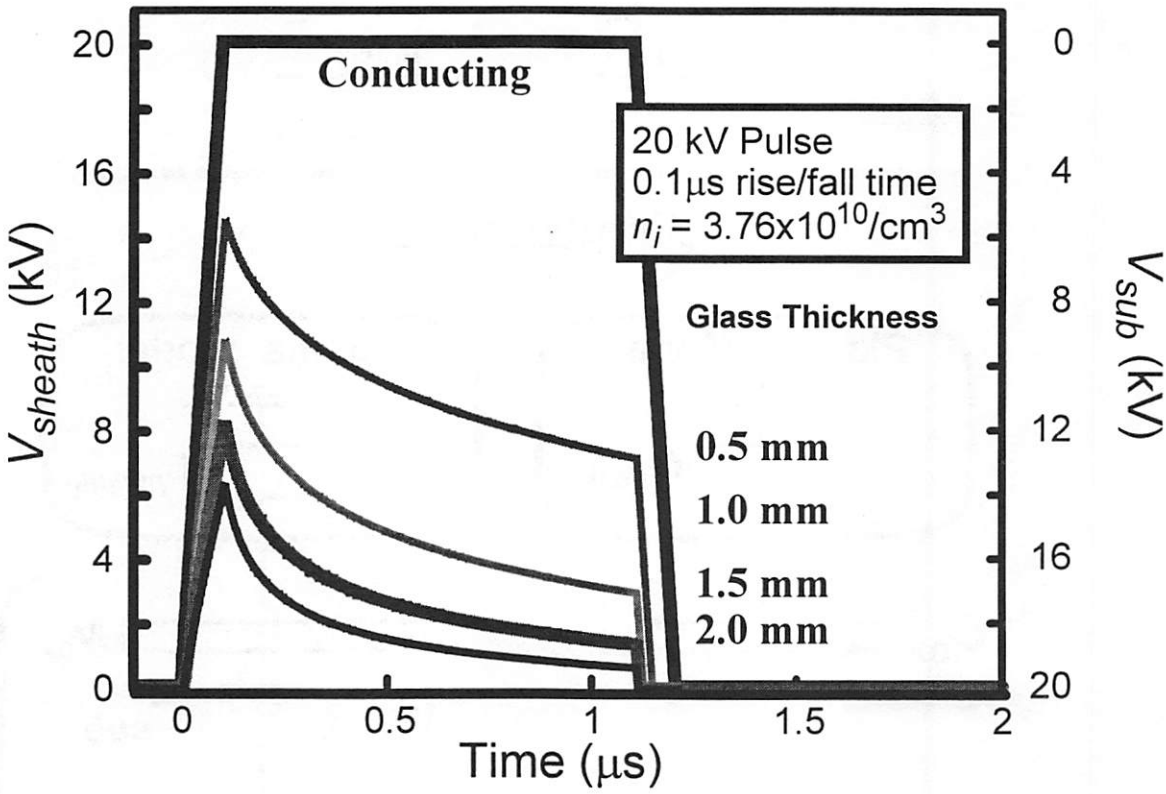

Figure 5-2 Sheath Evolution with Glass Substrates

V_{sheath} during implantation with varying substrate thicknesses for a 20 kV, 1μs pulse. During the rise time, the voltage pulse capacitively couples to the plasma and glass substrate. Implanted ions and ejected secondary electrons deposit positive charge on the wafer surface, degrading the sheath potential. For thick substrates or long pulse widths, the build-up of positive charge strongly attenuates the sheath voltage.

5.3.2 Hold and Fall Time

During the hold time, the sheath expands more slowly, decreasing J_i. Charge deposition continues, further degrading the sheath voltage and, consequently, the implant energy. The surface charge accumulation becomes so severe that the sheath voltage can be extinguished, defined as a sheath voltage of only 10% of the V_o, after only a couple of microseconds.

In the simulation, the falling edge of the pulse capacitively couples to the sheath, actually causing the sheath potential to become negative, an extremely non-equilibrium situation. To resolve this, the simulator limits the electron current to the electron saturation value (Equation (3-8)). This large J_e quickly neutralizes the surface charge. The initial equilibrium restores after the sheath fully collapses, which takes many microseconds for kilovolt pulses (Figure 4-5).

5.3.3 Dielectric Thickness

As shown in Figure 5-2, the sheath voltage evolution varies with the glass thickness. Thicker substrates worsen the capacitively coupling of V_o to the sheath and accelerate the sheath voltage degradation from charge accumulation. Combined, both of these effects reduce η and increase the voltage spread, δ (defined as the peak implant energy - implant energy at end of t_{on}).

5.4 Implant Energy Distribution

One of the interesting characteristics of pulsed bias PIII is the poly-energetic implant energy. Even with a conducting substrate there is a significant spread in the energy distribution of the implanted ions (Chapter 4). The large surface voltage buildup with a insulating substrate further disperses the energy distribution. Figure 5-3 depicts the implant energy distributions for each of the three stages of PIII. To reduce simulation noise, the implant dose is integrated over 400 V intervals, called bins.

During the rise time, the implant energy begins at 0 volts and ramps up to the maximum implant energy. After the initial current spike, J_i decreases, causing a slightly negative slope in *Dose/Bin* for the rise time. During the hold time, charge deposition reduces the implant energy. Because $\Delta V_{sheath}/\Delta t$ slows (Figure 5-3), the Dose/Bin increases with time (decreasing energy). During the fall time, J_i is small, yielding a comparatively small dose for the falling edge of the pulse, as shown in Figure 5-3. In this example, the fall time does not contribute much to implant, since the fall time in the simulation is much less than the critical fall time from Figure 4-5.

The energy spread is a sensitive function of the capacitance of the substrate. Doubling the substrate thickness nearly doubles the energy spread during the hold time. The value for δ is directly dependent on the total deposited positive charge. Therefore,

$$\delta \propto Dose/Pulse \cdot (1 + \gamma(V)) \tag{5-1}$$

where $\gamma(V)$ is the secondary electron yield as a function of ion impinging energy.

5.4.1 Applied Voltage Effect

Optimizing the implant voltage and achieving an acceptable level of energy spread requires a full understanding of the scaling of δ and η, with *Dose/Pulse*. Maximizing *Dose/Pulse* maximizes throughput, but at the cost of energy spread (Equation (5-1)). The *Dose/Pulse* for PIII is approximately:

$$Dose/Pulse \approx q n_i u_b t_w + q n_i s_{max} \tag{5-2}$$

where t_w and s_{max} are the pulse width and maximum sheath width, respectively. The first term represents the ions that cross the sheath boundary, while the second term corresponds to the uncovering of ions from the expanding sheath. The second term dominates for most implant conditions of interest with dielectric substrates.

Assuming a steady state Child Law current relationship, $s_{max} \propto V_o^{3/4}$ (Equation (3-4)). Therefore, the *Dose/Pulse* from Equation (5-2) increases sub-linearly with V_o. Fig-

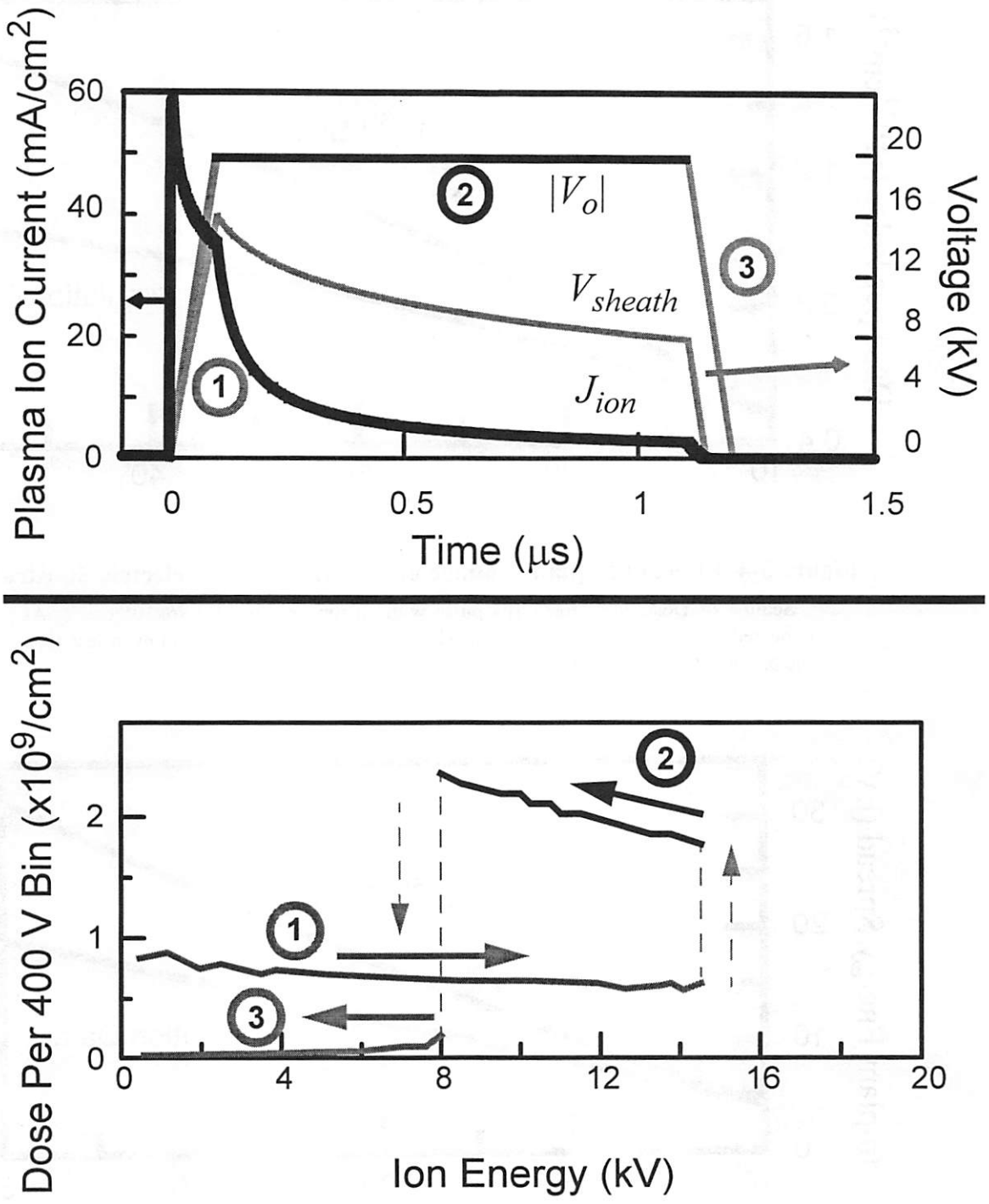

Figure 5-3 The Three Stages of PIII

(a) The time evolution of the applied voltage, plasma ion current and sheath voltage for a 0.5 mm glass substrate and a 20 kV, 1μs pulse. (b) The ion energy distribution for the same pulse, separated into the (1) rise time, (2) hold time, and (3) fall time.

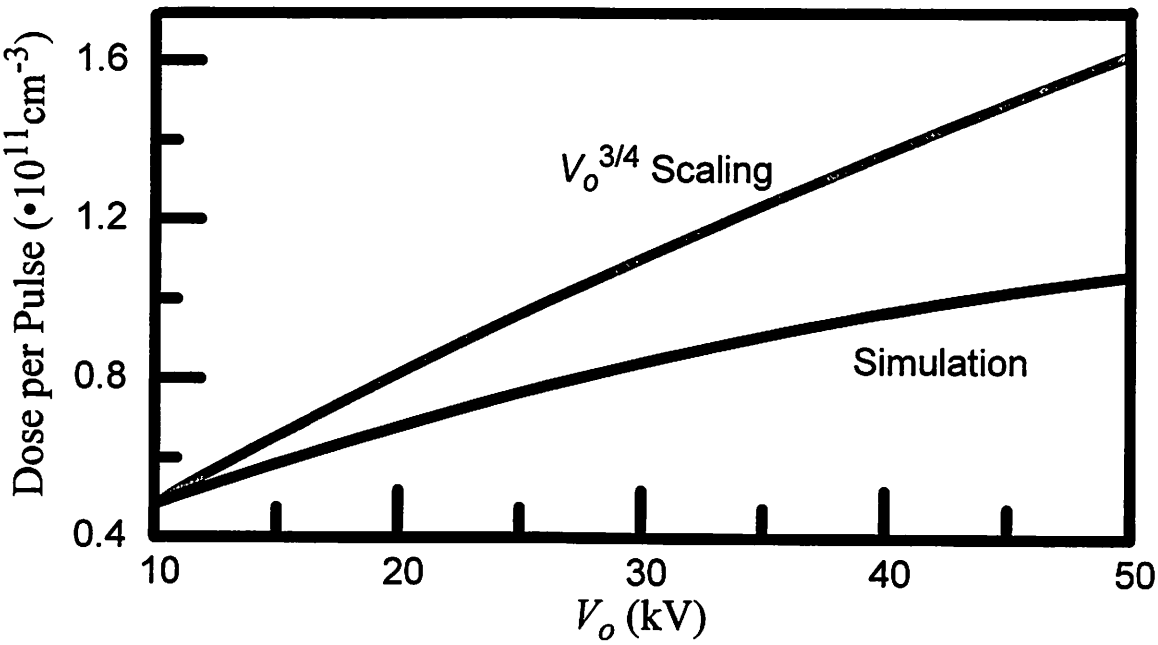

Figure 5-4 Effect of Implant Voltage on Dose/Pulse for Dielectric Substrates

Scaling of Dose/Pulse for a 1μs pulse with an ion density of 3.76×10^{10}cm^{-3}. As expected, the Dose/Pulse increases sub-linearly with implant bias, and even less than the predicted $V_o^{3/4}$ scaling.

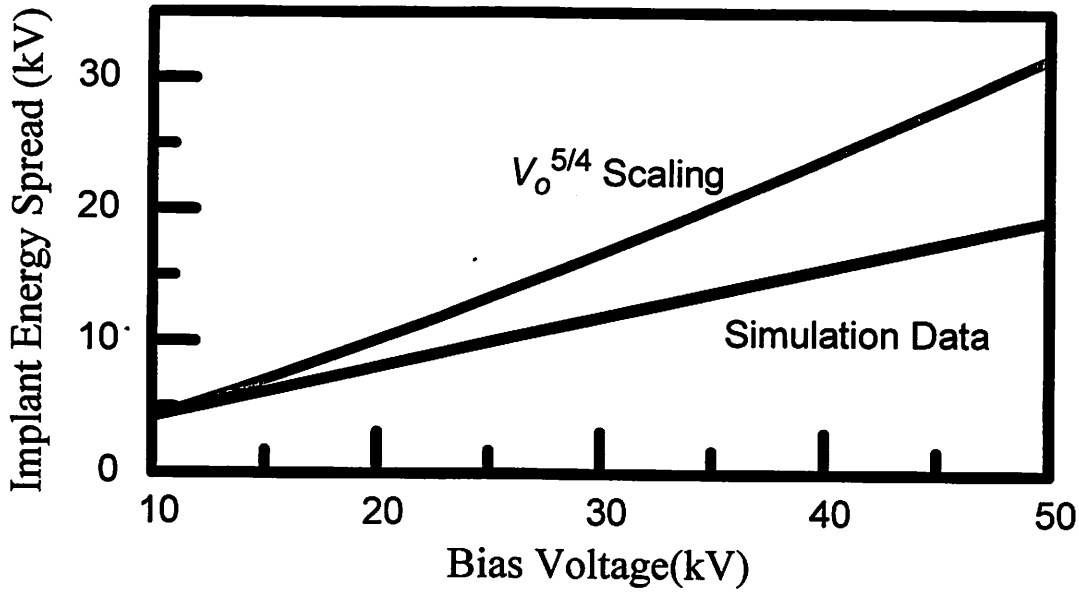

Figure 5-5 Scaling of Implant Energy Spread with Bias Voltage

The actual scaling of the implant energy spread from simulations is found to be slightly less than $V_o^{5/4}$.

ure 5-4 compares simulation with the predicted scaling. The *Dose/Pulse* indeed increases sub-linearly with applied bias; even less than the simple $V_o^{3/4}$ model. This is attributable to the sheath expansion rate not being proportional to the final Child Law sheath width.

To determine the scaling of δ with applied voltage, the scaling of $s_{max} \propto V_o^{3/4}$, and the scaling of secondary electron yield with voltage ($\gamma \propto V_o^{1/2}$), must be combined in Equation (5-1). This results in δ being proportional to $V_o^{5/4}$. The actual scaling is slightly less than $V_o^{5/4}$ (Figure 5-5). Since the implant voltage is not constant but degrades during the pulse, the secondary electron yield scales slower than assumed above. This reduces the increase of δ with V_o. Overall, increasing V_o boosts the *Dose/Pulse* while widening the voltage spread.

5.4.2 Ion Density Effect

The plasma ion density is another controllable parameter. Similar to altering V_o, changing n_i also affects the *Dose/Pulse* and δ. Increasing n_i raises the *Dose/Pulse* and decreases the processing time. But the higher dose, in turn, widens the energy spread. Equation (5-2) suggests that the dose rises linearly with n_i, but the full plasma model shows a sub-linear dependence on n_i (Figure 5-6). s_{max} decreasing with higher n_i, (the steady-state Child Law sheath scales as $s_{max} \propto n_i^{-1/2}$), accounts for most of this difference. In addition, the larger J_i reduces V_{sheath} during the implant, retarding the sheath expansion. Along these same lines, the increased current and the thinner sheath combine to reduce the coupling efficiency (Figure 5-7), lower the mean implant energy, and widen δ.

5.4.3 Pulsing Frequency and Pulse Width

The pulsing frequency (f_p) and pulse width (t_w) are two easily controllable variables in PIII. Both of these need to be optimized for maximum throughput and minimum substrate charging. The time-averaged ion impinging rate is independent of the wafer bias and is equal to:

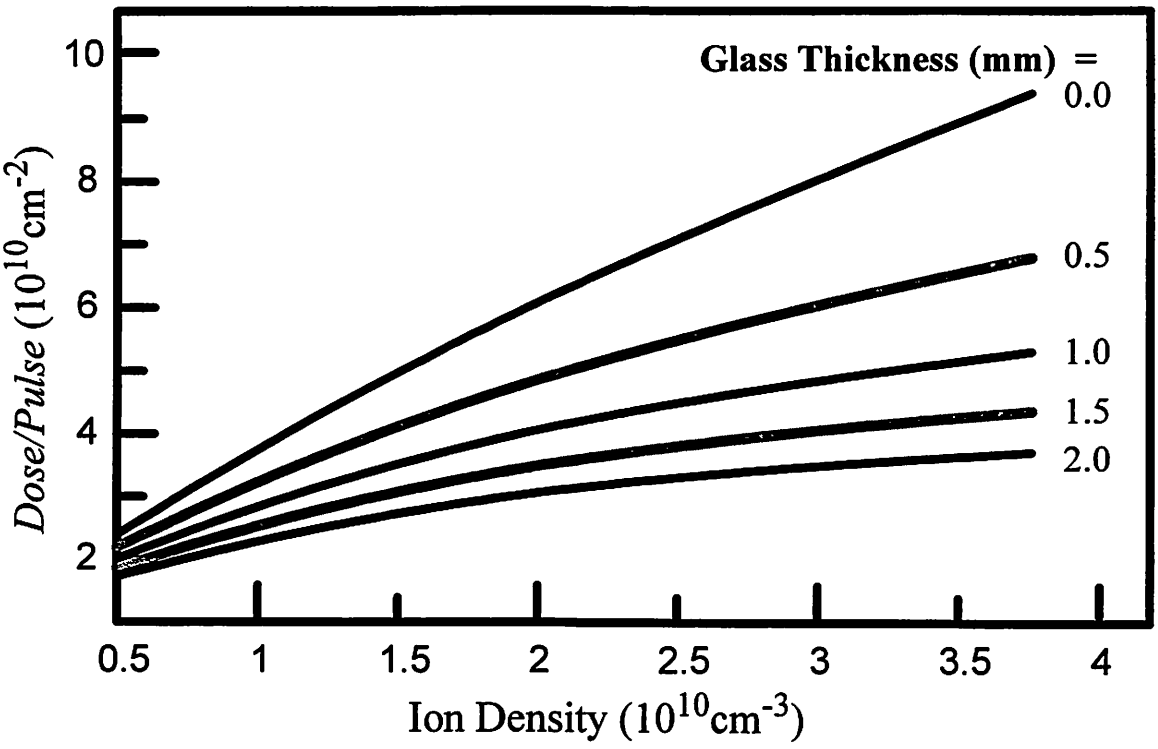

Figure 5-6 Scaling of Dose/Pulse with Ion Density

Dose Per Pulse scales sub-linearly with ion density. This arises from secondary effects, such as reduced bias coupling to the plasma, and increased charge build-up in the substrate. Simulation is for a 20 kV, 1μs pulse.

$$Maximum\ Dose\ Rate = q \cdot n_i \cdot u_b \tag{5-3}$$

With DC implantation, all of the impinging ions implant with the full energy, yielding the theoretical maximum throughput. With pulsed operation, the impinging rate is the same as DC operation, but a significant portion of the ions hit the wafer when the bias pulse is off, and therefore bombard the wafer with a minimal energy (~20 eV).

When the bias is pulsed on, the sheath expands, increasing the dose rate above this time-averaged value. When the pulse is turned off, the sheath collapses, and the dose rate temporarily goes below the time averaged value. The time averaged current, though, is always the same value, $q \cdot n_i \cdot u_b$. To maximize throughput, one wants as much of the

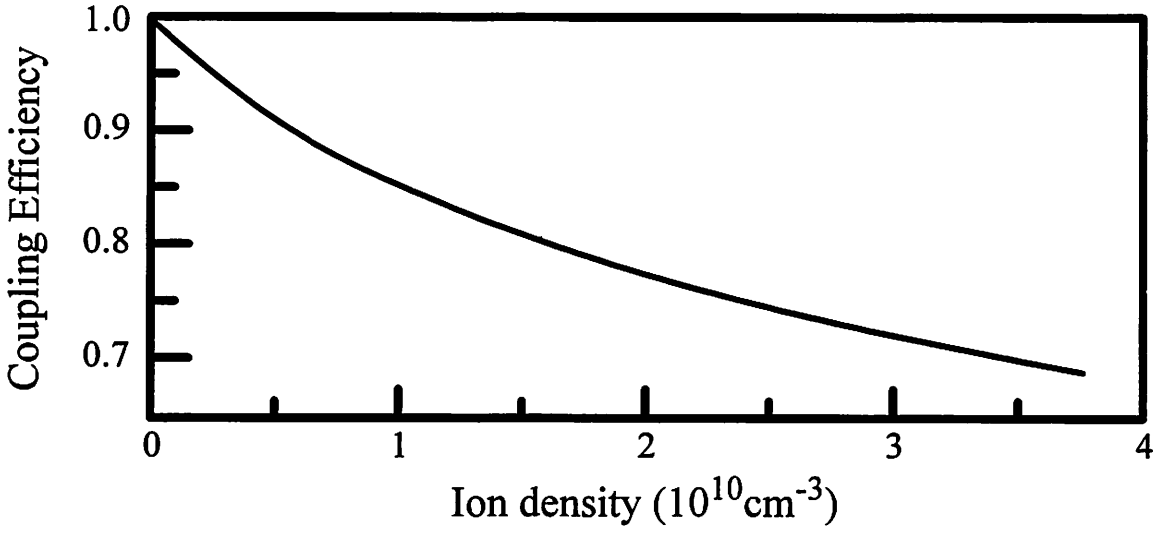

Figure 5-7 Scaling of Coupling Efficiency with Ion Density

The combination of thinner sheaths and increased charge deposition during the rise time leads to a dramatic reduction in coupling efficiency with higher plasma ion densities. The simulation is for a 20 kV, 1μs pulse with 0.1μs rise times, and a 0.5 mm glass substrate.

impinging ion flux to implant with higher energies, rather than hitting the surface at low voltages during the off time.

To maximize throughput, t_w is limited by the self-extinguishing time. If the pulse is on too long, the charge deposited by the plasma ions and ejected secondary electrons will completely counterbalance V_o, reducing the sheath voltage close to zero. Any time that the pulse is held on after self-extinguishment is wasted, since the ion impinging energy is so low. For the implant conditions of interest, this occurs in the 0.5μs to 10μs range, dependent on the ion density, substrate capacitance, and bias voltage.

The counterpart to t_w is f_p, the pulsing frequency. The period between pulses should be longer than the surface charge neutralization time. Therefore, the limitation for the off time is that it be longer than sheath collapse time, $t_{off} \geq s_{max}/u_b$. This ranges from less than 1μs to more than 50μs for the ion implant conditions of interest (Figure 4-5)

For example, with a 20 kV pulse, 0.5 mm thick glass substrate, and an n_i of $3.76 \times 10^{10} cm^{-3}$, the pulse extinguishes itself after 6.12 μs. The sheath fully collapses 3.15 μs after fall time commences. With a 2 μs fall time and a 1.15 μs off-time, the pulse frequency will be 109 kHz. This gives a 100% efficiency (defined as $\frac{Pulsed\ Implant\ Dose\ Rate}{DC\ Implant\ Dose\ Rate}$). If the maximum pulse frequency is 25 kHz (as is the case with our pulser), the off time would be 31.88 μs, yielding an implant efficiency of 23%. By optimizing t_w and f_p, it is possible to have an implantation current close to theoretical maximum DC rate dictated by Equation (5-3). If energy spread is more important parameter than implant efficiency, the pulse should be less than the extinguishing time.

5.5 Conclusions

Implanting into dielectric substrates introduces two new issues: bias coupling to the substrate and charge build-up. The coupling losses reduce the maximum implant energy, while charge build-up, a sum of the secondary electrons and the implanted ions, diminish the implant energy during the on-time. Optimizing the pulse width and pulse frequency allows the dose rate to remain high while controlling substrate charging.

References

[5-1] B. Szapiro and J. J. Rocca, "Electron Emission from Glow-discharge Cathode Materials due to Neon and Argon Ion Bombardment," *Journal of Applied Physics*, vol. 65, no. 9, pp. 3713-3716, 1989.

[5-2] Emmert, G.A., "Model for Expanding Sheaths and Surface Charging at Dielectric Surfaces during Plasma Source Ion Implantation", *Journal of Vacuum. Science and Technology B*, vol. 12, pp. 880-883, 1994.

6 Multiple Species PIII Model

6.1 Introduction

Models of PIII that accurately predict the implant dose and energy profile are needed for commercial use. Most of the previous work on PIII models concentrated on the implantation of single ion species plasmas, such as Argon. Most plasmas for implantation applications, however, contain a multiple ion species. Each ion species has its own mass, implant depth, and secondary electron yield, complicating the model. For proper dose, profile, and gate oxide charging prediction it is imperative to know the ratio of the implanted ions.

Chapter 3 derived the single species PIII model. The single species model may be extended with the derivation of an effective mass and effective Bohm velocity for the multiple positive ion species plasma. The model considers only positive ion species, ignoring negative ion species which won't implant under negative applied biases.

6.2 Multiple Species PIII Model

The crux of the multiple species model is the calculation of an effective mass and effective Bohm velocity for the plasma (Figure 6-1). The single species model calculates the sheath thickness and ion current as a function of time with these two effective quantities. The multiple species module solves for the individual ion currents and the secondary electron current.

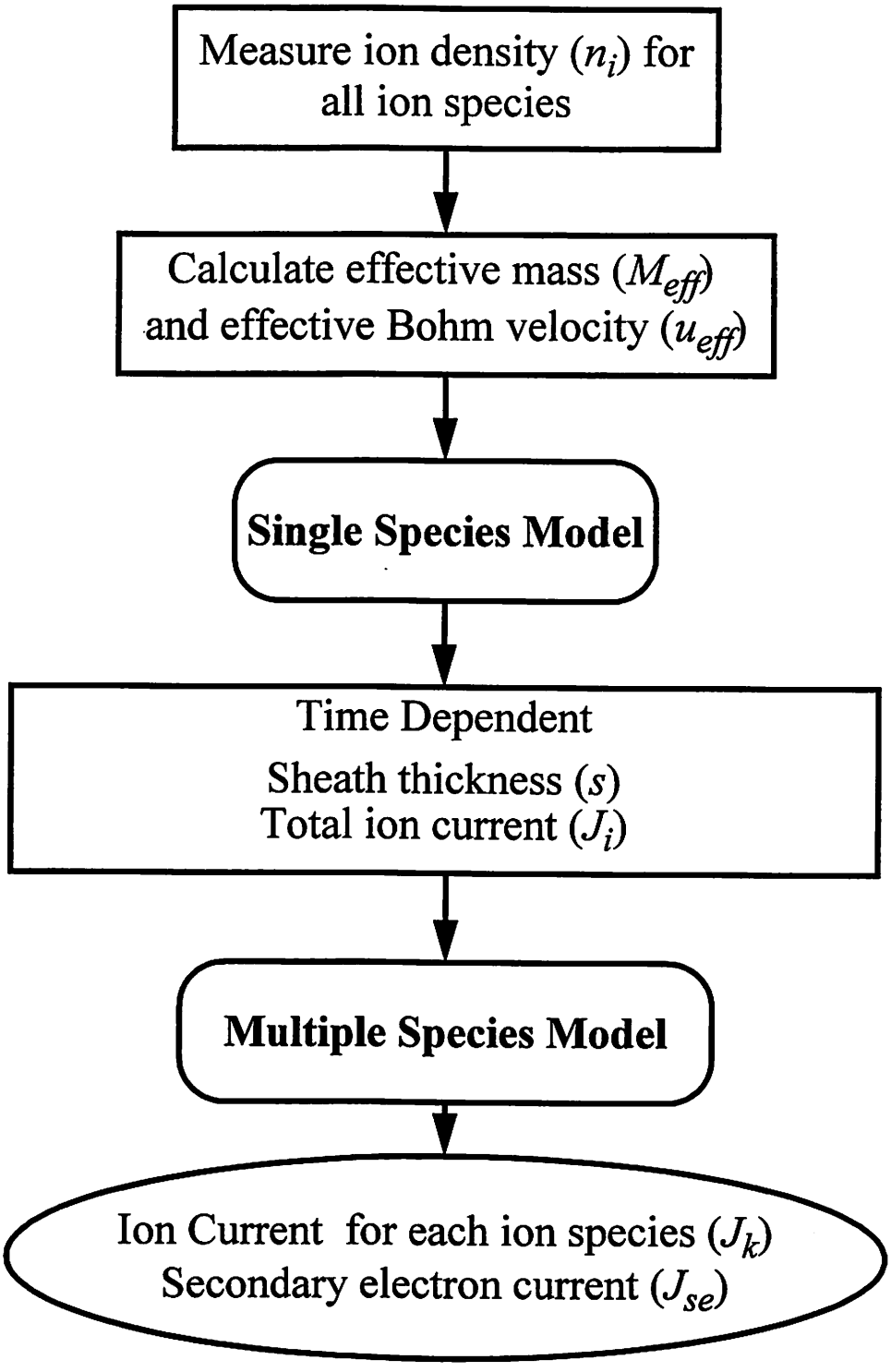

Figure 6-1 Flow Chart for PIII Multiple Ion Species Modeling

Calculating an effective mass and effective Bohm velocity for a multiple ion species plasma, the single species model may be utilized. The single species model calculates the sheath thickness and total ion current as a function of time, while the multiple species model, separates the ion current into its individual ion components.

This approach assumes that only one plasma front exists. This is valid during the rise and hold time for pulsed PIII, and for DC PIII. The single plasma front assumption breaks down during sheath collapse, or if the sheath does not fully collapse in between each pulse. Therefore, the model is not accurate in predicting currents for extremely fast pulsing PIII or during long fall times. Both of these situations are undesirable, and therefore this model applies for most practical PIII implants.

Extending the single species PIII model to accommodate multiple positive ion species requires re-deriving the ion current equations (equations 3-2 - 3-5) with multiple masses and multiple Bohm velocities. However, the basic framework of the derivation remains unchanged [6-4].

For each ion species the bulk plasma ion density is defined as:

$$n_k = \alpha_k n_o \tag{6-1}$$

where α_k is the ion density fraction of species k. Energy and flux conservation for the ions yields

$$\frac{1}{2} M_k u_{bk}^2(x) = -q\phi(x) \tag{6-2}$$

$$q n_k(x) u_{bk}(x) = j_k \tag{6-3}$$

where ϕ, and x are the voltage in the sheath, and distance from the sheath edge, respectively. Solving for $n_k(x)$ in Equation (6-2) and Equation (6-3) yields

$$n_k(x) = \frac{j_k}{q}\left(-\frac{2q\phi}{M_k}\right)^{-1/2} \tag{6-4}$$

Using Equation 6-4, the total ion density $n_o(x)$ is

$$n_o(x) = \frac{(-2q\phi)^{-1/2}}{q} \sum_{k=1}^{m} j_k M_k^{1/2} \tag{6-5}$$

where m is the number of positive ion species present in the plasma. Using Equation 6-5 in Poisson's equation results in

$$\frac{d^2\phi}{dx^2} = \frac{(-2q\phi)^{-1/2}}{-\varepsilon_o} \sum_{k=1}^{m} j_k M_k^{1/2} \tag{6-6}$$

Multiplying Equation (6-6) by $d\phi/dx$ and integrating from 0 to x, with the boundary conditions of $d\phi/dx = -E = 0$ at the plasma-sheath edge $\phi = 0$ ($x = 0$), yields:

$$\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 = 2\frac{(2q)^{-1/2}(-\phi)^{1/2}}{\varepsilon_o} \sum_{k=1}^{m} j_k M_k^{1/2} \tag{6-7}$$

Taking the (negative) square root, since $d\phi/dx$ is negative, and integrating once more, gives

$$\frac{4}{3}(-\phi)^{3/4} = 2\frac{(2q)^{-1/4}}{\sqrt{\varepsilon_o}}\left(\sqrt{\sum_{k=1}^{m} j_k M_k^{1/2}}\right)x \tag{6-8}$$

Employing the single plasma front assumption, and setting $\phi = -V_s$ at $x = s$ yields

$$\sum_{k=1}^{m} j_k M_k^{1/2} = \frac{4}{9}\varepsilon_o\sqrt{2q}\frac{V_s^{3/2}}{s^2} \tag{6-9}$$

Using this equation alone is not sufficient to solve for the sheath thickness as a function of time. The ion current density due to the k^{th} positive ion can also be described as the flux of ions crossing the sheath edge

$$j_k = qn_k\left(u_{bk} + \frac{ds}{dt}\right) \tag{6-10}$$

Inserting Equation (6-10) into Equation (6-9) and rearranging for the sheath thickness results in

$$\frac{1}{3}\sum_{k=1}^{m} qn_k M_k^{1/2}\frac{ds^3}{dt} + s^2\sum_{k=1}^{m} qn_k M_k^{1/2}u_{bk} = \frac{4}{9}\varepsilon_o\sqrt{2q}V_s^{3/2} \tag{6-11}$$

The sheath thickness can be solved, since s is the only unknown in Equation (6-11). Equation (6-10) and s determine the ion flux of each component and the total ion current.

Comparing Equation (6-11) with the single ion species equations allows an effective mass (M_{eff}) and an effective Bohm velocity (u_{eff}) to be extracted.

$$\sum_{k=1}^{m} n_k M_k^{1/2} = n_o\sqrt{M_{eff}} \tag{6-12}$$

$$\sum_{k=1}^{m} n_k M_k^{1/2}u_{bk} = n_o u_{eff}\sqrt{M_{eff}} \tag{6-13}$$

Using $n_k = \alpha_k n_o$ and solving for M_{eff} and u_{eff} gives

$$M_{eff} = \left(\sum_{k=1}^{m}\sqrt{\alpha_k \cdot M_k}\right)^2 \tag{6-14}$$

$$u_{eff} = \frac{\sum_{k=1}^{m}\alpha_k \cdot u_{bk} \cdot \sqrt{M_k}}{\sqrt{M_{eff}}} \tag{6-15}$$

Qin *et. al.* have performed a similar derivation [6-6], but did not include the Bohm velocity in their ion current equation (Equation (6-10)). The ion current during PIII consists of two parts, ions that are uncovered by the expanding sheath (n_{expand}), and ions that diffuse across the sheath boundary (n_{diff}). Figure 6-2 compares the contribution of n_{diff} to the total ion current for different implantation pulse widths. For pulse times less than 1μs, n_{expand} dominates the ion current, and n_{diff} contributes less than 10%. The contribution of n_{diff} rises for longer t_w, with n_{diff} accounting for 50% of the total ion current for 10μs

pulses. From this simulation, it is imperative to include n_{diff} in the multiple species model for pulse widths greater than 1μs. En [6-8] experimentally shows that both terms in Equation (6-10) are necessary for accurate modeling of PIII. Neglecting the second portion of the current severely underestimates ion current for longer pulse widths, and is completely inaccurate for DC applications. As will be shown in the next section, inclusion of both ion current terms significantly affects the estimates of the implant ion ratios.

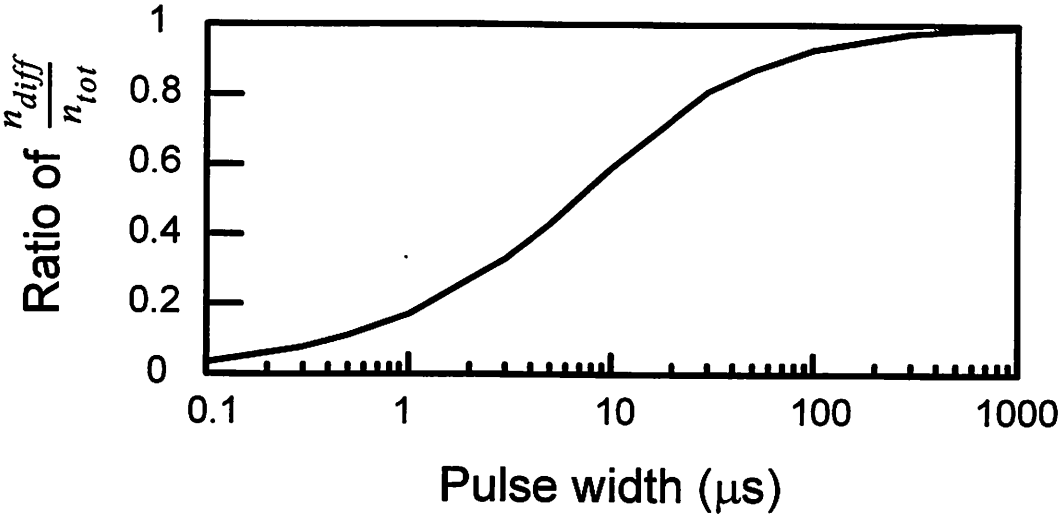

Figure 6-2 Contribution of Ions Diffusing across the Sheath Boundary

The ratio of the contribution of ions diffusing across the sheath (n_{diff}) to the total ion current (n_{tot}). For short pulse widths, the ions uncovered by the expanding sheath dominant the current, while for longer pulses the diffusing ions dominant the current. For pulses greater than ~1μs, the diffusing ions must be included in the model.

6.3 Simulation Results

The computation of the ion current for each species is possible after determining M_{eff} and u_{eff}. To calculate the implant ratios, the ion densities must first be known. Next, the effective mass and effective velocity are computed. Then the differential equation for sheath thickness as a function of time is solved. Finally, the individual ion currents are calculated.

The model results show that the ion implant ratios are a function of the pulse length. By analyzing the terms of the ion current in equation Equation (6-10), three different regimes of implantation as a function of pulse width are identifiable. For short pulses the expanding sheath term dominates ($j_k = qn_k\frac{ds}{dt}$). Therefore, the ion implant ratios will be equal to the ion density ratios in the plasma bulk. For longer implant times (and DC implants) the ions diffusing across the sheath boundary dominate ($j_k = qn_k u_b$) the total ion current. In this regime the implanted ion ratios are proportional to the product of the ion density and Bohm velocity. The latter depends on the square root of the ion mass. In between the two extremes is the mixed regime, where both terms contribute to the ion current.

Two different types of implants will be analyzed in more detail, the single molecular gas source plasma (e.g. BF_3), and the carrier gas plasma (e.g. PH_3/He).

6.3.1 Single Molecular Gas Source Plasma PIII

BF_3 gas is a commonly used for silicon p-type doping. A BF_3 plasma contains a number of different of boron containing ion species. For this simulation it is assumed only two ions exist, B^+ and BF_2^+. The ion density ratio for these two ion species are set at 10% and 90%, respectively, with a total ion density of $10^{10}/cm^3$. Since each ion has a different mass, knowing the exact ratios of implanted ions is imperative for calculating the depth profile of boron in the wafer. A 5 kV pulse is simulated with varying hold times, with a constant 1μs fall time. The implant ratio of $\frac{B^+}{BF_2^+}$ as a function of pulse width is shown in Figure 6-3. The ratio changes from ~10% for short pulse widths to ~20% for long pulse widths. The three regimes are easily identifiable. The short pulse regime, where the implant ratio is proportional to the ion density ratio, occurs for pulses less than 1μs. The mixed regime is for pulses from 1μs to ~300 μs, and the long pulse regime, where the implant ratio is inversely proportional to the square root of the mass, occurs for pulse widths greater than 300 μs. As Figure 6-3 demonstrates, most practical pulsed PIII

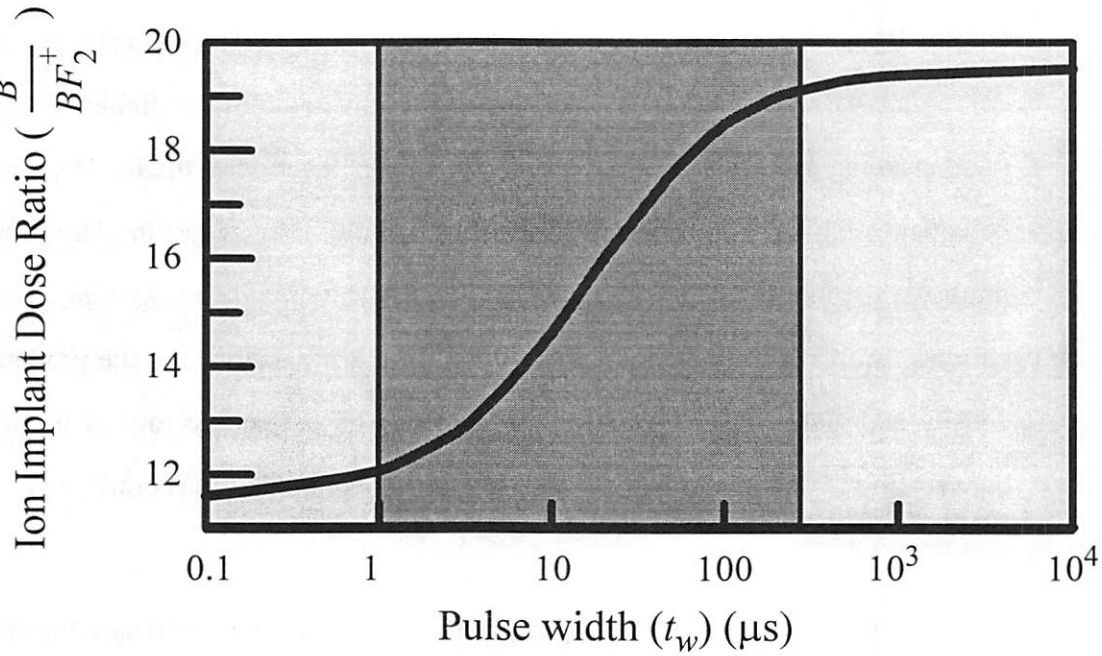

Figure 6-3 The Implant Ratio of B^+ to BF_2^+.

The ion density ratios are 10% and 90%, respectively, with a total ion density of 10^{10}/ cm^3. A 5 kV wafer bias is simulated with varying pulse widths. For short pulses, the implant ratios are proportional to the ion density ratios. For long pulses the implant ratio is modified by the mass ratio. Therefore, longer pulse widths contain a higher percentage of the lighter ion, B^+.

implants operate in the mixed regime, where both ion current terms contribute, reinforcing the need for including both terms in the model.

6.3.2 Carrier Gas Plasma

A carrier gas is often employed to dilute dangerous gases such as Phosphine. Both the carrier gas and dopant species ions are implanted. Accurate dose prediction from wafer current measurements requires full knowledge of the exact amount of carrier gas ions implanted.

A typical example is n-type doping implantation with a plasma of PH_3 diluted with helium. For this simulation only two ions are considered, PH_3^+ and He^+. The ion density ratios are set at 90% and 10% respectively, and the total ion density is $10^{10}/cm^3$. A 5 kV pulse is simulated with varying hold times, with a constant 1μs fall time. The implant

ratio of $\frac{PH_3^+}{He^+}$ is plotted in Figure 6-4. The ratio changes from ~90% with short pulses to 76% for long pulses. As in the BF_3 example, the mixed regime begins at ~1 μs, requiring consideration of both ion current terms expressed in Equation 6-10 for accurate modeling for practical implant conditions.

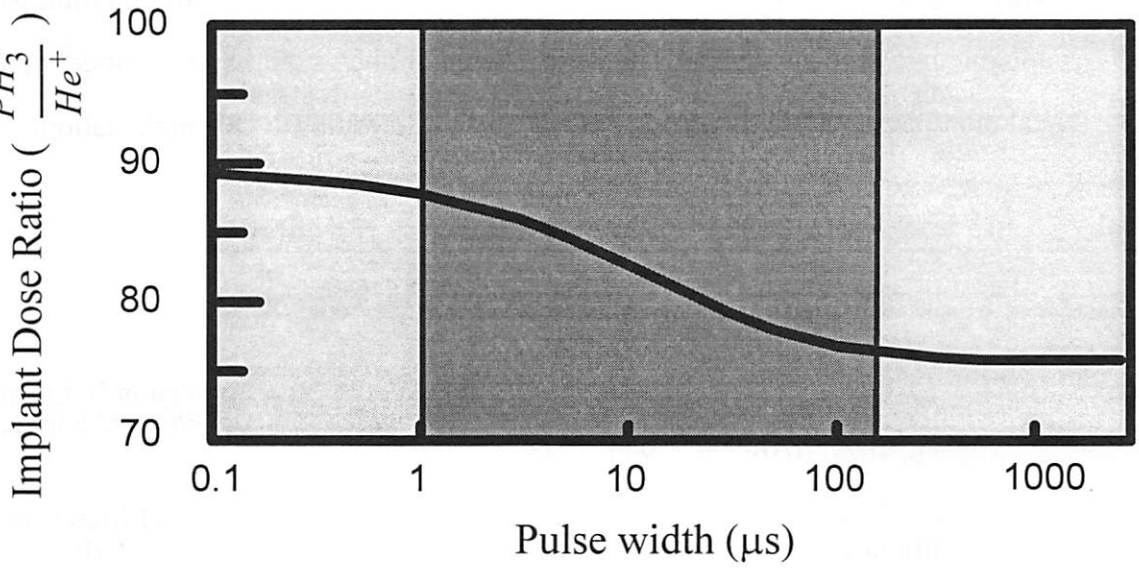

Figure 6-4 The Implant Ratio of PH_3^+ to He^+.

The ion density ratios are 90% and 10%, respectively, with a total ion density of 10^{10}/ cm^3. A 5 kV wafer bias is simulated with varying pulse widths. For short pulses, the implant ratios are proportional to the ion density ratios. For long pulses the implant ratio is affected by the ion mass ratio. Therefore, longer pulse widths contain a higher percentage of the lighter ion, Helium.

6.4 Conclusions

Almost all plasma sources for PIII contain multiple ion species. Through the use of an effective mass and an effective Bohm velocity, the single species model has been modified to account for multiple ion species.

The novel component of this model is the inclusion of the Bohm velocity in the ion current equation. This expands the range of validity of the model from short pulses to infinite pulses (DC implantation). A unique result of this model is that the ion implant ratios

are a function of the pulse width. For short pulses, the ratios are equal to the ion density ratios. For longer pulses (and DC implants), the implant ratios are also proportional to the Bohm velocity of the ions (and hence the inversely square root of the ion mass). Therefore, the implant ratios for the lighter ions are boosted at long pulse widths. For proper calibration, dosimetry, and depth profile prediction, it is imperative to understand that changing the pulse width will alter the implant ratios. Inclusion of this multiple species module into the Coupled PIII model in Chapter 3 allows the accurate modeling of practical multiple species implantations from short pulse widths to DC implantation.

References

[6-1] E. C. Jones, B. P. Linder, and N. W. Cheung, "Plasma Immersion Ion Implantation for Electronic Materials," *Japanese Journal of Applied Physics*, Part 1, vol. 35, pp. 1027-1036, Feb. 1996.

[6-2] J. K. Chester, "Effects of Expanding Sheaths on the Operation of Mercury-Arc Rectifiers", *Journal of Science and Technology*, vol. 37, no. 1, pp. 2-10, 1970.

[6-3] M. Shamim, J. T. Scheuer and J. R. Conrad, "Measurements of Spatial and Temporal Sheath Evolution for Spherical and Cylindrical Geometries in Plasma Source Ion Implantation", *Journal of Applied. Physics*, vol. 69, pp. 2904-2908, 1991.

[6-4] M. A. Lieberman, "Model of Plasma Immersion Ion Implantation," *Journal of Applied Physics*, vol. 66, no. 7, pp. 2926-2929, 1989.

[6-5] R. A. Stewart and M. A. Lieberman, "Model of Plasma Immersion Ion Implantation for Voltage Pulses with Finite Rise and Fall Times," *Journal of Applied Physics*, vol. 70, no. 7, pp. 3481- 3487, 1991.

[6-6] S. Qin, Z. Jin, and C. Chan, "Dynamic Sheath Model of Collisionless Multi-species Plasma Immersion Ion Implantation," *Journal of Applied Physics*, vol 78, no. 1, pp. 55-60, July 1995.

[6-7] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *Journal of Vacuum Science and Technology B*, vol. 12, no. 2, pp. 833-837, 1994.

[6-8] W. G. En, M. A. Lieberman, and N. W. Cheung, "Comparison of Experimental Target Currents with Analytical Model Results for Plasma Immersion Ion Implantation," *IEEE Transactions on Plasma Science*, vol. 23, no. 3, pp. 415-421, June 1995.

7 Collisional PIII

7.1 Introduction

The correlation between applied bias, implant dose rate, and implant energy distribution in PIII must be well understood. Chapter 4 formulated a 1st order model for calculating these characteristics of a PIII implant, for a mono-species, collisionless sheath, conducting substrate implant. Chapter 5 extended the basic model to dielectric substrates, while Chapter 6 modified the basic model for multiple species. This chapter adds another module to the PIII model, that is, the effects of ion-neutral collisions in the sheath.

Chester [7-1] first determined the flux of ions from a moving sheath region, while Scheur et al. [7-2] and Lieberman and Stewart [7-3, 7-4] extended the model to PIII. These first models all assumed a collisionless sheath. As will be shown in Section 7.3, most PIII implants occur in a moderately collisional regime, which requires modifications to the earlier PIII model for accuracy.

Davis [7-5] performed some of the earliest work with collisions, calculating ion impact energies on the cathode wall of a collisional DC glow discharge. Abril [7-6, 7-7] extended this work with formulations for the impact energies of fast neutrals. Tokonami [7-8] and Farouki [7-11] calculated both ion and neutral impinging energies for plasma etching applications. Qin [7-12] and Wang [7-13] extended the work to PIII, calculating ion energy distributions, while Vahedi [7-9, 7-10] computed ion impinging energies and angles for PIII conformal trench doping.

This chapter further analyzes the issue of ion-neutral collisions in the sheath. First, the phase-space of PIII implants will be explored. The results indicate that most PIII implants, even those at "low" pressures, operate in the moderately collisional regime. This contrasts sharply with previous work which predominantly addressed the fully collisional regime with pressures around 50 mtorr and above. Through Monte Carlo simulation, three new issues will be addressed: neutral energies for PIII implants, the effect of collisions on dose rate, and the effect of collisions on measured target current.

7.2 Collision Process

The PIII implantation including collisions consists of five distinct steps. First, the ion enters the plasma sheath region. Next, the electric field created by the wafer bias accelerates the ion. The third step is the ion-neutral collision. Two collision processes dominate in a PIII sheath region, elastic and charge exchange. Elastic collisions generally only change the direction of the ion trajectory and not the energy. With charge exchange, the ion's positive charge is transferred to a nearby thermal neutral. This creates a fast neutral (the old ion), and a new thermal ion. In the fourth step, the sheath electric field accelerates the new thermal ion, while the fast neutral coasts towards the wafer. The fast neutral may undergo neutral-neutral collisions, which divides the energy between the two neutrals, creating two fast neutrals of lower energy. At very high pressures, the neutral may undergo so many collisions that the original ion energy is divided among so many neutrals, that each individual neutral energy is attenuated back to the thermal energy. In the final step, the ion and the fast neutral implant into the wafer. Both ions and neutrals contribute to the implant dose. Both of them also eject secondary electrons, which has implications on substrate charging and total implantation current.

Collisions in the sheath alter the PIII implant in three ways. First, the dose now consists of both ions and neutrals, with the total dopant flux increasing. The higher dopant flux affects the implant energy, with collisions broadening the implant energy distribution,

and lowering the mean implant energy. Finally, collisions alter the measured pulser current.

7.3 Definition Of Moderately Collisional Regime

It is generally assumed that ion-neutral collisions in the sheath become significant when the sheath thickness (s) is equal to or greater than the mean free path (λ_i). Disregarding collisions allows a much simplified and straight-forward solution to the transient and DC sheath analysis during PIII. However, if $s = \lambda_i$, approximately 63% = $1\text{-}e^{-s/\lambda}$ of the ions collide before reaching the target. A more conservative definition of when collisions should be considered is when 10% of the ions collide before arriving at the target. This requires a sheath thickness one order a magnitude thinner than the ion mean free path.

Figure 7-1 compares the sheath width and the ion mean free path. The 10% threshold pressure is plotted for argon as a function of ion density and implant voltage. An operating condition above the appropriate curve exhibits collisional characteristics. This figure considers only charge exchange scattering with cross-sections from [7-6, 7-14]. For example, a 10 kV implant at an ion density of $10^{10}/cm^3$ requires a pressure less than ~0.3 mtorr. Increasing the ion density decreases the sheath width, allowing for higher pressures. Higher implant voltages increase the sheath thickness, requiring lower pressures. For example, a 50 kV implant with an ion density of $10^9/cm^3$, requires a pressure below ~60 μtorr. Because of the different cross-sections, the family of minimum collisional pressure curves are different for other gases. For comparison, the cross section for N_2^+ charge exchange is 50% larger than argon's (10 keV ions), while the cross-section for H_2^+ is 50% less than argon's [7-14, 7-15]. This means that for the same implant voltage and ion density, N_2 must operate at lower pressure, while H_2 can operate at higher pressures, as compared to argon.

Implants at UC Berkeley typically operate slightly above the collisional limit, such as a 5 kV argon implant with a $10^{10}cm^{-3}$ ion density. Published PIII operating conditions reported by other research groups similarly operate in the slightly collisional regime [7-16, 7-17]. Therefore, these implants require a collisional analysis for accurate models.

7.4 Computational Analysis

There are two general schemes for computing the effect of collisions, analytical and Monte Carlo methods. Analytical techniques provide noiseless and fast solutions but at the cost of necessitating additional simplifying assumptions. For example, charge exchange is usually the only ion collision process considered. Neutral-neutral collisions are ignored, and the mean free path must be independent of energy. Some simple analytical solutions are:

$$J_n = \frac{s}{\lambda} \cdot J_i \tag{7-1}$$

$$C = 1 - e^{-\left(\frac{s}{\lambda_i}\right)} \tag{7-2}$$

where J_n is the neutral flux, J_i is the ion flux and C is the percentage of ions that collide in the sheath.

Monte Carlo methods allow for fewer assumptions. For this work, charge exchange is still the only collisional process considered. Elastic scattering is ignored, since it generally only effects the angle of implantation, not dose, energy, or current [7-10]. Since the cross-sections of neutral-neutral collisions are usually an order of magnitude less than the charge exchange cross-section, they are neglected [7-14]. The mean free paths are calculated with cross-sections that depend on the ion energy, computing the energy distribution and neutral flux with greater accuracy. This contrasts with analytical methods which must use a fixed cross-section. Figure 7-2 plots the cross-section for argon as a function of ion energy, illustrating the large variation in cross-section. For the analy-

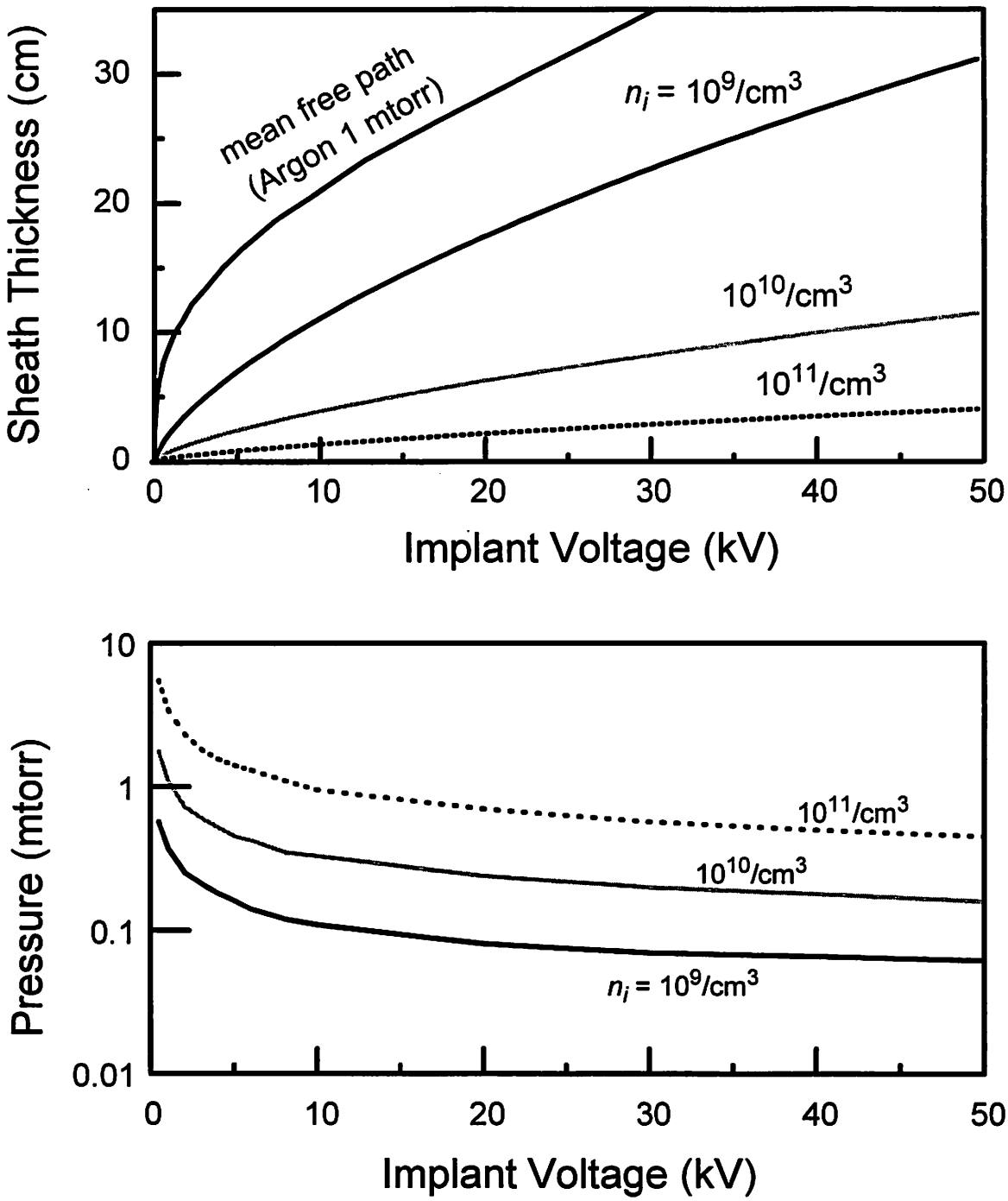

Figure 7-1 Minimum Pressure For Collisional PIII

a) The argon ion mean free path compared to the sheath thickness as a function of pressure and ion density. b) Using a) the pressure at which 10% of the ions collide before implanting into the target is calculated for argon is plotted. Implants with operating pressures above the curves exhibit collisional characteristics. Lower ion densities and higher implant voltages increase the sheath thickness, reducing the minimum pressure at which collisions become significant.

sis in this chapter, the sheath width is frozen, considering only the steady state. This enables the analysis of collisions in isolation, without the obfuscation of rise times, fall times, and matrix sheath implantation. Combining the output for many different sheath thicknesses yields results for pulse implantation.

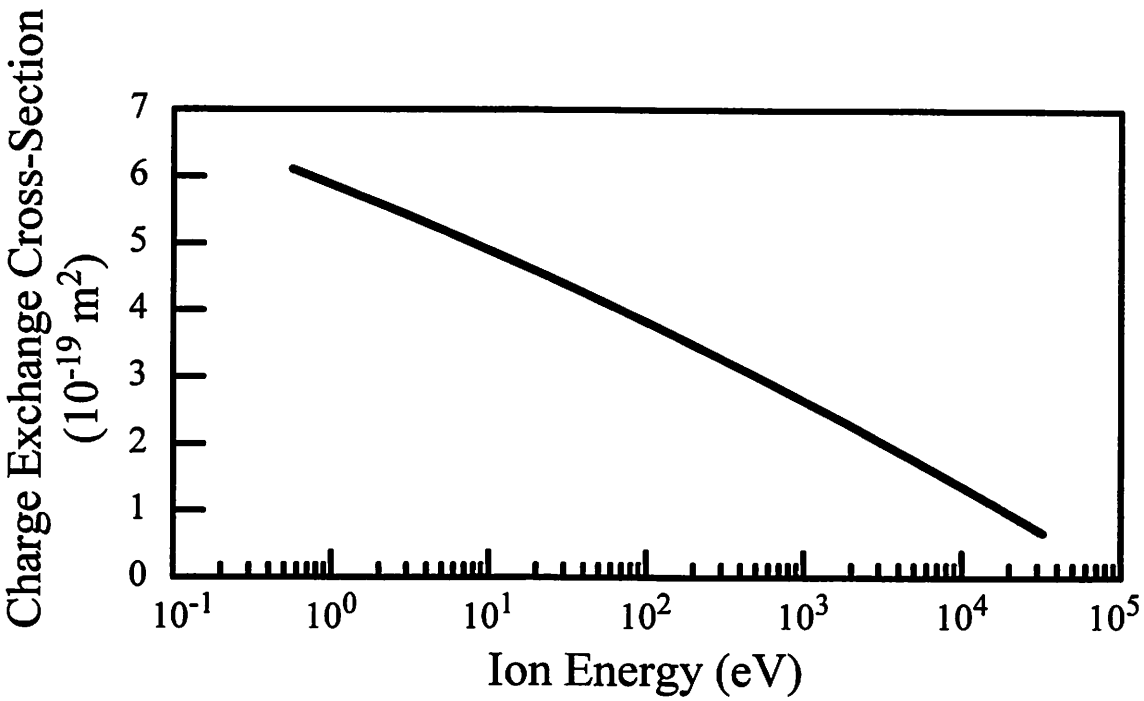

Figure 7-2 Argon Charge Exchange Cross Section

The cross section for ion-neutral charge exchange as a function of ion energy [7-14]. Higher ion energies (and velocities) reduce the interaction time between the ion and neutral, leading to a lower cross-section. Incorporating cross-sections as a function of energy is a critical component of the Monte Carlo analysis.

7.5 Simulation Approach

The effects of collisions are included in the PIII model by adding a module to the basic model of Chapter 3 (Figure 7-3). The only modification to the 1st order model is the use of a collisional sheath model instead of a collisionless one [7-18]. The modified current equation is

$$J_i = \frac{4\varepsilon_o}{9}\left(\frac{2q}{M}\right)^{1/2} \frac{V_s^{3/2}}{s^2[1+(12\pi/125)(s/\lambda_i)]^{1/2}} \tag{7-3}$$

where λ_i is the ion mean free path. The Bohm velocity is also modified by collisions

$$u_b = \sqrt{\frac{q \cdot T_e}{M}} \cdot \sqrt{\frac{1+\pi \cdot \lambda_{DE}}{2 \cdot \lambda_i}} \tag{7-4}$$

where λ_{DE} is the electron Debye length.

Equations 7-3 and 7-4 are valid from the collisionless to the highly collisional regimes.

The 1st order model calculates the sheath thickness, and ion current as a function of time. These outputs and the wafer bias are inputs for the Plasma Implantation Monte Carlo Analysis of Collisions module (PICMIC). The PICMIC module calculates the ion distribution, the neutral flux and distribution, and the total pulser current.

7.6 Simulation Results

Ion-neutral collisions in the sheath alter many attributes of the PIII implant. Collisions broaden the energy distribution of the implant, introducing lower energy ions. Charge-exchange collisions also produce fast neutrals, which also implant into the wafer. These neutrals have the energy of the colliding ion. Once created, they coast to the target without interaction (neutral-neutral collisions are ignored since their cross-sections are usually a magnitude less than charge exchange cross-sections). The total dose rate is equal to sum of the ion flux and the neutral flux. Therefore, fast neutrals increase the total dopant flux. Since total flux energy is conserved, the average impinging energy of both ions and neutrals is

$$Average\ Implant\ Energy = \frac{q \cdot V_{implant} \cdot Ion\ Flux}{(Ion\ Flux + Neutral\ Flux)} \tag{7-5}$$

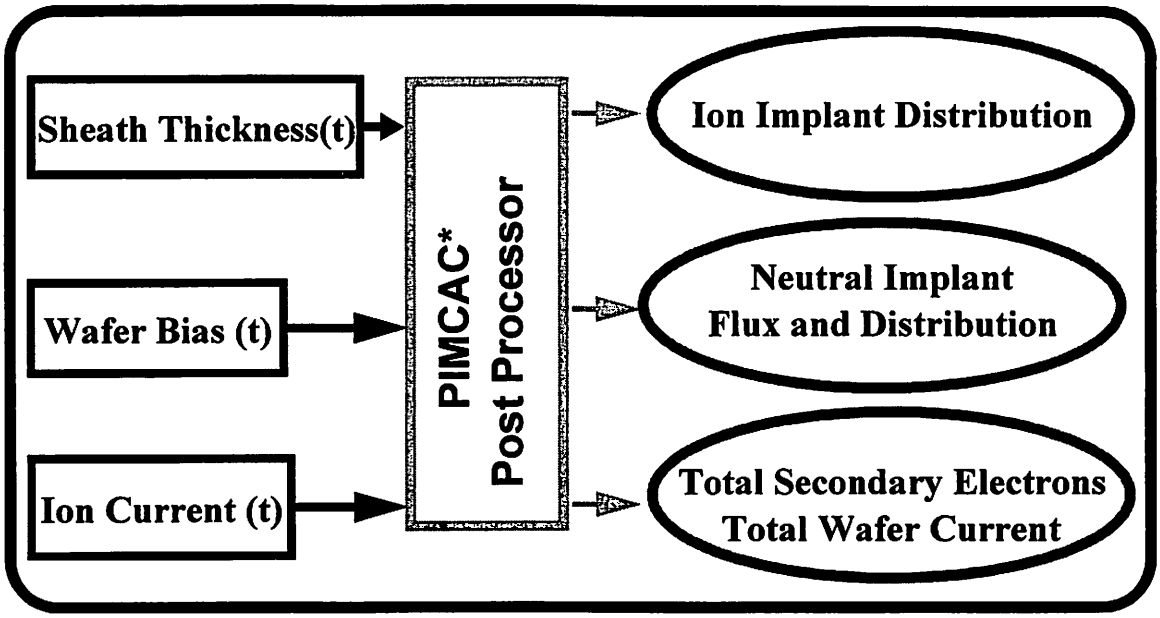

Figure 7-3 Plasma Implantation Monte Carlo Analysis of Collisions

The PICMIC post processor uses the results from the 1st order PIII model and calculates the ion implant distribution, the neutral implant flux, and distribution and the total wafer current.

where $V_{implant}$ is the implant voltage. The full cumulative distribution functions (CDF) of the neutrals, ions, and total flux are shown in Figure 7-4 for a 20 kV, 1 mtorr, argon implant. The ion density is $10^{10}/cm^3$. Even at this low pressure, and moderate ion density, 30% of the total impinging flux is neutrals. Only 45% of the total flux implants with the full 20 keV, while a full 30% implant with less than 10 keV.

The effect of including the energy dependency of the cross-section is apparent in Figure 7-4. The cross-section decreases significantly at higher energies (Figure 7-2), causing the ion CDF to tail upwards at high energies. Once an ion accelerates to high energies, the scattering probability is reduced, giving the ion a higher chance of continuing to the target unperturbed. Since most of the colliding ions are at low energies, the majority of the fast neutrals are also at low energies. For this example, the neutral mean energy is only 3 keV. Utilizing a fixed cross-section, which is necessary with analytical

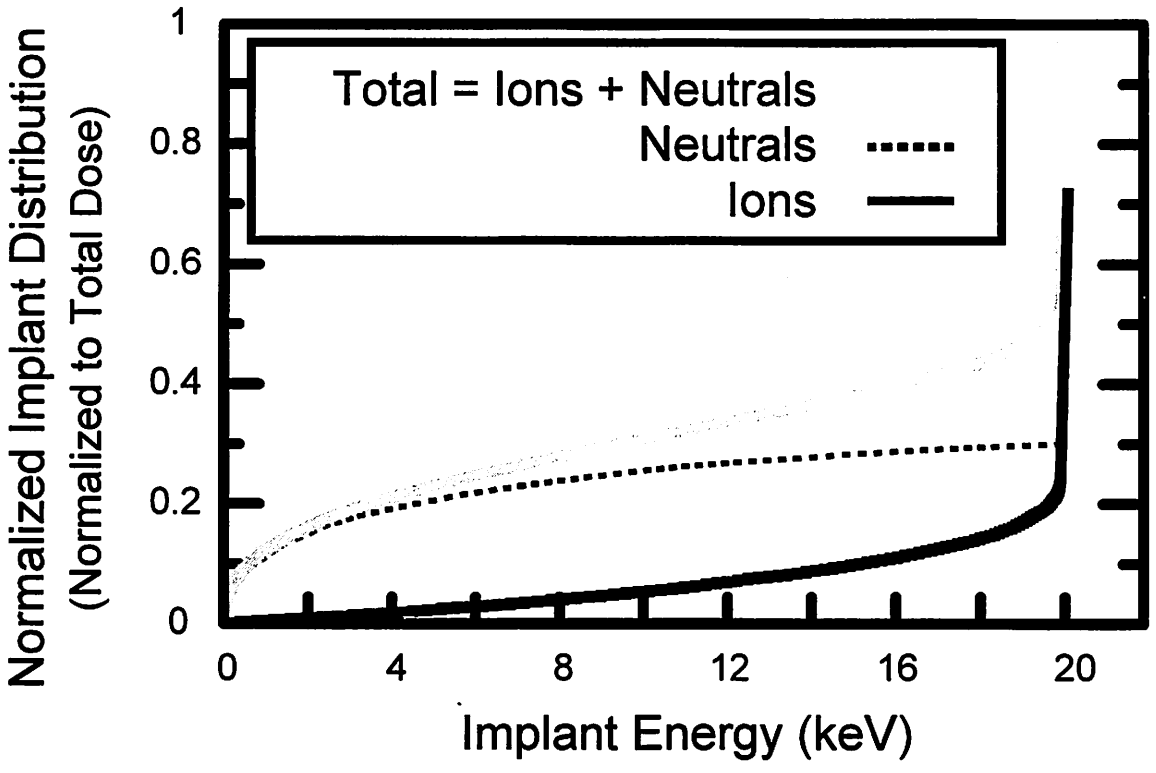

Figure 7-4 Implant Energy Cumulative Distribution Function

The implant energy cumulative distribution function is graphed for a 1 mtorr, 20 kV argon implant. The ion density is $10^{10}cm^{-3}$. The total dose consists of 30% neutrals, and 70% ions. Only 45% of the total implant is at the full 20 keV, while 30% implants with less than 10 keV.

methods, the fast neutral production being biased towards low energies is ignored. Finally, the large cross-section at low energies increases the total neutral flux considerably, which would be overlooked by a constant cross-section (usually assumed to be the value of the cross-section at about 1 keV).

In addition to broadening the implant energy distribution, and increasing the dopant flux rate, collisions increase the wafer current. Although neutrals themselves do not contribute to the current, they do eject secondary electrons. For high energy implants, the secondary electron yield is proportional to the square root of the impinging energy (proportional to the particle velocity). The total wafer current is the sum of the ion current, plus the secondary electrons from ion and neutral impacts. The total current must be

greater than the ion current alone without collisions. Energy flux conservation, with charge exchange as the dominant collision interaction, insists that

$$\sum_{k=1}^{n+i} E_k = i \cdot q \cdot V_{pulse} \tag{7-6}$$

where n, i, E_k are the number of neutrals, ions, and the particle implant energy, respectively. Since secondary electron current is proportional to the square root of the particle energy, splitting each ion's energy among a larger population of lower energy particles ejects more secondaries than a smaller population of higher energy.

$$\sum_{k=1}^{n+i} \sqrt{E_k} \geq i \cdot \sqrt{q \cdot V_{pulse}} \tag{7-7}$$

Therefore, collisions increase the total target current. Appendix D explains Equation (7-7) in more detail.

Figure 7-5 plots the Monte Carlo simulation results for the current contributions of ions, neutrals, and the total target current, for a 3 kV argon implant. The ion density is $10^{10}/cm^3$, while the pressure is varied. Appendix section C.2 lists the Matlab source code for the simulation. Collisions increase with pressure, thus higher pressures will have an increased target current (Equation (7-7)). At 10 mtorr, collisions have raised the current by 20% over the collisionless case. For higher implant voltages, the effect is even more prominent.

7.7 Experimental Results

As discussed in the previous section, collisions broaden the implant energy distribution, increase the dopant flux rate, and increase the target current. Of these effects, the simplest to measure is the increase in target current.

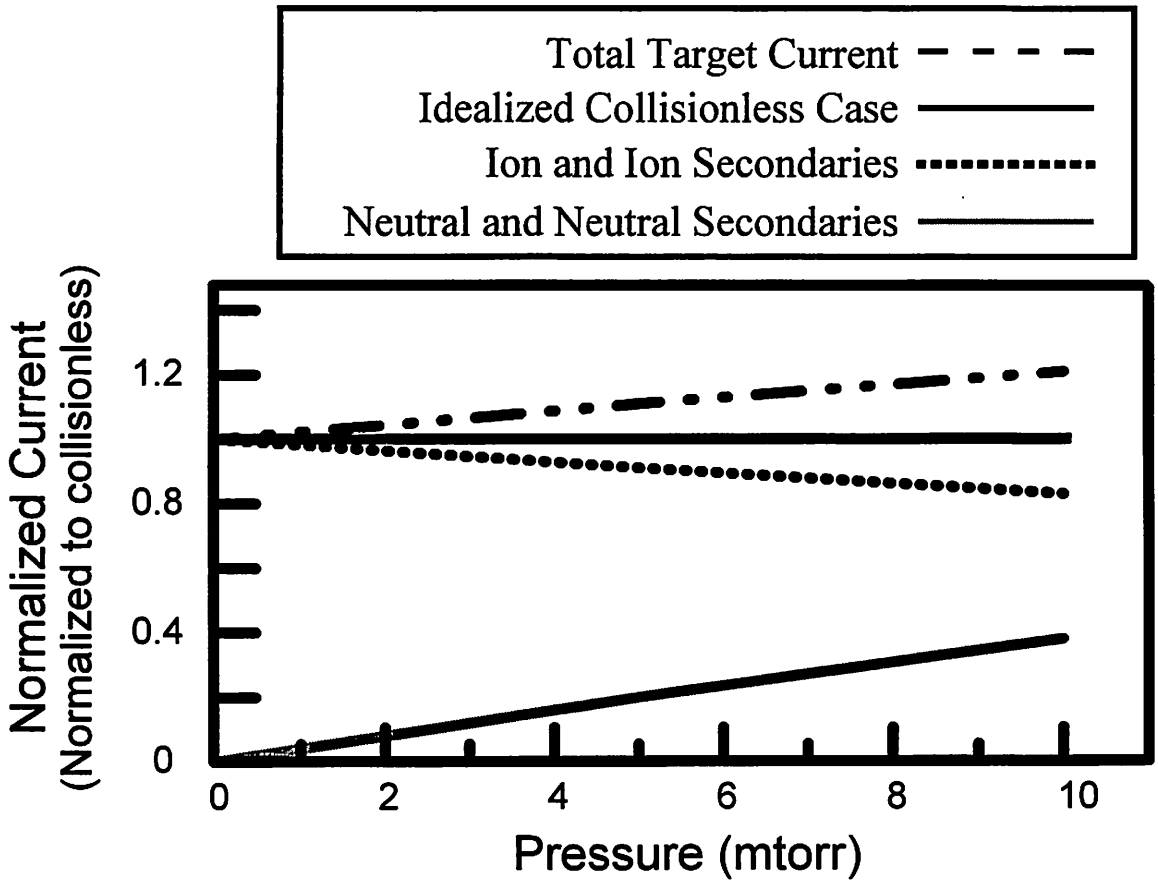

Figure 7-5 Simulation of Target Current with Collisions

The target current including charge exchange collisions is calculated through Monte Carlo simulation. Collisions are more prominent at higher pressures. Since collisions produce fast neutrals, which eject secondary electrons, the total target current is increased at higher pressures (with constant ion density). The simulation is for an 3 kV argon implant with an ion density of $10^{10}/cm^3$.

For this experiment, the gas pressure and target bias control the degree of collisionality. To accurately compare target currents for different pressure plasmas, the base ion flux should be kept constant for all conditions. The ion flux is equal to the ion density multiplied by the Bohm velocity. Since the Bohm velocity decreases with higher pressures, the ion density must be increased to maintain a constant ion flux. The ion flux is monitored by measuring the target current at low implant voltages (around 100 V). At low implant voltages, the sheath thickness is minuscule and collisions are negligible. The

input microwave power is adjusted for each pressure to maintain the same ion density and Bohm velocity product (target current), ensuring equal ion fluxes at all pressures in the collisionless regime. Higher pressure plasmas also have a lower electron temperature, which is factored into the simulation. Ensuring equal collisionless ion fluxes at all pressures, allows an accurate comparison of target current with higher implant voltages (the collisional regime).

Figure 7-6 plots the measured and simulated target currents for an argon plasma with an ion density of $10^{10}/cm^3$ for three pressures (1, 5, and 10 mtorr) and three voltages (1 kV, 3 kV, and 6 kV). Raising the target bias increases the number of collisions. This effect is more significant at higher pressures. At 6 kV, the current for 10 mtorr is ~35% greater than the current for the 1 mtorr case. As shown in Figure 7-6, the simulation results match the data fairly well.

7.8 Conclusion

Plasma Immersion Ion Implantation is an alternative implantation technique for high dose, large area, and low energy applications. In contrast to conventional beam line implantation, PIII implantation occurs at the same pressure as plasma generation (in the mtorr range). Therefore, ion-neutral collisions during ion acceleration become significant for a large portion of PIII implants. Even implants at 1 mtorr and below operate in the slightly collisional regime (greater than 10% of the accelerated ions undergo collisions).

With a Monte Carlo simulation program, the effect of collisions on the implant energy distribution, dose rate, and target current is computed. Charge-exchange collisions in the sheath produce fast neutrals. These fast neutrals implant into the target, increasing the total dose rate of the implant. The increased dose rate comes at the cost of implant energy, since the total flux energy must be conserved. Therefore the higher dose rate reduces the average implant energy, broadening the implant energy distribution. Although the fast neutrals themselves do not contribute to the target current, they eject

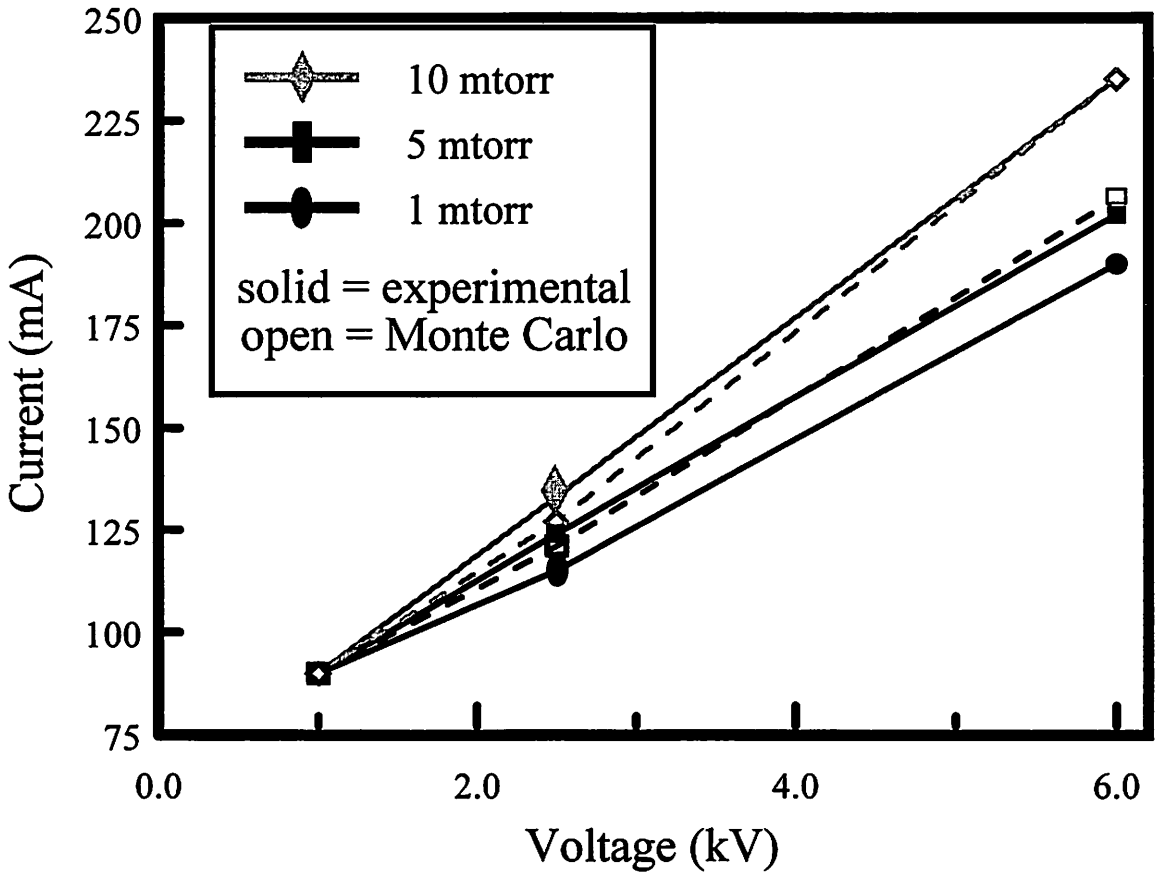

Figure 7-6 Target Current Increasing with Collisions.

Raising either the pressure or the implant voltage increases the number of ion-neutral collisions in the sheath. Because of the increased secondary electron current from fast neutral implantation, collisions lead to higher target currents. This effect is measured experimentally with an argon plasma. The 10 mtorr line, which is most effected by collisions, has the largest slope, with a current 35% higher than the 1 mtorr case at 6 kV, thus revealing the collisions in the sheath.

secondary electrons which enhance the measured wafer current. This increase in current has been experimentally measured as a function of pressure and implant voltage.

Typical PIII implants operate in the collisional regime. Hence, collision effects must be included in comprehensive PIII models to estimate the implant energy distribution, dose rate, and target current, more accurately.

The full model developed in the previous chapters accounts for rise and fall times, matrix sheath implantation, dielectric implantation, and now ion-neutral charge-exchange

collisions in the sheath. Combining all the modules creates a fairly comprehensive and accurate one dimensional PIII dose and implant simulator.

References

[7-1] J. K. Chester, "Effects of Expanding Sheaths on the Operation of Mercury-Arc Rectifiers", *Journal of Science and Technology*, vol. 37, no. 1, pp. 2-10, 1970.

[7-2] M. Shamim, J. T. Scheuer, and J. R. Conrad, "Measurements of Spatial and Temporal Sheath Evolution for Spherical and Cylindrical Geometries in Plasma Source Ion Implantation", *Journal Applied Physics*, vol. 69, pp. 2904-2908, 1991.

[7-3] M. A. Lieberman, "Model of Plasma Immersion Ion Implantation," *Journal Applied Physics*, vol. 66, no. 7, pp. 2926-2929, 1989.

[7-4] R. A. Stewart and M. A. Lieberman, "Model of Plasma Immersion Ion Implantation for Voltage Pulses with Finite Rise and Fall Times," *Journal Applied Physics*, vol. 70, no. 7, pp. 3481- 3487, 1991.

[7-5] W. D. Davis, and T. A. Vanderslice, "Ion Energies at the Cathode of a Glow Discharge," *Physical Review*, vol. 131, no. 1, pp. 219-228, July 1, 1963.

[7-6] I. Abril, A. Gras-Marti, and J. A. Valles-Abarca, "The Contribution of Fast Neutrals to Cathode Erosion in Glow Discharges," *Journal of Physics D*, vol. 17, pp. 1841-1849, 1984.

[7-7] I. Abril, A. Gras-Marti, and J. A. Valles-Abarca, "Energy Distributions of Particles Striking the Cathode in a Glow Discharge," *Physical Review A*, vol. 28, no. 6, pp. 3677-3678, December, 1983.

[7-8] T. Tokonami, and T. Makabe, "Modeling of Physical Etching based on the Two-dimensional Velocity Distributions of Ions and Fast Neutrals using the Boltzmann Equation," *Journal Applied Physics*, vol. 72, no. 8, pp. 2008-2014, October 15, 1992.

[7-9] V. Vahedi, M. A. Lieberman, M. V. Alves, J.P. Verbonocoeur, and C. K. Birdsall, "A One-dimensional Collisional Model for Plasma Immersion Ion Implantation," *Journal Applied Physics*, vol. 69, no. 4, pp. 1275-1282, February 15, 1991.

[7-10] V. Vahedi, R. A. Stewart, and M. A. Lieberman, "Analytic Model of the Ion Angular Distribution in a Collisional Sheath," *Journal of Vacuum Science and Technology A*, vol. 11, no. 4, pp. 2664-2681, July, 1993.

[7-11] R. T. Farouki, S. Hamaguchi, and M. Dalvie, "Monte Carlo Simulations of Space-charge-Limited Ion Transport through Collisional Plasma Sheaths," *Physical Review A*, vol. 44, no. 4, pp. 3323-3329, August 15, 1991.

[7-12] S. Qin, C. Chan, and N. E. McGruer, "Energy Distribution of Boron Ions during Plasma Immersion Ion Implantation," *Plasma Sources Science. Technology.*, vol. 1, no. 1, pp. 1-6, March, 1992.

[7-13] D. Wang, T. Ma, and Y. Gong, "A Monte Carlo Simulation Model for Plasma Source Ion Implantation," *Journal Applied Physics*, vol. 73, no. 9, pp. 4171-4175, May 1, 1993

[7-14] A. V. Phelps, "Cross Sections and Swarm Coefficients for Nitrogen Ions and Neutrals in N_2 and argon Ions and Neutrals in Ar for Energies from 0.1 ev to 10 keV," *J. Phys. Chem. Ref. Data*, vol. 20, no. 3, pp. 557-573, 1991.

[7-15] A. V. Phelps, "Cross Sections and swarm Coefficients for H^+, H_2^+, H, H_2, and H^- in H_2 for Energies from 0.1 ev to 10 keV," *J. Phys. Chem. Ref. Data*, vol. 19, no. 3, pp. 653-675, 1990.

[7-16] R. W. Thomae, B. Seiler, and et. al., "High Current Ion Implantation by Plasma Immersion Technique," *Nuclear Instruments and Methods in Physics Research*, vol. B99, pp. 569-572, May, 1995.

[7-17] S. Qin, Y. Zhou, T. Nakatsugawa, I. F. Husein, C. Chan, and T. J. King, "Plasma Ion Implantation Hydrogenation of Poly-Si CMOS Thin-film Transistors at Low Energy and High Dose Rate using an Inductively-coupled Plasma Source," *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1324-1328, June 1998.

[7-18] K.U. Riemann, "Consistent Analysis of a Weakly Ionized Plasma and its Boundary Layer," *Journal of Physics D*, vol. 25, pp. 1432-1442, 1992.

8 Universal Plasma Charging Damage Model

8.1 Introduction

Plasma charging damage has emerged as an important reliability concern. One of the most important issues is the relationship between gate oxide scaling and plasma damage, with groups presenting seemingly conflicting results. Many report damage increasing with gate dielectric scaling [8-1 - 8-3], while others report the opposite trend [8-4 - 8-7]. Multiple papers also describe the effect of altering the plasma parameters and wafer bias on the extent of charging damage [8-8 - 8-10]. To unify the published results, this chapter develops a generalized universal plasma charging damage (UPCD) model that predicts plasma charging damage as a function of the plasma parameters, gate oxide thickness, circuit design, and wafer bias.

8.2 Model Framework

Calculating plasma charging damage requires the determination of the electrical stress to the oxide. The UPCD achieves this through a loadline analysis comparing the impedances of the plasma and gate oxide. Figure 8-1 depicts the flow chart outlining the steps for calculating plasma damage.

It is a three step procedure to calculate the amount of damage. First, the impedance of the plasma and gate dielectric as a function of voltage must be determined. The second step computes the electrical stress created by the plasma process by comparing the

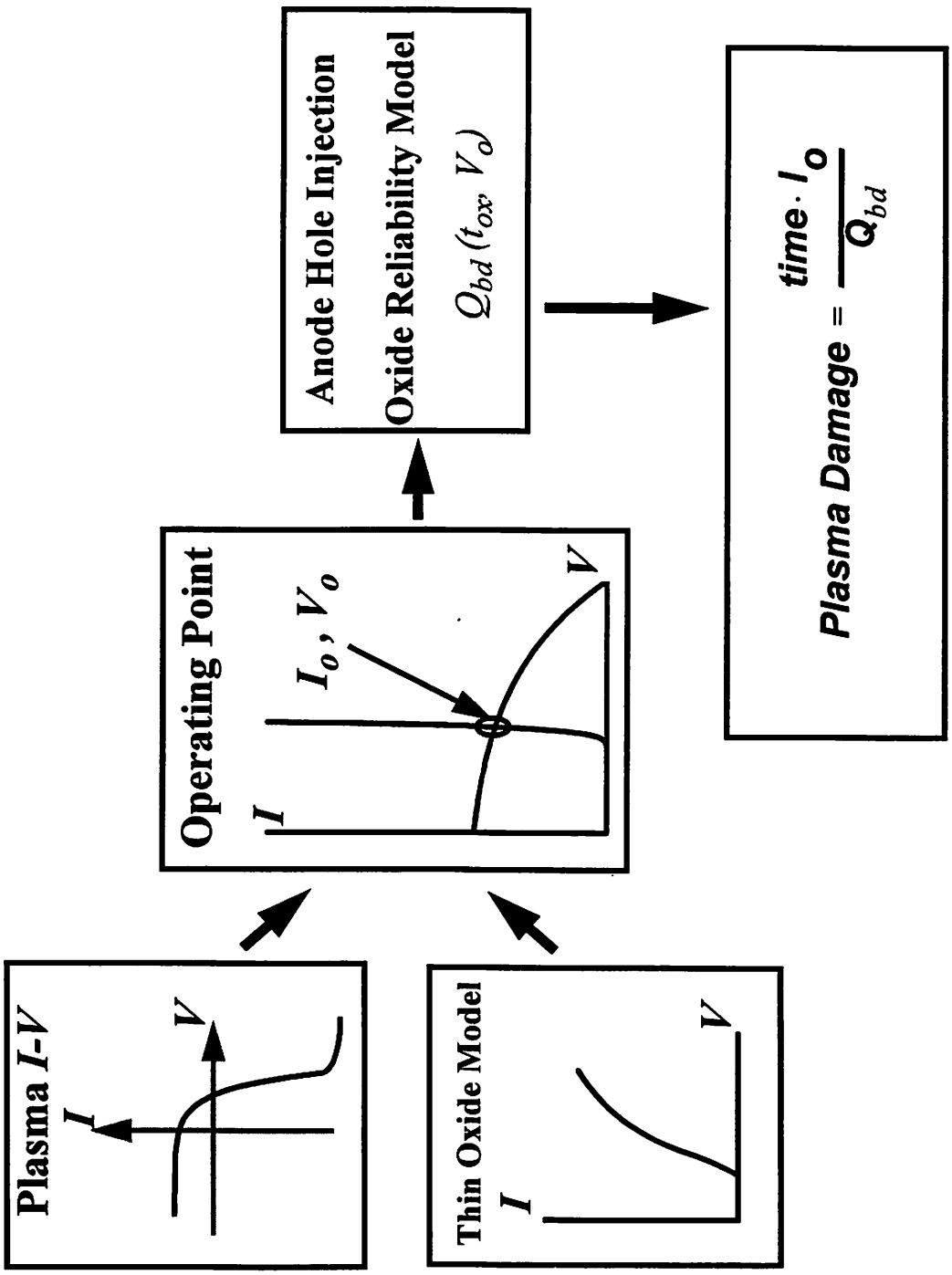

Figure 8-1 Universal Charging Damage Flowchart

Combining the plasma *I-V* and the thin oxide tunneling models in a loadline analysis determines the DC stress condition during the plasma process. Damage is calculated by comparing the plasma stress condition to an oxide reliability mobility. The damage metric is the percent of the lifetime of the oxide consumed during the plasma process.

plasma *I-V* curve with the thin oxide tunneling current load line. The last step correlates the plasma stress with oxide damage. This calculation uses the Anode Hole Injection model as the electrical stress model [8-11]. The metric for comparing different oxides thicknesses is the percentage of the total charge to breakdown of the oxide (Q_{bd}) consumed during the plasma process. The following sections describe in detail each of the modules in the flow chart (Figure 8-1).

8.2.1 Plasma Impedance

The first step in calculating plasma damage is determining the plasma impedance as a function of voltage. The full Langmuir equation for the plasma current is

$$I = I_{Ion} + I_{electron} = qn_i u_b + \frac{1}{4} qn_i v_e \exp\left(\frac{V_p - \Phi_p}{T_e}\right) \quad (8\text{-}1)$$

where n_i, u_b, v_e, V_p,Φ_p, and T_e, are the ion density, the Bohm velocity, the electron velocity, the plasma potential, the probe voltage, and the electron temperature, respectively. Three parameters, n_i, V_p, and T_e govern the plasma *I-V*. Machine design, excitation method, and the process conditions control the value of these parameters. Figure 8-2 shows a typical plasma *I-V*. Three regimes are identifiable: electron saturation, ion saturation, and mixed. With a positive bias, the probe collects the entire electron flux, and the negative current saturates. With a highly negative bias, the probe collects the entire positive flux, while repelling all the electrons, saturating the positive current. In between the two extremes, the probe collects a varying amount of electrons and ions. The theoretical and experimental plasma impedances are described in Chapter 2 in more detail.

The wafer actually interacts with a shifted version of the plasma *I-V* curve (Figure 8-3). The wafer bias, plasma non-uniformities, electron shading, and other effects regulate the magnitude and sign of the shift. The shift from plasma non-uniformities is well known

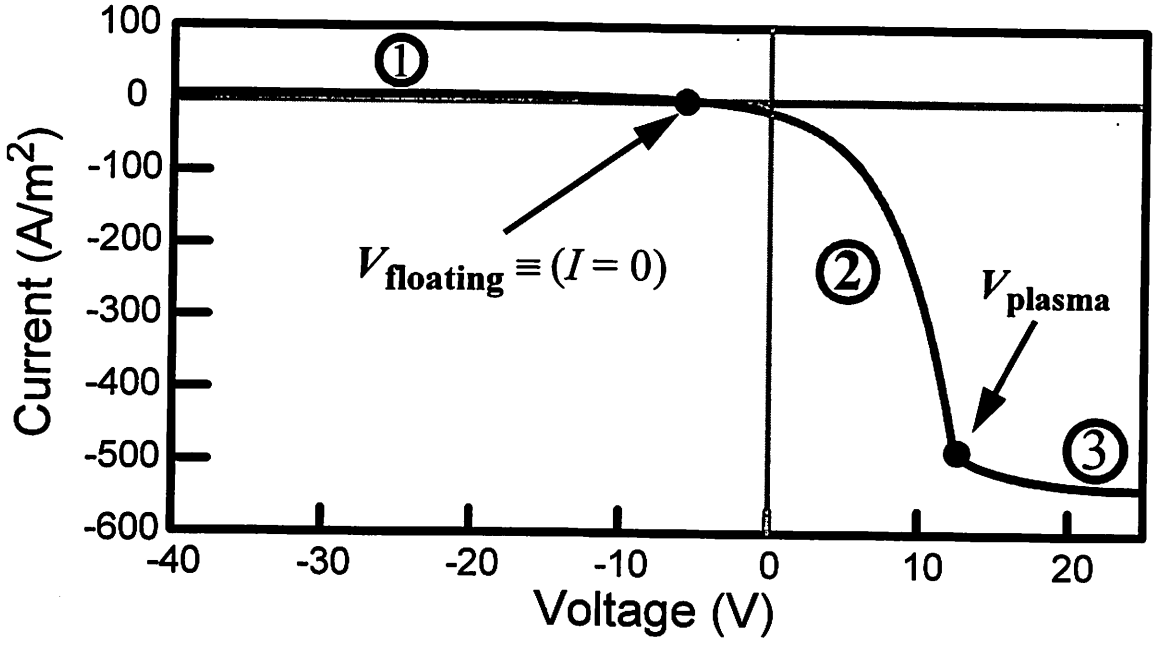

Figure 8-2 Plasma *I-V* Characteristic

The plasma *I-V* contains three regimes: 1) ion saturation, 2) mixed, and 3) electron saturation. The floating potential is defined as the zero current voltage, while the plasma potential is the knee in the curve that separates the electron saturation and mixed regimes.

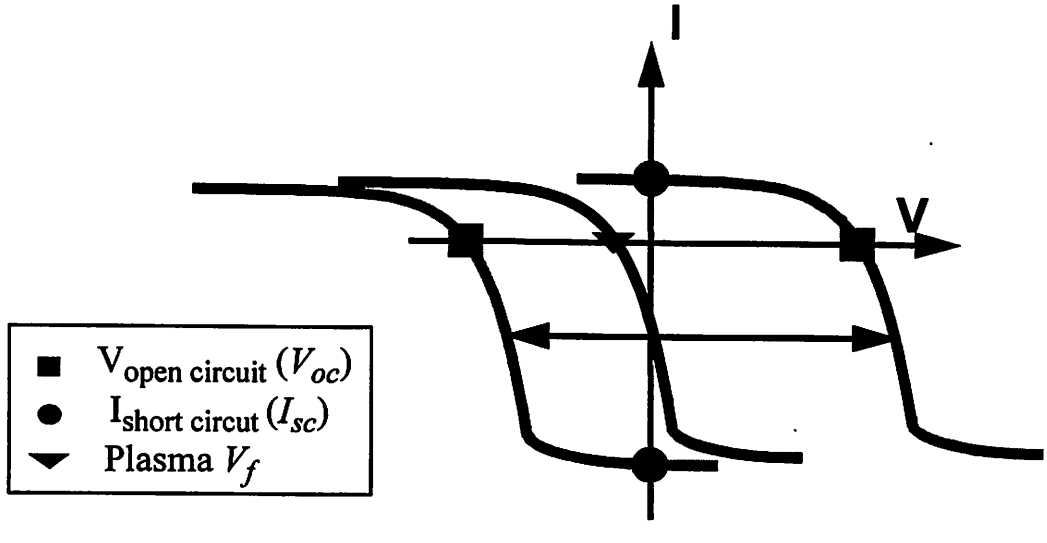

Figure 8-3 Plasma Loadline Shifting

The gate oxide interacts with a shifted version of the plasma Langmuir *I-V* curve. Various factors determine the magnitude of the shift, including the plasma non-uniformities and the wafer bias. The open circuit voltage and the short circuit current are key parameters of the shifted *I-V*.

$$V_{shift} = T_e \ln\left(\frac{n_{s1}}{n_{s0}}\right) \tag{8-2}$$

where V_{shift}, n_{s1}, and n_{s2} are the voltage shift, the maximum ion density, and the minimum ion density, respectively. The shift from electron shading is similar,

$$V_{shift} = T_e \ln\left(\frac{k_i}{k_e}\right) \tag{8-3}$$

where k_i and k_e are the shadowing factors for ions and electrons [8-11]. The shifting of the plasma curve simply modulates the open circuit voltage, or, in other words, the effective V_p of the plasma. For most plasma applications (RIE and PECVD), the plasma curve typically shifts in the positive direction. For other processes, e.g. Plasma Immersion Ion Implantation (PIII) [8-12, 8-13, 8-14], the plasma curve may shift in either the positive or negative directions. Determining the amount of shift for more general situations requires computer simulation. Chapter 11 details the simulation that calculates the shift for the PIII process.

Three parameters fully characterize the shape of the shifted *I-V*. The effective floating potential specifies the open circuit voltage (V_{oc}) of the system. The ion density correlates with the saturation currents, while the electron temperature determines the curvature of the plasma *I-V* in the mixed regime. T_e and n_i both influence the short circuit current ($I_{sc}(V=0)$). V_{oc} is the maximum stress voltage during the plasma process, while I_{sc} is the maximum stress current.

8.2.2 Thin Oxide Tunneling Current

The two relevant conduction mechanisms that govern thin oxide current are Fowler-Nordheim (F-N) tunneling and direct tunneling (DT) (Figure 8-4). The tunneling mechanisms are generally mutually exclusive, with only one dominating. Since the oxide barrier height (Φ_b) is 3.2 eV for a Poly Si - SiO_2 system, F-N tunneling dominates for

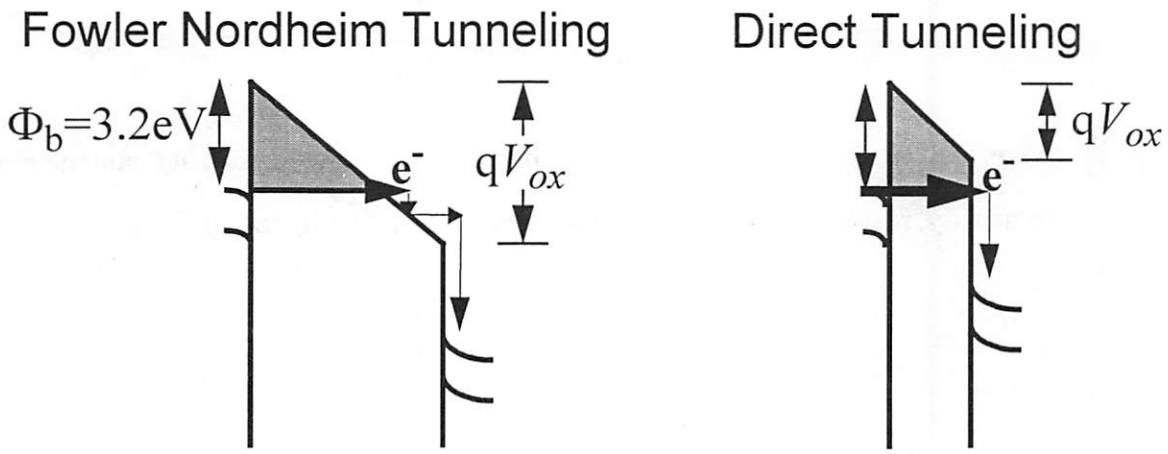

Figure 8-4 Tunneling Mechanisms in Thin Oxides

Fowler-Nordheim and Direct tunneling are the two dominant conduction mechanisms in thin oxides.

oxide voltages greater than 3.2 V. With F-N tunneling, the electrons tunnel through a triangular barrier into the conduction band of the gate dielectric. The electron undergoes scattering as it travels towards the anode with a characteristic interaction length, λ [8-19]. Finally, the electron enters the anode as a hot electron. The F-N tunneling current depends only on the electric field:

$$I_{FN} = K \times Area \times \left(\frac{V_{ox}}{t_{ox}}\right)^2 \times exp\left(\frac{-24.7}{V_{ox}/t_{ox}}\right) \quad (8\text{-}4)$$

The tunneling distance for the electrons is always $\frac{3.2}{E_{ox}}$, independent of the oxide thickness. Significant damaging F-N currents flow with electric fields greater than about 6 MV/cm.

The second relevant tunnelling mechanism, direct tunneling, dominates when the oxide voltage is less than 3.2 V. In this case, the electrons tunnel through the entire oxide, entering the anode without scattering in the oxide. The DT currents are dependent on both the oxide thickness, and the oxide electric field.

$$I_{DT} = K \times Area\left(\frac{V_{ox}}{t_{ox}}\right)^2 \times exp\left(\frac{-24.7 \times \left(1 - \left(1 - \frac{qV_{ox}}{\Phi_b}\right)^{3/2}\right)}{V_{ox}/t_{ox}}\right) \quad (8\text{-}5)$$

With DT tunneling, the energy of the hot electron equals the oxide voltage. This contrasts with F-N tunneling, in which the hot electron losses some energy when scattering in the oxide. Since Equation (8-5) depends on both t_{ox} and E_{ox}, the minimum electric field for conduction depends on the oxide thickness, and decreases dramatically with oxides thinner than 4 nm. Significant DT tunneling currents flow with electric fields even less than 5 MV/cm. Figure 8-5 graphs Equation (8-4) and (8-5) for 2.5 nm to 10 nm oxides. The dashed lines are extrapolations of the F-N equation into the DT regime. For the same applied voltage, the DT mechanism yields much higher currents than predicted by the F-N extrapolation. Conversely, for the same current, DT requires a comparatively smaller voltage (and electric field). The lower required electric field for DT currents affects plasma damage, as discussed in Section 8.3.

8.2.3 Operating Point

With a known plasma impedance and gate dielectric impedance, a loadline analysis may be performed. Combining the shifted plasma curve with the oxide load line determines the voltage (V_o) across the oxide and current (I_o) through the oxide during the plasma process (Figure 8-6).

8.2.4 Oxide Reliability Model

With the stress conditions (I_o and V_o) established, an oxide reliability model calculates the damage sustained during the plasma process. For this work, the Anode Hole Injection model correlates the stress condition and the amount of damage sustained This $1/E$ model is physically based, and has been extensively tested against electrical stress

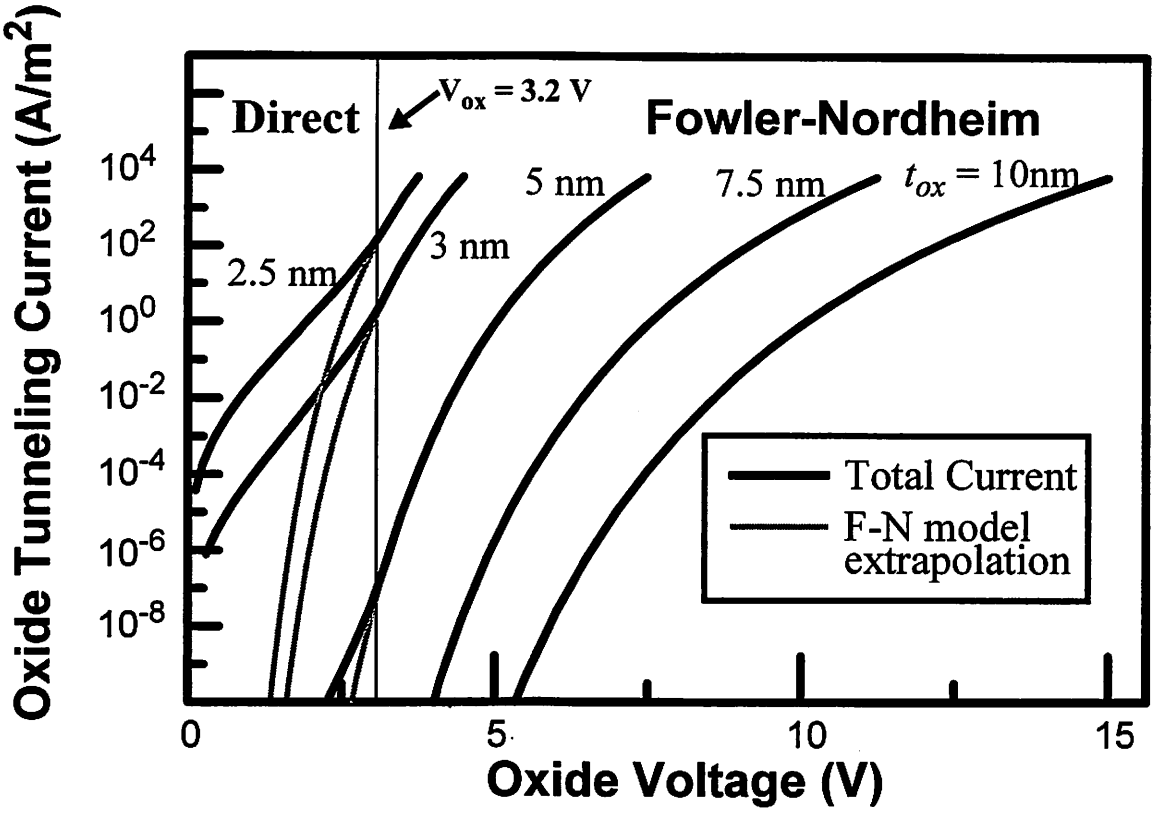

Figure 8-5 Tunneling Current in Thin Oxides

With a poly-silicon gate, Fowler-Nordheim tunneling occurs with oxide voltages greater than 3.2 V, while direct tunneling dominates with voltages less than 3.2 V. The F-N extrapolation into the DT regime shows that DT allows much more current than the F-N mechanism at lower electric fields.

measurements. The model fits data for oxide thicknesses from 2.5 nm to over 12 nm (Figure 8-7). Figure 8-8 depicts the three step damage mechanism for this model. First, an electron tunnels from the cathode to the anode by either F-N or DT tunneling mechanisms. If the gate voltage is greater than 3.2 V, the electron tunnels into the conduction band of the oxide. Drifting towards the anode, the electron scatters with a characteristic length, λ. The electron reaches the anode with energy E_{gain}:

$$E_{gain} = \begin{cases} \Phi_b + E_{ox}\lambda\left(1 - \exp\frac{-1}{\lambda}\left(t_{ox} - \frac{\Phi_b}{E_{ox}}\right)\right) & V_{ox} > \Phi_b \\ V_{ox} & V_{ox} \le \Phi_b \end{cases} \quad (8\text{-}6)$$

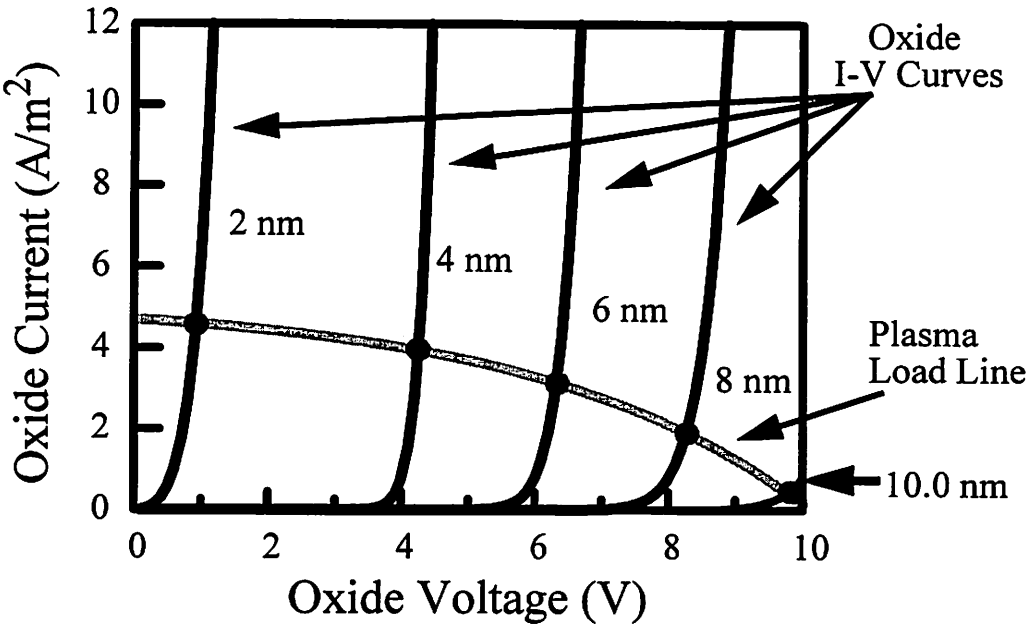

Figure 8-6 Loadline Analysis

The stress condition during plasma processing is the intersection between the plasma I-V curve and the oxide load line.

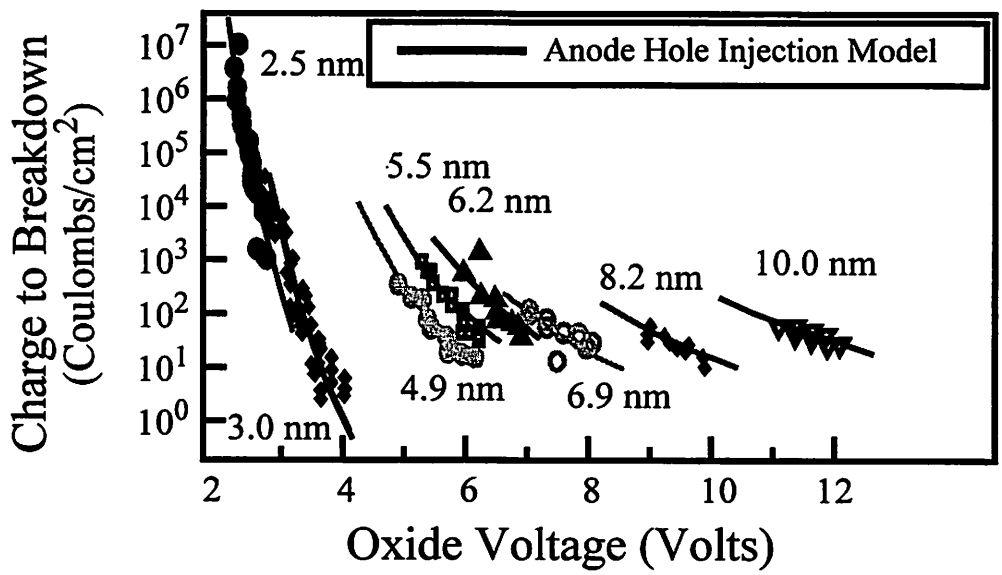

Figure 8-7 Anode Hole Injection Model Fit

The Anode Hole Injection model accurately fits high stress field oxide breakdown data from 2.5 nm to 10 nm. This figure is from [8-19].

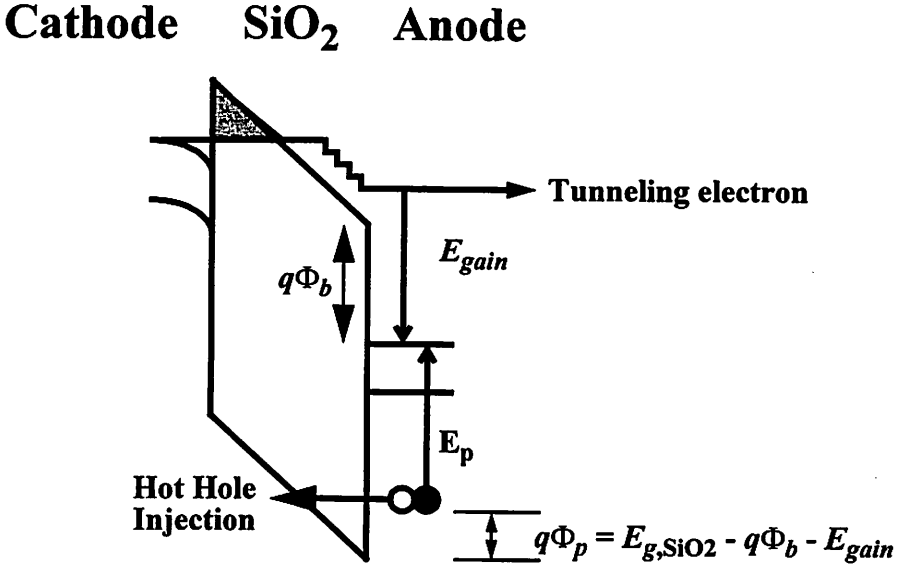

Figure 8-8 Anode Hole Injection Damage Mechanism

In the Anode Hole Injection model, an electron tunnels from the cathode to the anode. In the anode, the hot electron promotes a deep valence electron to the conduction band, creating a hot hole. This hot hole injects into the oxide layer, creating damage.

where $\lambda = 1.5$ nm [8-17]. The second step is the elastic collision of the hot electron with a deep valance band electron. The collision promotes the valance band electron to the bottom of the conduction band, creating a hot hole. Experiments determine the hot hole production efficiency, α, to be independent of the hot electron energy and approximately equal to 0.08 [8-19]. Finally, the hot hole injects into the oxide, tunneling through a barrier Φ_p.

$$\Phi_p = E_{g,SiO_2} - \Phi_b - E_{gain} = 8eV - 3.2eV - E_{gain} \quad (8\text{-}7)$$

The Anode Hole Injection model predicts that breakdown transpires when the hot hole fluence reaches a critical value (Q_p), independent of stress voltage and electric field. For a particular stress condition, one can define a critical electron fluence (Q_{bd}), and the time to breakdown (t_{bd}) which are related to Q_p

$$Q_{bd} = \frac{Q_p}{0.08}\exp\left(\frac{B}{E_{ox}}\left[\Phi_p \langle V_{ox}\rangle^{\frac{3}{2}}\right]\right) \tag{8-8}$$

$$B = \frac{8\pi\sqrt{2m_{p,ox}}}{3hq} \tag{8-9}$$

where h is Planck's constant, and $m_{p,ox}$ is the effective hole mass in the oxide which is assumed to be 0.2 m_o [8-18]. Q_p decreases with oxides thinner than ~6 nm. Figure 8-9 plots the experimentally determined values of Q_p. The reduction of Q_p implies that thin-

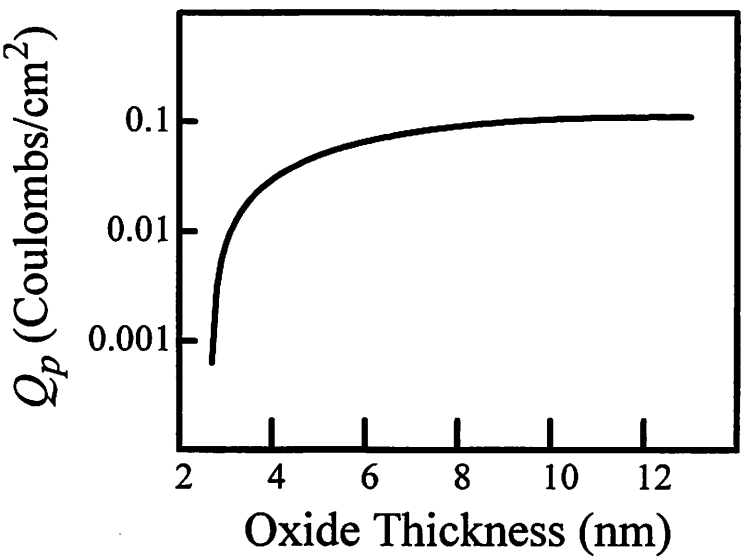

Figure 8-9 Hole Injection until Breakdown

For a give oxide thickness, oxide breakdown occurs after a certain hole fluence, Q_p. The critical hole fluence is constant above 8 nm, but drops quickly for thicknesses less than 4 nm. This is an experimental fit [8-19].

ner oxides require fewer traps before breakdown. This agrees with recent percolation models for ultra-thin oxide breakdown [8-15, 8-16]. The reduction in the critical hole fluence is overwhelmed by the diminished hot hole tunneling current (since Φp is greater), such that thinner oxides are more robust. Refer to Scheugraf [8-19] for more information on this model.

In summary, the salient points of the model are:

1) A critical hole fluence, Q_p, determines breakdown, independent of stress condition.

2) Q_{bd} and t_{bd} depend on the stress voltage and electric field. Reducing either the voltage or the field increases the oxide lifetime.

Figure 8-10 plots t_{bd} as a function of gate oxide thickness under constant voltage and constant current stress. With constant voltage stress, the thinner oxides experience a higher electric field, and subsequently lower lifetimes. With constant current stress all the oxides are stressed at the same electric field (for oxides with F-N tunneling), but the thinner oxides are stressed at lower voltages, and therefore have longer lifetimes. The difference between constant current and constant voltage stress is crucial in understanding plasma damage as a function of gate oxide thickness.

Over the past decade there has been much debate over the best oxide reliability model for extrapolating low field, 10 year lifetimes (most recently [8-20]). Since plasma charging stresses oxides at high fields, greater than 6 MV/cm, all of the reliability models perform well. Therefore, the choice of the Anode Hole Injection model does not affect the plasma damage simulation results.

8.2.5 Plasma Damage Metric

Combining the operating stress condition and the oxide reliability model determines the plasma charging damage. The damage metric for this work is:

$$Plasma\ Damage \equiv \frac{time \cdot I_o}{Q_{bd}} \tag{8-10}$$

which is the total charge conducted through the oxide during the plasma process divided by the total charge to breakdown. The ratio is also equal to the fraction of the oxide lifetime consumed during the plasma process.

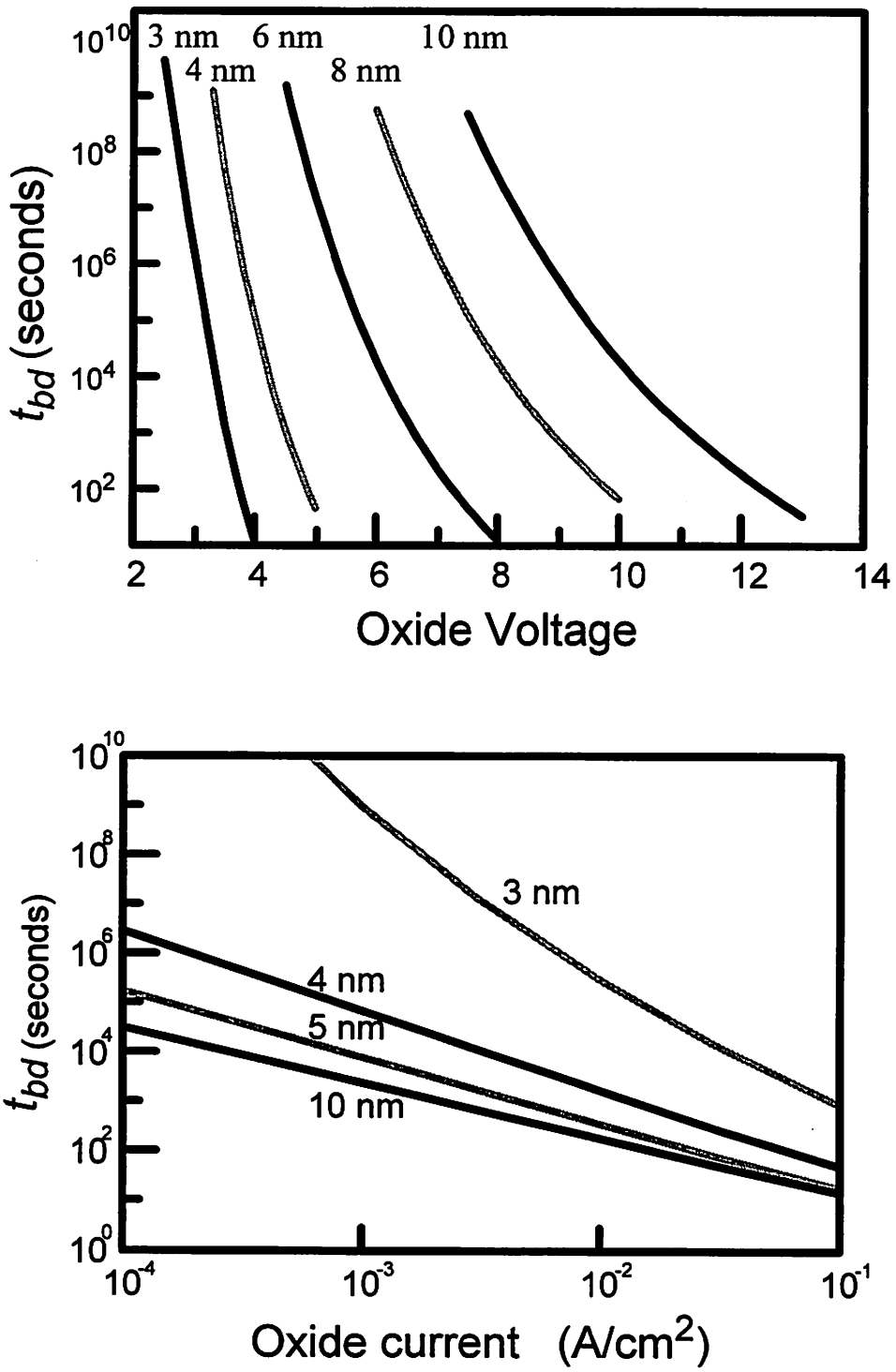

Figure 8-10 Time to Breakdown for Constant Current and Voltage Stress

With constant voltage stress (top), thicker oxides having longer lifetimes, while with constant current stress (bottom), thinner oxides are more robust.

8.3 Simulation Results

The methodology developed above is now applied to compute the damage trends as a function of gate oxide thickness, plasma ion density, circuit antenna ratio, plasma electron temperature, and open circuit voltage. For simplicity, an Argon plasma is analyzed.

8.3.1 Load Line Analysis

Figure 8-6 depicts the load line analysis for four oxide thicknesses and an Argon plasma with an ion density of $10^{10}/cm^3$, an electron temperature of 4 eV, and an open circuit voltage of 10 V. The operating points (indicated by circles) specify the stress voltages and stress currents. To calculate plasma damage, these stress conditions are input into the oxide reliability model. The semi-log plot in Figure 8-11 shows the percent of the total Q_{bd} consumed during the simulated plasma process. When scaling the thickness down from 15 nm, the damage initially increases, until peaking at ~6 nm. The damage drops steeply for oxide thinner than 6 nm. Oxides thinner than 3 nm suffer little damage in this example.

The stress voltages and stress electric fields as a function of gate oxide thickness explain the damage trends (Figure 8-12). Three stressing regimes are identifiable in Figure 8-12. For thicker oxides, the plasma acts like a constant voltage source. With oxide scaling, the stress voltage remains constant, but the electric field follows a $1/t_{ox}$ dependence, since $E_{ox} = \frac{V_{ox}}{t_{ox}}$. For this simulated plasma, all oxides thicker than 10 nm are stressed by the V_{oc} of 10 volts. With constant voltage stressing, thinner oxides are damaged more (Figure 8-10). This is reflected in Figure 8-11 by the increase in damage from 15 to 10 nm.

Further thinning of the oxide advances the stress conditions from the constant voltage regime to the nearly constant current region. Under constant current stressing, the electric field remains constant, but the voltage drops linearly. From 10 nm down to 3 nm,

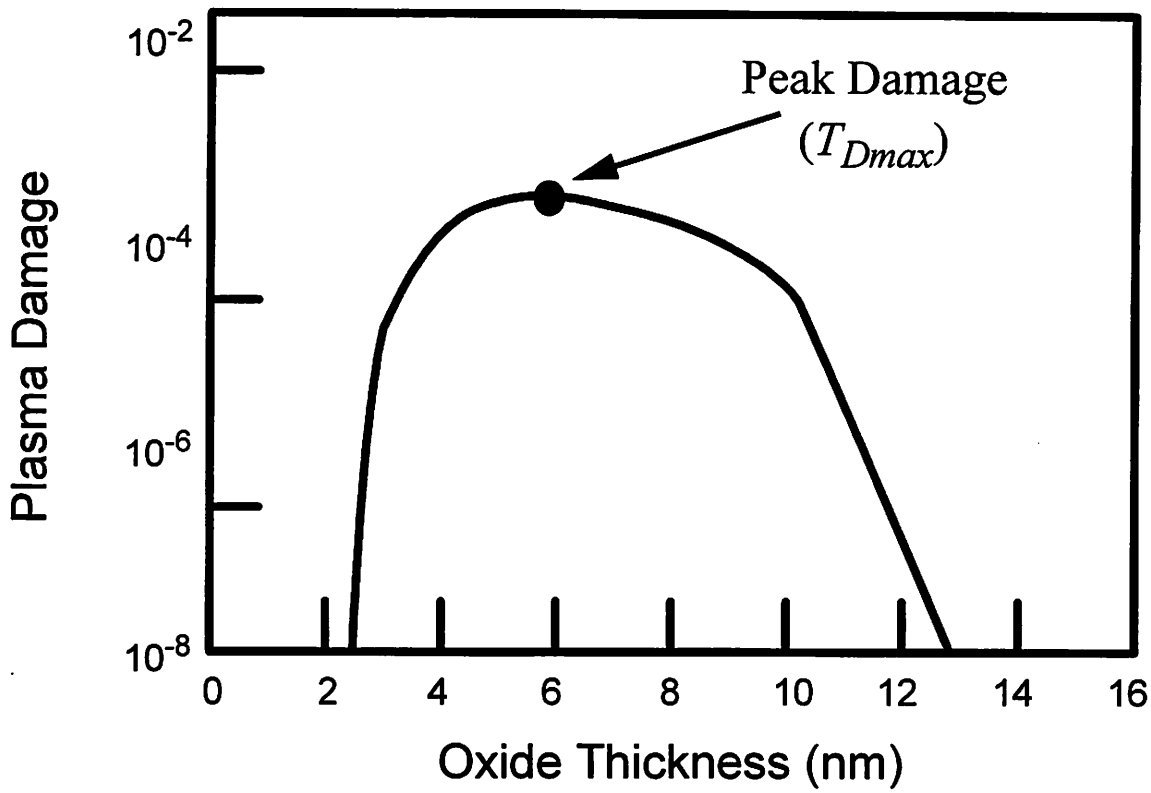

Figure 8-11 Plasma Damage and Gate Oxide Thickness

With gate oxide scaling, plasma damage initially increases, peaking at ~6 nm. The damage metric is the percent of the total oxide lifetime consumed in the plasma process.

the electric field rises slightly from 10 to 11 MV/cm, while the voltage drops almost linearly. The Anode Hole Injection model dictates that either a lower voltage or a lower electric field diminishes damage. Since the electric field rises slowly, but the voltage decreases linearly, the damage peaks somewhere in this regime. For the simulated plasma, the damage peaks at ~6 nm.

Finally, the transition to direct tunneling conduction with oxides thinner than 3 nm (Figure 8-5), reduces both the electric field and voltage, further diminishing the damage. The direct tunneling transition is seen is Figure 8-11 by the change in slope at ~ 3 nm.

Ultra-thin oxides are more robust for two reasons: lower stress voltages and lower stress electric fields. It is important to understand that the primary force reducing damage

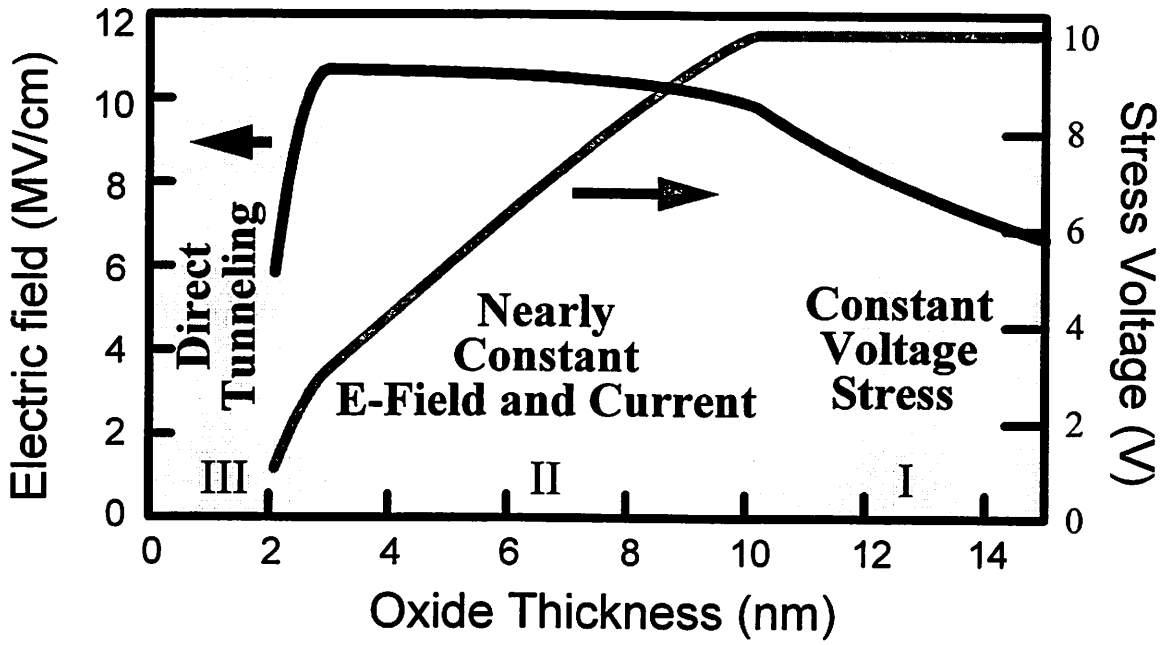

Figure 8-12 Stress Voltage and Stress Electric Field

The voltage and electric field stress as a function of gate oxide thickness. As the oxide thins, the electric field increases at first, levels off, and then drops precipitously in the direct tunneling regime. The voltage, at first, remains constant, then decreases nearly linearly in the nearly constant current regime, and drops abruptly in the direct tunneling regime. The reduction of stress voltage in the nearly constant current regime is the force behind the reduction in damage for ultra-thin oxides.

for oxides thinner than 6 nm is the drop in stress voltage from the constant current source effect, and not direct tunneling. Direct tunneling simply augments the slope of the damage reduction.

8.3.2 Antenna Effect

During IC processing, the plasma usually does not impinge upon the oxide directly, but rather hits a surface conductor. This conductor collects charge and funnels it down to stress the gate oxide. The conductor acts as an antenna since it collects charge from a large area, multiplying the effective impedance of the plasma. Antennas multiply the effective current density for the gate oxide, magnifying plasma damage.

Since antennas amplify damage, large antenna test structures are the preferred method for detecting and studying charging damage. Extremely large antennas (some-

times as large a 1 million) are unquestionably sensitive to charging damage, but as will be discussed below, may lead to erroneous conclusions.

8.3.2.1 Antenna Types

There are two broad categories of antennas, area and edge structures (Figure 8-13).

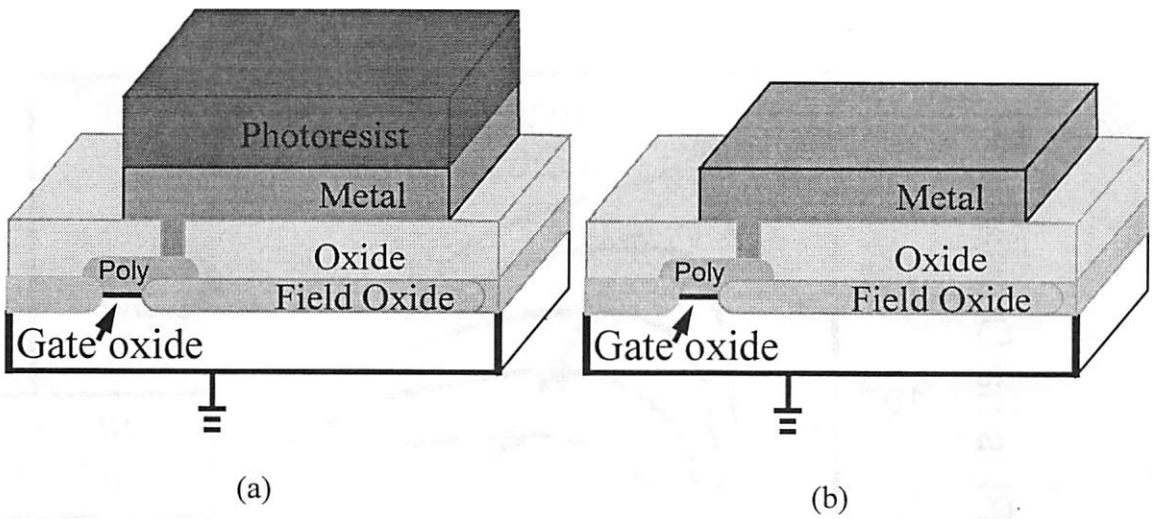

Figure 8-13 Edge vs. Area Antennas

The surface conductor collects charge, funneling it towards the gate. The large area of the surface conductor compared to the gate oxide area effectively multiples the plasma impedance curve. The multiplication factor is either proportional to the exposed edge area of the surface conductor (a), or with the entire area of the surface conductor, including the edges (b).

The antenna type depends on whether photoresist covers the antenna during the plasma process. For example, during the overetch of photoresist ashing, the top of the antenna is exposed to plasma, allowing the entire antenna to collect charge. In other processes, such as metal etch, the PR covers the antenna, and only the antenna's edges collect charge.

Area antennas multiply the plasma impedance by the antenna area to the gate area ratio (*AR*). The *AR* is not easily defined for edge antennas, but is necessarily less than the area ratio. In fact, the effective *AR* for edge sensitive processes could be an order of magnitude less than the actual AR. The antenna sizes for this work are all area antennas. The

sizes must be significantly adjusted for edge dependent processes such as metal etch. For example, a 1000X antenna in a metal etch is approximately equivalent to a 100X area antenna (dependent on the area to edge ratio of the antenna).

8.3.2.2 Antenna Effect Simulation

Figure 8-14 reveals two plasma damage trends as a function of *AR* for area antenna

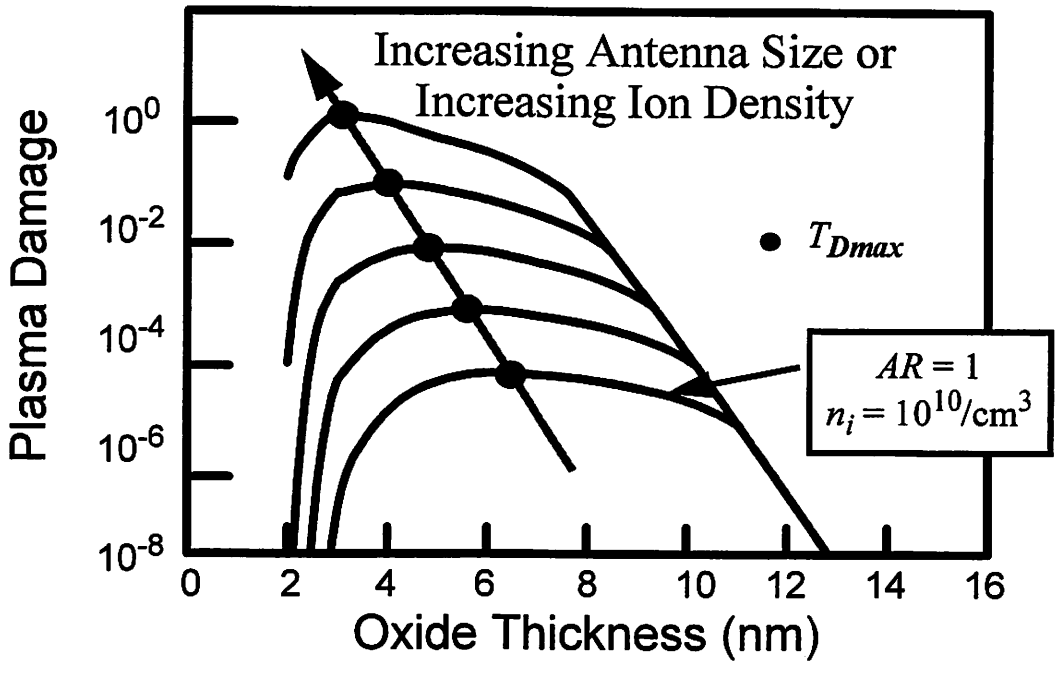

Figure 8-14 Antenna Effect on Plasma Damage

Damage peaks at about 6.0 nm for small antennas. Larger antennas magnify the damage, while shifting T_{Dmax} to thinner oxides. Each line represents a magnitude increases in the product of n_i and *AR*.

ratios up to 1 million. Larger AR's create more damage. Surprisingly, though, the damage curve also shifts T_{dmax}, the oxide thickness which undergoes the most plasma damage, towards thinner oxides. For moderate antennas (~100X), the damage peaks at ~5 nm, with a reduction in damage for ultra-thin oxides. The *AR* data are re-plotted in Figure 8-15 as a function of antenna ratio. For small antennas, the 6.0 nm oxide suffers the most damage. At larger antenna ratios, the thinner oxides sustain more damage. The transition of

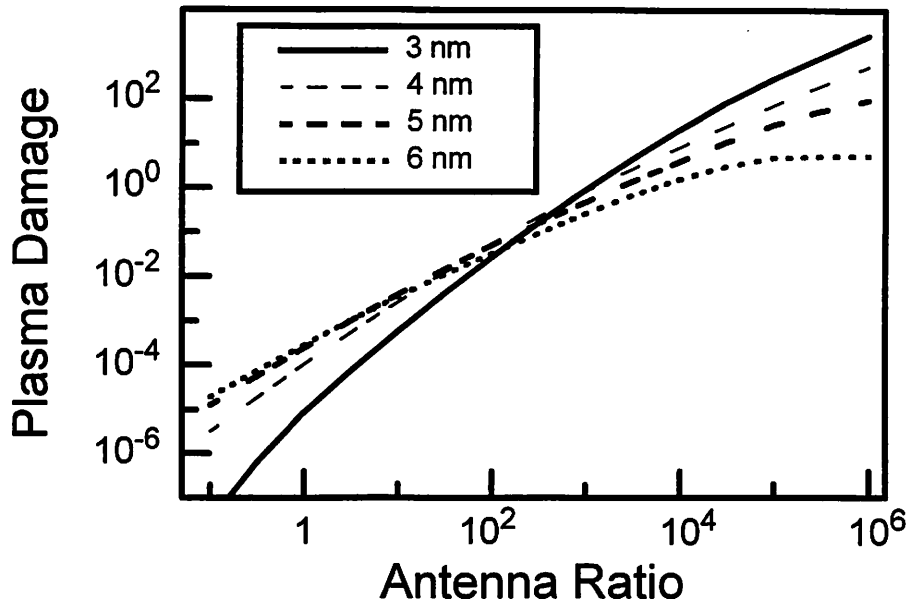

Figure 8-15 Antenna Ratio Effect on T_{Dmax}

The 3.0 nm oxide is least damaged with antennas below 100X, and damaged the most for antennas greater than 10,000X. Extremely large antennas are not good predictors of damage for realistically sized device layouts.

T_{dmax} from thicker oxides to thinner oxides occurs with *AR*s between 100X and 10,000X. With a 1 million AR, T_{dmax} shifts to below 3 nm. Since larger AR's supply more current (by many orders of magnitude), they push the transition to the constant current regime towards thinner oxides. This in turn shifts T_{dmax} to thinner oxides.

Figure 8-15 also reveals a non-linearity between *AR* and damage for ultra-thin oxides. Enlarging the *AR* from 10X to 10,000X enhances the damage for the 3.0 nm oxide by 35,000 (35X greater than proportional). For the same *AR* enlargement, the 6.0 nm oxide damage scales by only 1400. Conventionally, it has been assumed that damage increases proportionally with AR. This assumption is reasonable for thicker oxides (6.0 nm), but extremely inaccurate for ultra-thin oxides. This non-linearity precludes the use of super-large antennas for predicting damage for ultra-thin oxides, since they over-estimate damage that might occur in production design layouts by more than an order of magnitude.

The crossover in T_{Dmax} around 100X - 10000X also resolves the disparate results in the literature. As stated above, some report a reduction in damage for ultra-thin oxides, while others show an increase. The majority of papers utilize large antennas within the cross-over regime. Since the exact crossover point depends on the particular plasma process, small process differences will switch T_{Dmax} from thicker to thinner oxides. Furthermore, large AR data are not directly applicable to real IC layouts in which the area antennas average less than 300 and are seldom above 1000 [8-21].

The crossover and non-linearity effects indicate that utilizing large antennas for detecting plasma damage must be done with care. One should fully understand that large AR results are not easily correlated to actual IC layout AR's.

8.4 Plasma Parameter Effects

For the simulated plasma condition above, T_{Dmax} occurs at ~6 nm. The actual value of T_{dmax} damage peak, though, is a function of the plasma process. It is advantageous to understand how the plasma condition affects the peak location. The following sections analyze the effect of ion density, open circuit voltage, and electron temperature on the magnitude of plasma damage and T_{Dmax}.

8.4.1 Ion Density

Over the past decade there has been a continuous shift toward higher plasma density tools, with ion densities often exceeding $10^{11}/cm^3$. Higher ion densities increase the electron and ion saturation currents, magnifying the potential for plasma damage. Actually, higher ion densities, keeping all else equal, are analogous to larger AR's. The only practical difference is the ability of AR to be scaled by huge amounts, up to 10^6 on a single wafer, while the ion density has a practical range closer to three orders of magnitude. In a similar manner to larger AR's, higher ion densities delay the transition from constant voltage to constant current stress (since the plasma can supply more current), shifting T_{Dmax} to

thinner oxides. In Figure 8-14, each tenfold increase in n_i, with all else equal, shifts T_{Dmax} by about 0.8 nm. Altering the ion density affects the open circuit voltage and the electron temperature, further modifying the damage magnitude and peak.

8.4.2 Open Circuit Voltage

The open circuit voltage generated by the plasma is the main driving force behind charging damage. A variety of factors including non-uniformities and electron shading (governed by Equations (8-2) and (8-3)) determine V_{oc}. Ideally, if V_{oc} equals zero, the damage will necessarily be zero. Altering V_{oc} closer to zero, achieved by adjusting the plasma condition, substrate bias, and reactor geometry, will minimize damage. The sign of the open circuit voltage also affects the maximum amount of damage. With a negative V_{oc}, the plasma stress during processing will be in the electron saturation regime where the current levels are up to 100 times greater than the ion saturation regime (Figure 8-2).

Lowering V_{oc} does not reduce damage evenly for all oxide thicknesses. Figure 8-16 compares the plasma impedance line for two conditions, each with T_e= 2 eV, but with different V_{oc}'s of 5 and 10 V. Halving V_{oc} significantly lowers the stress voltage and current for oxides greater than 6 nm. Consequently, the plasma damage drops dramatically for these oxides. On the other hand, the stress condition does not change greatly for oxides thinner than ~4 nm. The stress voltage is essentially unchanged, while the current drops by at most a factor of two. These oxides are operating in the nearly constant current regime, and are not significantly affected by the drop in V_{oc}. For the shift in V_{oc} to affect damage for these ultra-thin oxides, V_{oc} must be reduced further, such that $V_{oc} < t_{ox} \cdot 10\frac{MV}{cm}$.

This relationship between V_{oc} and oxide thickness may also explain the disparate reports in the literature concerning oxide scaling and plasma damage. Overall, minimizing the open circuit voltage is an effective means of reducing plasma damage, but often must be reduced close to zero volts to affect ultra-thin oxides.

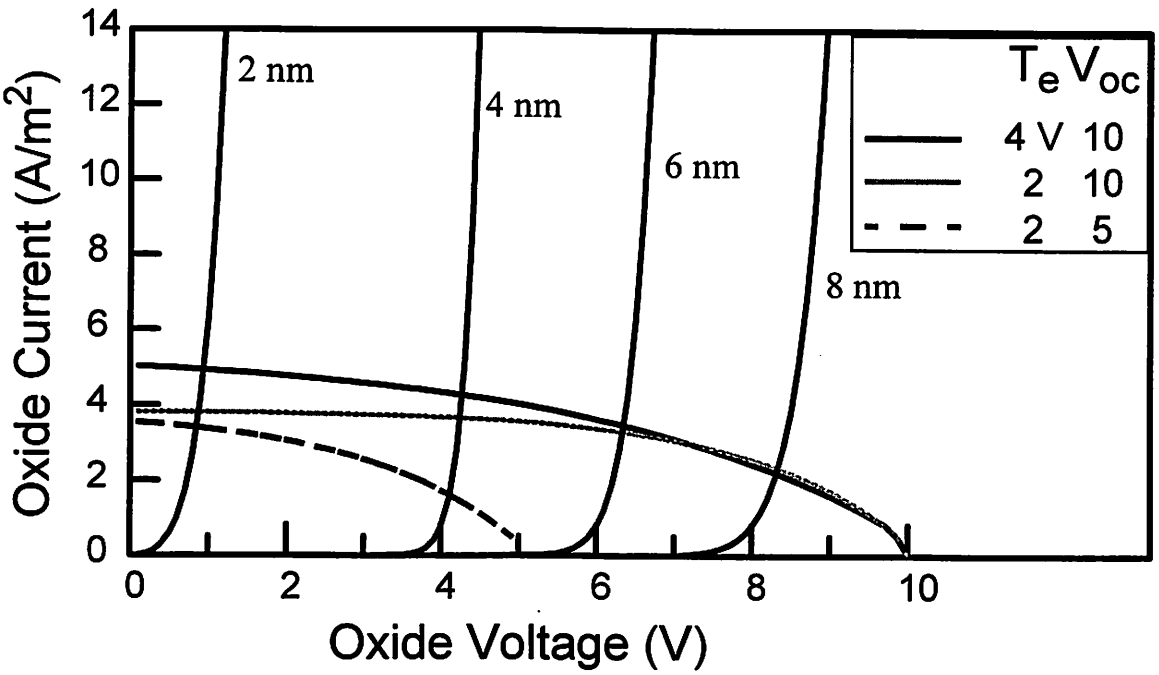

Figure 8-16 Effect of Electron Temperature and Open Circuit Voltage

The plasma V_{oc} and T_e directly control the plasma impedance curve. Halving V_{oc} from 10 to 5 V, reduces I_{sc} by less than 10%. While halving T_e, drops I_{sc} by more than 20%.

8.4.3 Electron Temperature

It is well known that a lower T_e produces less damage. Decreasing T_e abates damage in two ways. First, a smaller T_e, all else equal, reduces the ion flux, by lowering the Bohm velocity (Equation (3-4)). This is apparent by comparing I_{sc} in Figure 8-16 for the two conditions with V_{oc} = 10 V. The halving of T_e cut I_{sc} by more than 20%.

Second, a drop in V_{oc} often accompanies a reduction in T_e. This is an evident in Equations (8-2) and (8-3), in which V_{oc} is proportional to T_e. Consequently, a drop of T_e will proportionally reduce V_{oc}. Moreover, the shadowing factors in Equation (8-3) also depend on T_e. Minimizing T_e is an effective method for reducing damage, since both I_{sc} and V_{oc} are lowered.

8.4.4 Plasma Parameter Implications

The plasma conditions controls the magnitude and peak thickness of plasma damage. The key to reducing damage is to minimize I_{sc} and V_{oc}. Reducing the ion density and the electron temperature minimize damage by directly affecting I_{sc}. Lowering T_e should also proportionally reduce V_{oc}. Reducing V_{oc} is only effective in reducing plasma damage if $V_{oc} < t_{ox} \cdot 10\frac{MV}{cm}$.

8.5 Conclusions

Understanding and preventing plasma damage has been a source of intense research over the past decade. Gate oxide scaling below 5 nm has initiated debates between the relationship between oxide thickness and plasma damage. Along these lines, a procedure is developed to theoretically calculate plasma damage as a function of gate oxide thickness. A load line analysis between the plasma impedance and gate conduction establishes the stress condition during the plasma process. An oxide reliability model (for this work the Anode Hole Injection model) correlates the stress condition with the level of oxide damage. Oxide reliability models reveal that constant voltage stressing damages thinner oxides more, while thin oxides fair better under constant current stressing.

With oxide scaling, the plasma stress transforms from constant voltage stressing to constant current stressing. The simulation identifies the oxide thickness at which the plasma behaves as a constant current source during processing. The simulation reveals three distinct scaling regimes: constant voltage stressing, constant current stressing, and direct tunneling. Simulations with typical plasma parameters indicate that the transformation to constant current stressing is nearly complete at around 5 nm. Hence, plasma damage peaks at this thickness, (T_{Dmax}). Thinner oxides suffer less damage because constant current stressing decreases the oxide voltage. It should be noted that the main driving force that reduces damage in ultra-thin oxides is the plasma constant current source effect, not direct tunneling, as many have purported.

T_{Dmax} is a strong function of antenna ratio. Large ARs shift T_{Dmax} to thinner oxides. Moreover, the damage for ultra-thin oxides increases super-linearly as a function of antenna size. For example, scaling the AR by 1,000X with a 3 nm oxide, raises the damage by 35,000. Both of these effects reveal that large antenna data are difficult to extrapolate to smaller, and more realistic antenna sizes. Extrapolation may lead to erroneous conclusions.

The simulations trends suggest ways to prevent or reduce plasma damage. Most importantly, plasma non-uniformity must kept to a minimum. After that, the key to reducing damage is to lower the plasma V_{oc} and I_{sc}. Reducing I_{sc} is always effective, while V_{oc} must be less than $t_{ox} \cdot 10\frac{MV}{cm}$ to significantly alter the damage. Thus, the requirement on V_{oc} becomes more stringent with thinner oxides. Decreasing the ion density, directly lowers I_{sc} proportionally. Minimizing T_e is quite effective, since a smaller T_e affects both I_{sc} and V_{oc}. A combination of lower ion density and reduced T_e achieves the best results, especially for ultra-thin oxides.

The simulation reveals two reasons for the disparate results in the literature concerning oxide scaling and plasma damage. Many groups report decreasing damage with thinner oxides, while others report the opposite. First, the use of different ARs shifts T_{Dmax}. In addition, a different V_{oc} will alter the threshold thickness for damage. Plasma damage is much less significant if $V_{oc} < t_{ox} \cdot 10\frac{MV}{cm}$. Therefore, groups with different processing conditions will operate with different damage thresholds and T_{Dmax}.

To summarize, the simulations reveal a peak in damage with oxide scaling that shifts with the exact processing conditions. Chapter 10 presents experimental verification of T_{Dmax}, and the shifting of T_{Dmax} with antenna ratio and ion density. Before that, the next chapter, Chapter 9, discusses the various methods for measuring charging damage.

References

[8-1] A. Bayoumi, S. Ma and et. al., "Scalability of Plasma Damage with Gate Oxide Thickness," *2nd International Symposium on Plasma Process-Induced Damage*, pp. 11-14, May 1997.

[8-2] S. Krishnan, W. W. Dostalik, K. Brennan, S. Aur, S. Rangan, and S. Ashok, "Inductively Coupled Plasma (ICP) Metal Etch Damage to 35-60 Å Gate Oxide," *Proceedings of the 1996 International Electron Device Meeting*, pp. 731-734, 1996.

[8-3] K. P. Cheung, C. T. Liu, and et. al., "Charging Damage in Thin Gate-Oxides - Better or Worse," *3rd International Symposium on Plasma Process-Induced Damage*, pp. 34-37, June 1998.

[8-4] D. G. Park and C. Hu, "Plasma Charging Damage on Ultra-thin Oxides," *2nd International Symposium on Plasma Process-Induced Damage*, pp. 15-18, May 1997.

[8-5] M. Alavi, S. Jacobs, S. Ahmed, C. Chern, and P. McGregor, "Effect of MOS Device Scaling On Process Induced Gate Charging," *2nd International Symposium on Plasma Process-Induced Damage*, pp. 7-19, May 1997.

[8-6] B. P. Linder, and N. W. Cheung, "Calculating Plasma Damage as a Function of Gate Oxide Thickness," *3rd International Symposium on Plasma Process-Induced Damage*, pp. 42-45, June 1998.

[8-7] T. Brozek, and T. Dao, "Effect of Device Type and Plasma Process on the Oxide Thickness Dependence of Plasma-Induced Charging Damage", *3rd International Symposium on Plasma Process-Induced Damage*, pp. 46-49, June 1998.

[8-8] K. Noguchi, H. Ohtake, S. Samukawa, and T. Horiuchi, "Suppression of Topography-Dependent Charging Damage to MOS Devices using Pulse-Time-Modulated Plasma," *3rd International Symposium on Plasma Process-Induced Damage*, pp. 176-179, June 1998.

[8-9] K. Hashimoto, F. Shimpuku, A. Hasegawa, Y. Hikosaka, and M. Nakamura, "General Dependence of Electron Shading Damage on the Wafer Bias Power," *2nd International Symposium on Plasma Process-Induced Damage*, pp. 49-52, May 1997.

[8-10] K. Tokashiki, M. Araki, M. Nagase, K. Noguchi, H. Miyamoto, and T. Horiuchi, "Effect of Electron Temperature and Electron Density on Topography Dependent Charging (TDC) Damage in Inductively Coupled Plasma Etching Tool," *3rd International Symposium on Plasma Process-Induced Damage*, pp. 183-186, June 1998.

[8-11] V. Vahedi, N. Benjamin, and A. Perry, "Topographic Dependence of Plasma Charging Induced Device Damage," *2nd International Symposium on Plasma Process-Induced Damage*, pp. 41-44, May 1997.

[8-12] E. C. Jones, B. P. Linder, and N. W. Cheung, "Plasma Immersion Ion Implantation for Electronic Materials," *Japanese Journal of Applied Physics,* Part 1, vol. 35, pp. 1027-1036, Feb. 1996.

[8-13] W. En. B.P. Linder, and N. W. Cheung, "Modeling of Oxide Charging Effects in Plasma Processing," *Journal of Vacuum Science and Technology B*, vol. 14, no. 1, pp. 552-559, 1996.

[8-14] B. P. Linder, W. En, and N. W. Cheung, "The Effect of Wells on Gate Oxide Charging Damage during Plasma Processing," to be published in *Transactions on Electron Devices*, Dec. 1998.

[8-15] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, "A Consistent Model for the Thickness Dependence Of Intrinsic Breakdown in Ultra-thin Oxides," *1995 International Electron Device Meeting*, pp. 343-346, 1995.

[8-16] J. H. Stathis, and D. J. Dimaria, "Reliability Projection for Ultra-thin Oxides at Low Voltages," *Proceedings of the 1998 International Electron Device Meeting*, pp. 167-170, 1998.

[8-17] G. Lewicki and J. Maserjian, "Tunneling in MOS Structures," *Journal of Applied Physics*, vol. 46, no. 7, p. 3032, 1975.

[8-18] B. Eitan and A. Kolodny, "Two components of Tunneling Current in Metal-Oxide-Semiconductor structures," *Applied Physics Letters*, vol. 43, no. 1, p. 106, 1983.

[8-19] K. Scheugraf and C. Hu, "Reliability of Thin $Si0_2$," *Semicond. Science and Technology, vol.* 9, pp. 989-1004, 1994.

[8-20] J. McPherson, V. Reddy, K. Banarjee, and H. Le, "Comparison of E and 1/E TDDB Models for SiO2 under long-term/low field test conditions," *Proceedings of the 1998 International Electron Device Meeting*, pp. 171-174, 1998.

[8-21] J. Luchies, P. Simon, F. Kuper, and W. Maly, "Relation Between Product Yield and Plasma Process Induced Damage," *3rd International Symposium on Plasma Process-Induced Damage*, pp. 7-10, 1998.

9 Charging Damage Measurement Techniques

9.1 Introduction

A number of methods have been developed to quantify gate oxide plasma charging damage. The ideal method should be easy to execute, simple to analyze and interpret, exhibit a monotonic correlation with oxide damage, highly sensitive to damage with a large dynamic range without saturation, and provide a fast turnaround.

All of the measurement techniques discussed in this chapter fall into three general categories: oxide integrity, transistor performance degradation, and non-MOS devices. Within each of these categories are a number of methods which are briefly describe in the following sections. No one method satisfies all of the desirable characteristics, requiring compromise.

9.2 Oxide Integrity Measurements

Direct measurements of the oxide degradation due to plasma charging damage provides the best correlation with actual oxide damage. Moreover, they generally only require MOS capacitors, and not fully processed transistors. There are three sub-categories of oxide integrity measurements: lifetime, interface traps, and leakage current.

9.2.1 Oxide Lifetime Measurements

The oxide lifetime may be determined by Charge-to-Breakdown (Q_{bd}), Time-Dependent Dielectric Breakdown (TDDB), or Voltage Breakdown (V_{BD}). Q_{bd} measures

cumulative tunneling charge under constant current stress before breakdown. Subtracting the damaged Q_{bd} from the virgin breakdown value yields the plasma stress. TDDB is similar to Q_{bd}, but utilizes a constant voltage stress. Finally, V_{bd} measurements apply a fast ramp voltage stress until breakdown.

Breakdown is an intrinsically statistical quantity relying on the random distribution of oxide defects to form a conduction path. Therefore, large statistical variations plague these methods, and their destructive nature preclude further evaluation of the oxide.

9.2.2 Oxide Interface Trap Measurements

An alternative to destructive oxide lifetime testing, are non-destructive interface traps measurements. During oxide stress, interface traps accumulate, altering the capacitance of the MOS system. These traps are detectable with *C-V* or conductance measurements. *C-V* measurements quantify the change in capacitance as a function of voltage, while the conductance method monitors the frequency response of the oxide. For good resolution, both methods usually require relatively large MOS devices of at least 50μm x 50μm. The conductance method is generally more accurate, but *C-V* measurements provide a faster turnaround time. Therefore, this work utilizes *C-V* measurements. Section 9.7 describes the method in more detail.

9.2.3 Leakage Current Measurements

In addition to creating interface traps, plasma charging damage creates bulk traps. These bulk traps provide "hopping" sites for electrons from the cathode to the anode. This increases the low-field leakage currents, called Stress Induced Leakage Current (SILC). SILC increases with the number of bulk traps, and can analyze oxide integrity in a wide range of MOS device areas from less than 5 μm^2 to greater than 400 μm^2. Besides creating SILC currents, the bulk traps also increase the $1/f$ gate conduction noise, providing another parameter to analyze.

The main drawback with SILC measurements are that the extra leakage currents are difficult to detect outside the range of 3 nm to 8 nm [9-7]. SILC is below detectable limits (10^{-15} A) in oxides thicker than this. While oxides thinner than 3 nm suffer from large direct tunneling currents that swamp out the SILC. Even so, SILC is an excellent characterization method for plasma damage, and is utilized in this work. Section 9.7 describes this technique in more detail.

9.3 Transistor Performance Degradation Measurements

Monitoring the transistor performance degradation is an alternative to measuring the oxide degradation directly. The bulk and interface traps generated by oxide damage shift the threshold voltage (V_t) of the MOS capacitor/transistor. The V_t shift correlates with the plasma damage. The interpretation is sometimes difficult since the shift depends on the charge of the traps created and hence V_t does not always show a monotonic degradation with plasma damage [9-1]. Other parameters that show DC performance degradation with plasma damage are transconductance, inversion carrier mobility, and subthreshold slope. Plasma damage also corrupts the AC transistor performance. Drain current noise power ($1/f$ noise), and random telegraph signal (RTS) both reveal plasma damage well, and are especially useful for analog CMOS applications [9-2].

Transistor degradation is measurable in transistors of all sizes, a major advantage over oxide integrity measurements. Moreover, the extraction of the transistor characteristics is routine, especially for the IC industry, and is easy to interpret. The downside is that they require fully processed transistors, and the results are not direct measurements of the oxide reliability.

9.4 Alternative Non-MOS Measurements

To assess the plasma charging damage capability of a tool, both oxide integrity and transistor performance measurements require destructive exposure of fully processed

wafers. In an effort to reduce cost and decrease turnaround time, several re-usable wafer techniques with alternative structures have been developed.

Most popular, are CHARM-2 wafers, with EEPROM devices [9-3]. The voltage stress from the plasma programs the EEPROM, storing the damage as a threshold voltage shift. By adding resistors in parallel to the EEPROM devices, a two dimensional map of the plasma impedance as a function of voltage may be determined. After each use, the EEPROMs may be erased, and re-used. This is the fundamental advantage of the CHARM-2 wafers. However, since CHARM-2 is an indirect measurement of plasma damage, the results may not correlate well to product MOS devices. Moreover, EEPROM devices are programmed within milliseconds, and therefore measure short transients which may not result in damage.

Another emerging technology is Surface Potential Measurement (SPM) or Contact Potential Measurement (CPM). This method utilizes a silicon wafer with 100 nm thermal oxide. After processing, the SPM measures the residual charge, producing a voltage map across the wafer. Although some report good correlation with plasma damage [9-4], others question the usefulness of the method, showing that even water cleans "create" damage with SPM [9-5].

In an attempt to provide in-situ monitoring of charging, cantilever MEMS structures have been devised [9-6]. The cantilever structure deflects towards the wafer as it charges. A laser measures the deflection angle, which is correlated with charging voltage. Even though the technique is elegant and provide in-situ charge-up data, the need for correlation to actual devices structures precludes anything but academic exploration.

In summary, non-MOS re-usable devices may have their niche applications, but are plagued by doubts of their correlation to actual product devices.

9.5 Gate Oxide Thickness Dependence

Over the last decade, gate oxides have scaled from greater than 10 nm to 3 nm and below. Many of the traditional methods for detecting plasma damage are no longer effective with oxides thinner than 5 nm. The methods for thicker oxides typically detect the presence and effects of oxide traps. The defect density at breakdown is a function of t_{ox}, with 3 nm oxides containing 10^{-5} cm^{-2} less traps than 6 nm oxides at breakdown [9-7]. Therefore, all methods that rely on traps for detecting damage will be less useful for ultra-thin oxides. Even if there are enough traps to measure, the trapped charges de-trap easily in ultra-thin oxides. The excessive leakage currents render *C-V* methods inappropriate for oxides much thinner than 4 nm.

The main method that is suitable for plasma damage measurement in ultra-thin oxides is SILC, which relies on the excessive gate conduction as the damage monitor. SILC is challenging to measure in oxides thinner than 3 nm, but with careful technique detects damage in oxides as thin as 2.2 nm [9-7]. Some groups report that $1/f$ gate conduction noise detects damage and soft breakdown [9-8] in oxides thinner than 2 nm [9-9].

9.6 Post-Anneal Reveal Stress

After plasma processing, the damage may be passivated with a hydrogen or forming gas anneal. Hydrogen passivation is temporal, and may be liberated with F-N or channel hot carrier stress. Approximately 70% of the original traps re-form with nominal additional stress [9-10]. If the wafers are subjected to any annealing conditions after the plasma process, they must undergo a post-anneal damage reveal stress before characterization. For this work, the wafers are never subjected to temperatures above 25 °C after plasma exposure, and therefore do not require a reveal stress step.

9.7 *C-V* and SILC Measurement Procedures

This work utilizes *C-V* interface trap extraction for oxides thicker than 5 nm and SILC measurements for thinner oxides. The following sections detail the measurement procedures.

9.7.1 MOS Capacitance

The MOS capacitance system model, including the effect of interface traps, is depicted in Figure 9-1. The small signal capacitance of the system is determined by where

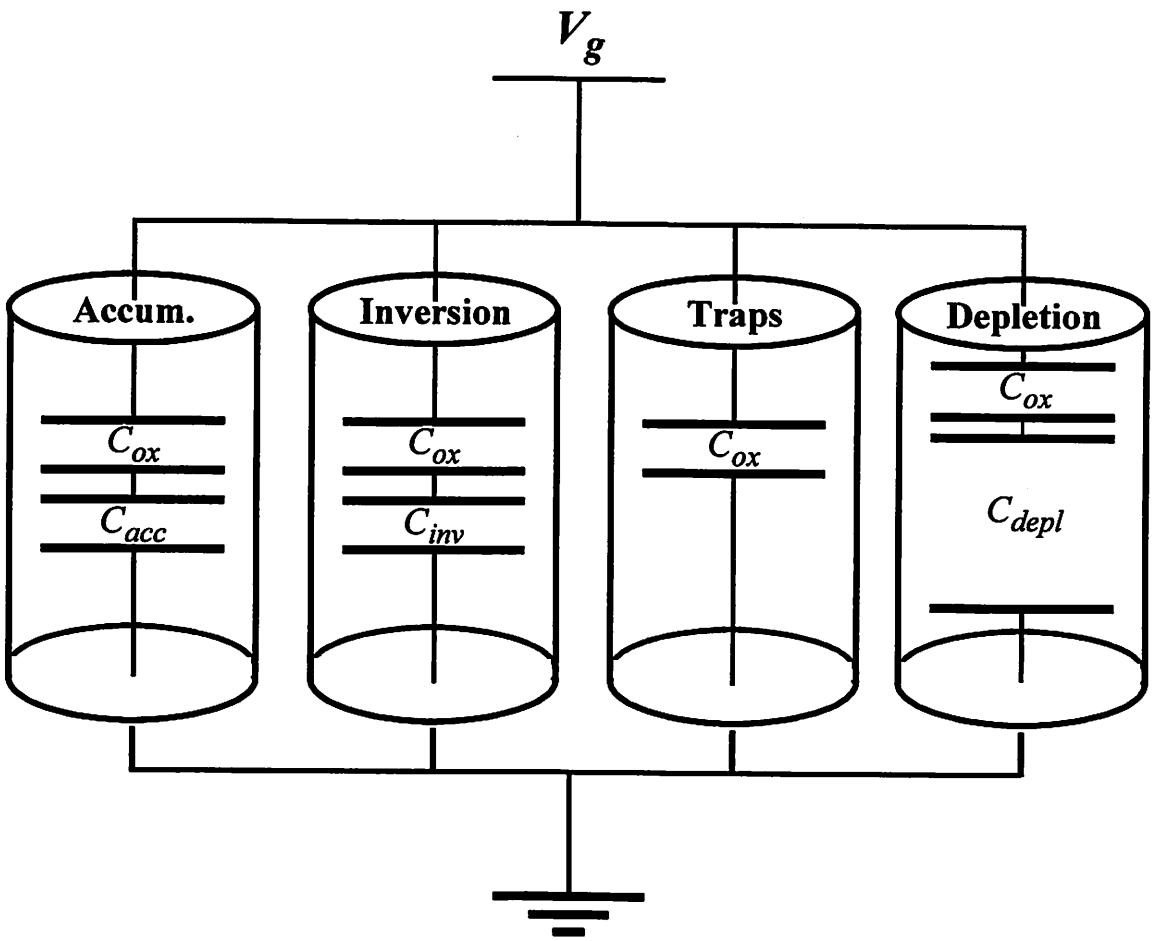

Figure 9-1 The MOS *C-V* System Model

The total capacitance of a MOS system is composed of four different parallel contributors. The small signal system capacitance is computed by keeping track of which barrel the new charge is stored.

the electric fields lines terminate, or by the location of the modulating charge. Generally, the charge is stored in either the inversion layer, accumulation layer, depletion layer, or in interface traps. When the gate voltage changes by ΔV, the system capacitance is determined by which bin the extra charge accumulates. If charge forms in either the inversion layer or the accumulation layer, the system capacitance is equal to C_{ox} in series with the inversion capacitance (C_{inv}) or accumulation capacitance (C_{acc}), respectively. If the charge fills an interface trap, the system capacitance is equal to C_{ox}. If the charge is stored in the depletion region, the capacitance is the series combination of the oxide capacitance and the depletion capacitance. The total system capacitance consists of the sum of the contribution of the four parallel capacitors. For example, if the additional charge is split between interface traps and the depletion region, the total capacitance is $0.5C_{ox} + 0.5\left(\frac{C_{ox} \cdot C_{depl}}{C_{ox} + C_{depl}}\right)$.

9.7.2 Capacitance Measurements

There are two main methods for measuring the capacitance of a MOS system: quasi-static and high-frequency. The frequency dependence of a MOS system arises from the frequency response of inversion charge generation and interface trap filling and emptying. Inversion charge requires milliseconds or more to generate, while depletion charge storage is nearly instantaneous. A high frequency sweep does not generate inversion layer charge, but only modulates depletion charge. Therefore, at high frequencies the depletion region determines the capacitance, and not the inversion region. Along the same lines, interface traps have significant time constants, and do not respond to fast (Mhz) signals. All capacitances contribute to the quasi-static (low frequency) measurement, while only accumulation and depletion charge respond to the high frequency measurement.

9.7.2.1 Quasi-static Measurement

The Quasi-static (QS) measurement ramps the gate voltage to determine the *C-V* of the test structure. Neglecting leakage, the current is proportional to the voltage ramp rate.

$$I = C \cdot \frac{dV}{dt} \tag{9-1}$$

where I, C and dV/dt are the current, capacitance, and ramp rate, respectively. The ramp rate is kept low (< 0.1 V/s) assuring system equilibrium. A sample QS measurement is shown in Figure 9-2.

9.7.2.2 High Frequency Measurement

During a high frequency (HF) measurement, a small amplitude signal (~ 0.026 V) is applied on top of a DC bias. Since high frequencies prevent the inversion and interface trap regions from responding, HF *C-V* only measures the capacitance of the depletion and accumulation region. This usually requires signal rates exceeding 1 MHz. Stepping the voltage, while allowing ample time for carrier equilibrium, produces a full *C-V* curve. A sample HF measurement is shown in Figure 9-2.

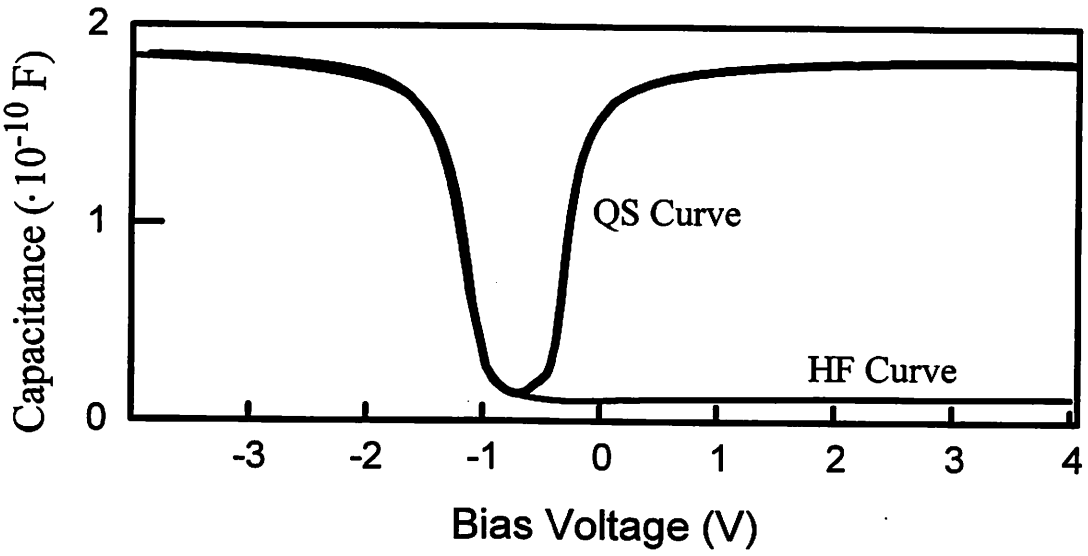

Figure 9-2 Quasi-Static and High Frequency *C-V* curves for an Undamaged Sample

Representative QS and HF *C-V* curves for interface trap extraction. The high frequency curve remains low at the positive voltages because the inversion layer charge can not respond. This is a measurement for a n-MOS structure.

9.7.3 Interface Trap Extraction

There are four main methods for extracting the interface trap density (D_{it}) from the *C-V* measurements. The first two compare either the theoretical QS or HF curve and the corresponding measured curve. The difference between the capacitances is assumed to be due to oxide traps. This method assumes apriori knowledge of the depletion capacitance, which is a function of the doping density underneath the gate oxide. Any errors in the pre-supposed doping density will unacceptably propagate through to the extracted interface trap density. The two other methods compare only measured curves, and do not make any assumptions of the doping density.

9.7.3.1 Quasi-Static and High Frequency Comparison

The most common extraction method compares the measured quasi-static and high frequency *C-V* curves, eliminating many of the errors associated with the use of theoretical *C-V* curves. The QS measurement allows all of the interface traps to affect the capacitance, while the HF measurement operates at frequencies above the interface trap response rate. By manipulating the capacitance relations for the two measurements, the interface trap capacitance as a function of voltage is solved:

$$C_{it}(V_g) = \left\{\left(\frac{1}{C_{qs}(V_g)} - \frac{1}{C_{ox}}\right)^{-1} - \left(\frac{1}{C_{hf}(V_g)} - \frac{1}{C_{ox}}\right)^{-1}\right\} \tag{9-2}$$

where C_{it}, C_{qs}, C_{ox}, and C_{hf} are the interface trap capacitance, the quasi-static CV capacitance, the oxide capacitance, and high frequency capacitance at a particular gate bias [9-11]. With the interface capacitance calculated, the interface trap density as a function of V_g per eV is simply:

$$D_{it}(V_g) = \frac{C_{it}(V_g)}{q \cdot A} \tag{9-3}$$

where A is the capacitor area. Typically Equation 9-3 is integrated over a region in the bandgap yielding the aerial interface trap density. This requires the transformation of V_g into Φ_s, the potential at the oxide-silicon interface. This is accomplished by Berglund's Method [9-12].

$$\Phi_s = \int_{V_{fb}}^{V_g} \left(1 - \frac{C_{qs}}{C_{ox}}\right) dV \tag{9-4}$$

where V_{fb} is the flatband voltage. Finally, the transformed D_{it} is integrated across the bandgap as shown in Figure 9-3. Extraction error near the band edges dramatically increases the measured D_{it}. Hence, it is customary to confine the integration to the midgap region. The silicon bandgap is ~1.12 eV, so it is customary to integrate symmetrically around 0.56 eV, the midgap. The extracted interface trap density necessarily depends on the range of integration, and therefore all extractions must have the same limits for comparison's sake.

9.7.3.2 Quasi-Static Only Comparisons

Small stray capacitances severely affect the HF *C-V* measurement. Choosing a frequency such that none of the traps respond, but which is low enough so that the stray capacitances do not dominate, proves difficult. Another method avoids the difficulties of the HF measurement by comparing before and after stress QS curves. Any generated interface traps will increase the capacitance at a given Φ_s for the QS measurement. Integrating the *C-V* deviation yields the cumulative change in interface trap density. With this method, before and after QS curves are transformed from a function of V_g to a function of Φ_s as in the previous section. Then solving for D_{it}

$$D_{it}(\Phi_s) = \frac{\left\{\left(\frac{C_{qs2} \cdot C_{ox}}{C_{ox} - C_{qs2}}\right) - \left(\frac{C_{qs1} \cdot C_{ox}}{C_{ox} - C_{qs1}}\right)\right\}}{q \cdot A} \tag{9-5}$$

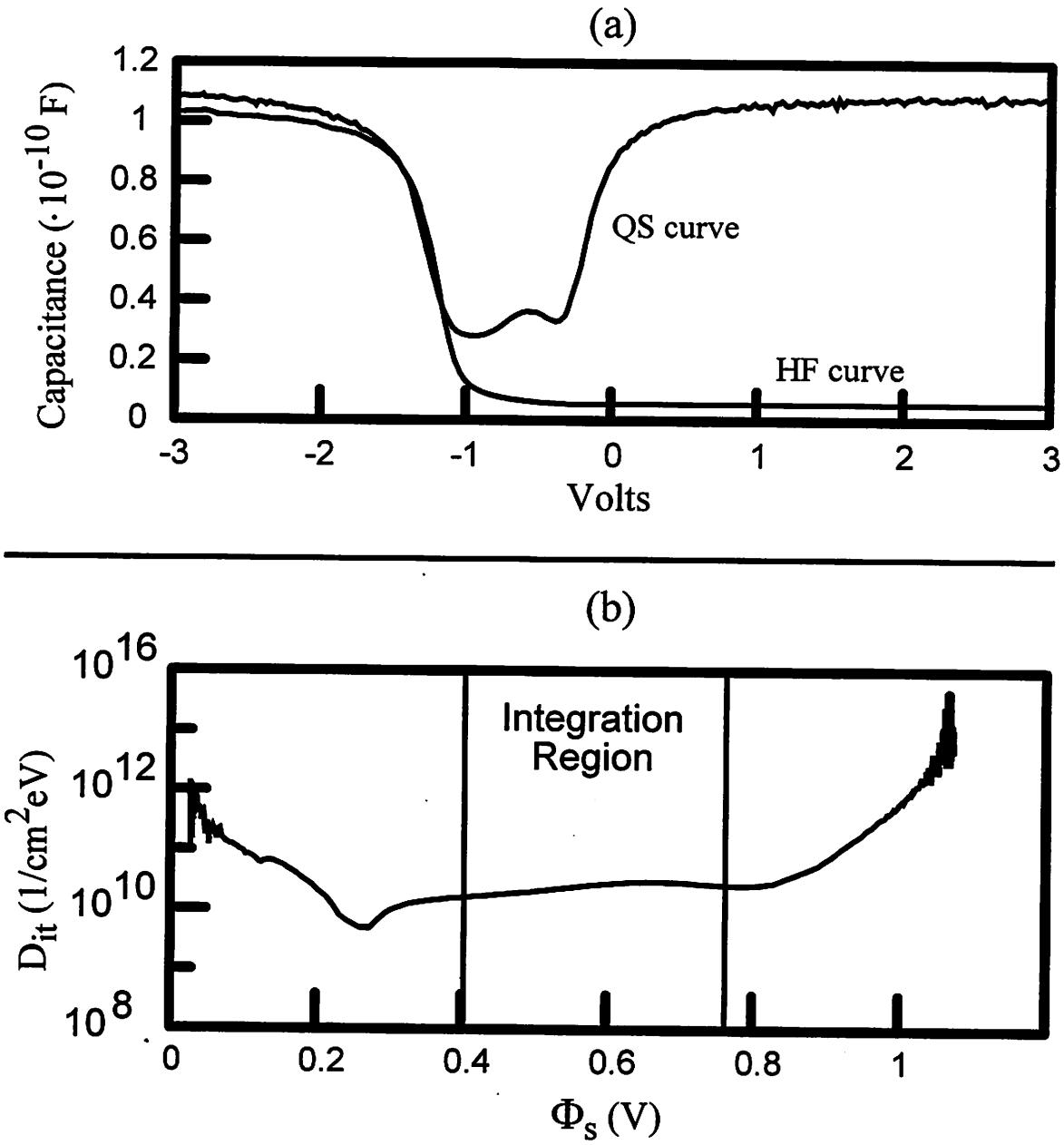

Figure 9-3 *C-V* **Extraction of Interface Traps**

(a) "Damaged" QS and HF curves from the same capacitor. (b) extracted interface trap density as a function of the semiconductor surface potential (Φ_s). Integrating over the mid-gap yields interface traps per cm^2. Integrating from 0.4 to 0.72 and dividing by an area factor yields a interface trap density of $1.94 \cdot 10^{11} cm^{-2}$ for this sample.

where C_{qs1} and C_{qs2} are the before and after stress QS *C-V* measurements, respectively. As in the last section, the total interface trap value is an integration of $D_{it}(\Phi_s)$ over the midgap region.

The only drawback of this technique is the need for before and after stress measurements. This contrasts with the QS/HF technique which requires only after stress measurements. Because of the increased ease of interface trap extraction, the Quasi-Static Only method is employed throughout this paper.

9.7.4 Measurement Technique and Errors in Interface Trap Extraction

To determine the amount of oxide damage, the capacitance measurements must be executed with extreme care. Any errors in the capacitance values tend to be magnified by the extraction methods. There are two broad categories of errors, those concerned with the measurement set-up and those physically inherent in the technique.

9.7.4.1 Measurement Conditions

The Quasi-Static capacitance measurements were made with an HP 4140B pico-ammeter. The 4140B features a constant ramping of the output voltage, a requirement for the Quasi-Static measurement. With a constant voltage ramp rate, the capacitance is simply the current divided by the ramp rate (Equation (9-1)). The ramp rate is user controllable from 0.01 V/sec to 0.1 V/sec. Faster ramp rates average out the noise, but a slow ramp rate is necessary to guarantee that the MOS system is in equilibrium, a requirement of the method. Furthermore, the ramp rate must be slow enough for the ammeter to change scales near the onset of inversion, when the current may change by an order of magnitude. The ramp rate for the capacitance measurements in this work was 0.03 V/s, a compromise which yields low noise and reasonably maintains thermal equilibrium. To further ensure equilibrium, the MOS capacitor is ramped from inversion to accumulation, eliminating minority carrier generation from the measurement. Nitrogen gas flowing across the capacitor reduces moisture, minimizing leakage currents. Leakage was monitored before

each voltage sweep by observing the DC current at a constant applied bias. Ideally the current should be zero, and for low leakage situations is less than 10 fA. During the sweep, 250 data points were taken, with intermediate values calculated by simple linear interpolation. Under proper conditions, the accuracy of Quasi-Static measurements are near 1%.

The high frequency measurements were made with an HP 4192 Impedance Meter. The main user parameter for HF measurements is the frequency. The frequency must be high enough so that the inversion layer and the interface traps can not respond to the small-signal oscillation. For the measurements, the frequency was set at 1 Mhz, which was a compromise between the inversion layer generation rate, and the limit dictated by stray capacitances (that dominate above 10 MHz). To properly exploit the internal compensation for the coaxial line reflections, the wires must be exactly 1 meter long. Under proper measurement conditions, typical measurement HF C-V errors are on the order of 1%, or 1 pF, whichever is larger.

9.7.4.2 Extraction Errors

With the accuracy levels of the HF and QS CV measurements, the authenticity of the interface trap extraction is limited by the intrinsic error in the extraction calculations. Nicollian and Brews [9-13] discuss these errors in detail, and they will be summarized here. First, assumption that a 1 MHz HF measurement is a true high frequency measurement creates error. Some interface traps will respond at 1 MHz, especially near flatband, where the trap capture time is the most rapid [9-13]. With a $10^{15}/cm^3$ doped substrate, errors in excess of 10% occur from approximately flatband to 0.1 V away from flatband. Higher doped substrates lead to more error, with a $10^{18}/cm^3$ doped substrate inducing 10% errors up to 0.25V from flatband.

With the Quasi-Static measurement, the onset of inversion translates to errors in the interface trap density. The inversion layer generation (and its associated capacitance)

will be attributed to interface traps, artificially increasing the interface trap value. For a $10^{15}/cm^3$ doped substrate, an additional value of $10^{10}cm^{-2}/eV$ will be added 0.8 eV into the bandgap. Higher substrate dopings decrease the error by delaying the onset of inversion. In a $10^{18}/cm^3$ doped substrate, a $10^{10}cm^{-2}/eV$ error arises 0. 9 eV into the bandgap.

Another source of error, round-off error, occurs with the calculation of the reciprocal of the difference of two nearly equal numbers in Equation (9-2) and Equation (9-5). When either the quasi-static or high frequency capacitances are close to the oxide capacitance, the measurement errors will be magnified considerably. The 10% error level for $10^{15}/cm^3$ doped substrates with a 10 nm gate oxide, occur nearer than 0.1V away from flatband. Measurements further from flatband than this are more accurate since the measured capacitance is substantially lower than the oxide capacitance. Thicker oxides and higher doped substrates worsen this effect, and a 100 nm gate oxide with a $10^{18}/cm^3$ doped substrate will have 10% errors until 0.45V away from flatband. Although lower substrate dopings reduce the round-off error, the large series resistance may introduce other errors. With the high frequency measurement, an additional series resistance may translate into a significantly lower capacitance at flatband. This may, however, be compensated if the series resistance value is known.

Generally, the interface traps are integrated near midgap, and therefore, the errors should be less than 10%, as long as the integration level does not extend close to flatband or the onset of inversion. This becomes difficult for thick oxides with high substrate doping, which might preclude the use of capacitance techniques for D_{it} extraction.

Round-off errors also limit the overall sensitivity of the capacitance extraction technique. With a doping level of $10^{15}/cm^3$ and a C_{LF} measurement accuracy of 1%, the minimum extractable interface trap density near midgap is $10^{10}cm^{-2}/eV$. This increases to $3 \cdot 10^{11}cm^{-2}/eV$ with a $10^{18}/cm^3$ substrate doping.

For the interface trap extraction done in this chapter, the minimum sensitivity is approximately $10^{10}cm^{-2}/eV$. Since the integration range is 0.32V, the minimal detectable density is $3.2 \cdot 10^{9}/cm^{2}$. For higher values of interface traps, the accuracy is expected to be better than 10%. With before and after stressing quasi-static measurements, the minimum sensitivity is not as limiting as the accuracy, since the before capacitance usually has trap densities near (or even above) the minimum sensitivity.

9.7.5 Stress Induced Leakage Measurements

Measuring the low-field gate leakage current is the preferred method for damage detection in oxides thinner than 5 nm. SILC is much more straight forward than *C-V* interface extraction, with no data transformations necessary. Comparing SILC data for different oxides, though, is complicated by the varying intrinsic leakage currents. SILC is a relatively new measurement, and physical models and data interpretation are still evolving.

9.7.5.1 SILC Mechanism

The most accepted mechanism for SILC is trap assisted tunneling. Figure 9-4 depicts a model for the SILC mechanism. Neutral bulk traps act as hopping points for the electron. Since tunneling rates depend on tunneling distance, there is a higher probability for an electron to tunnel two short distances than one long distance. The SILC current is always positively correlated to damage, with the leakage proportional to the trap density [9-14, 9-15]. Figure 9-5 shows the leakage current for a 4.5 nm oxide for increasing levels of oxide stress. Each increase in damage produces larger SILC currents.

9.7.5.2 SILC and Gate Oxide Scaling

The amount of detectable SILC is a complicated function of gate oxide thickness. The total conduction through the oxide is the summation of all the parallel conduction paths: 1) F-N and D-T tunnelling, 2) Single trap conduction, and 3) Multiple trap conduc-

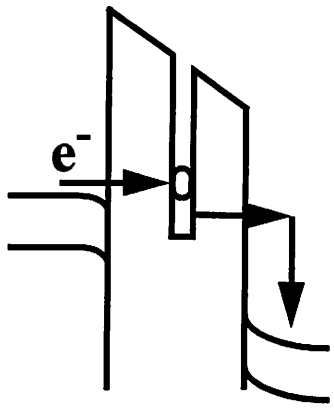

Figure 9-4 Stress Induced Leakage Mechanism

Oxide stress creates neutral bulk traps, that act as hopping points for tunneling electrons. This increases low-field oxide conduction.

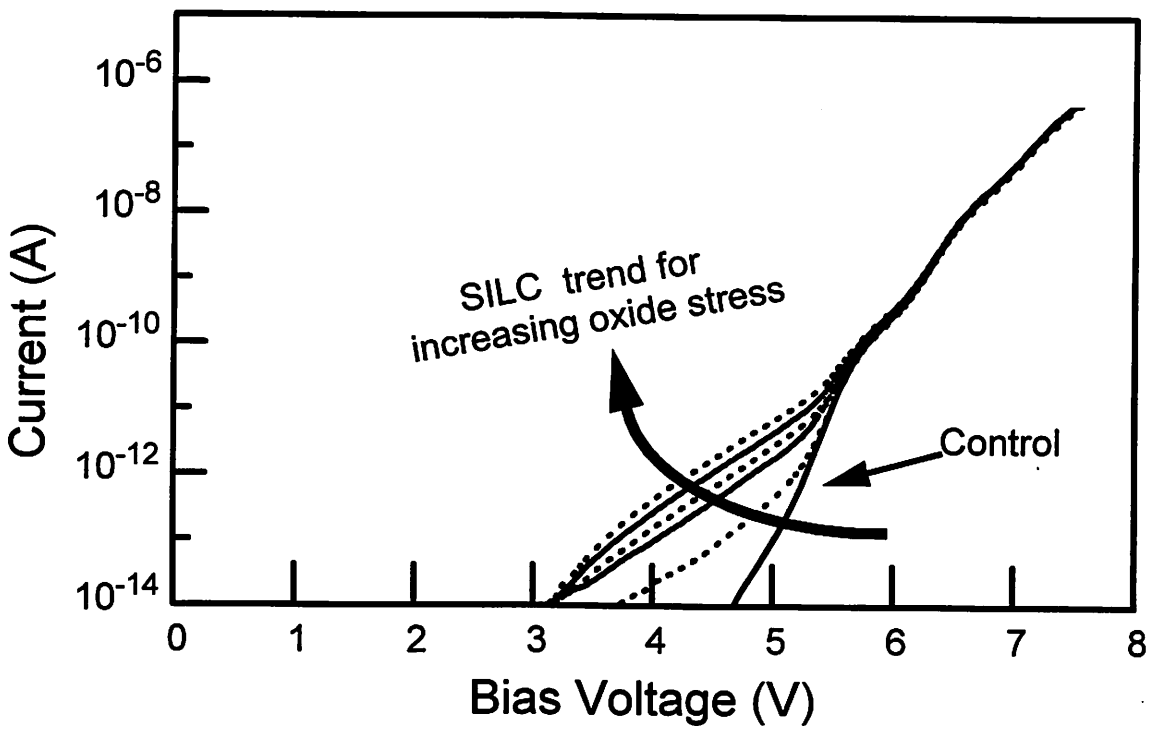

Figure 9-5 Stress Induced Leakage Current Trend

SILC measurements with increasing oxide stress for a 4.5 nm oxide of area 20x5 μm^2.

tion. The ratio of the trap conduction currents to the intrinsic tunneling mechanisms determines the ease of SILC detection. For the same trap density, thinner oxides have a higher trap leakage current. SILC is difficult to detect in oxides thicker than ~8 nm [9-16]. For ultra-thin oxides less than 3 nm, the trap density necessary for breakdown is drastically reduced. Without a large trap density, SILC current never becomes detectable. Recent data show that SILC provides approximately only 1% extra current at breakdown for oxides thinner than 3 nm [9-7]. Therefore, very careful data collection with minimal noise and leakage is necessary for SILC measurements with ultra-thin oxides. Overall, SILC is an ideal measurement for oxides between 3 nm and 7 nm, and has reduced capabilities for other oxide thicknesses.

9.7.5.3 SILC Sense Voltage

To compare different devices, the SILC I-V measurements must be consolidated into a single value. Generally, the leakage is compared at a single voltage, V_{sense}. V_{sense} should be optimized, such that SILC current is large compared to the intrinsic leakage currents, and the current level allows a high level of resolution. There are two methods for adjusting V_{sense} for different oxide thicknesses. The most popular is constant E-Field scaling, adjusting V_{sense} such that E_{ox} is constant for all t_{ox}. This, however, results in non-optimized SILC measurements for thinner oxides. Since thinner oxides have larger intrinsic leakage currents for the same E_{ox}, the V_{sense} should be lowered. Along these lines, the second method selects V_{sense} such that the intrinsic leakage current is the same for all t_{ox}. With this method, the V_{sense} produces a lower E_{ox} in thinner oxides. Since this work compares different t_{ox}, the second method is preferred.

9.7.5.4 Correlation with Circuit Reliability

For SILC current to be an effective measure of plasma damage, it must correlate well with circuit reliability. For sub 0.1 μm scaling, oxide performance may either be leakage limited [9-17] or breakdown limited [9-7]. Fortuitously, SILC measurements

directly determine oxide yield in the leakage limited regime. For the breakdown limited regime, recent modeling has shown a good correlation between SILC and predicted life-time of the device. Therefore, SILC, in addition to being a sensitive measure of plasma damage, may be directly correlated to circuit and device reliability.

9.7.5.5 Measurement Conditions

The SILC measurements were made with an HP 4140B pico-ammeter, with a minimum current sensitivity down to 10^{-15} A. With nitrogen flowing across the test die, background leakage could be suppressed to the detection limit, with random noise below $2x10^{-15}$. To ensure low-noise, the integration time was set to long. The applied voltage was stepped in 0.1 V increments, with a five second delay preceding measurement. This ensured that displacement currents for the pad and oxide were below the 10^{-15} A detection limit. Larger area devices would require a longer delay time. If the measurements are to be repeated, the maximum measurement voltage must be kept as low as practical, so as not to stress the device.

9.8 Conclusion

There are numerous methods for detecting plasma damage. They generally fall into three main categories: oxide integrity, transistor performance, and non-MOS methods. Oxide integrity measurements directly determine oxide plasma charging damage, while transistor performance methods are easier to interpret and automate. Non-MOS methods offer attractive features, such as re-usability, non-full flow devices, or in-situ monitoring, but may not correlate well with damage to MOS devices. As devices scale below 5 nm, many of the traditional methods are no longer effective. The most popular detection methods for ultra-thin oxides are Stress Induced Leakage Current and gate conductance $1/f$ noise. This work uses *C-V* interface extraction for oxides thicker than 5 nm, and SILC for thinner oxides.

References

[9-1] C. T. Gabriel, and J. L. Educato, "Application of Damage Measurement Techniques to a Study of Antenna Structure Charging," *2nd Plasma Process Induced Damage Conference*, pp. 92-95, 1997.

[9-2] Z. J. Ma, H. Shin, P. K. Ko, and C. Hu, "Effects of Plasma Charging Damage on the Noise Performance of Thin-Oxide MOSFETs," *IEEE Electron Device Letters*, vol. 15, no. 6, pp. 224-226, June, 1994.

[9-3] W. Lukaszek, "Quantifying Wafer Charging Damage during Via Etch," *1st Plasma Process Induced Damage Conference*, pp. 30-33, 1996.

[9-4] K. Nauka, J. Theil, J. Lagowski, L. Jastzebski, and S. Sawtchouk, "A Complete Approach to the In-line Monitoring of Materials Defects Introduced in Dielectric and Si by Plasma Processing," *2nd Plasma Process Induced Damage Conference*, pp. 127-130, 1997.

[9-5] K. P. Cheung, J. I. Colonell, K.G. Steiner, et. al., "Is Surface Potential Measurement (SPM) a Useful Charging Damage Measurement Method?" *3rd Plasma Process Induced Damage Conference*, pp. 18-21, 1998.

[9-6] K. Pangal, S. L. Firebaugh, and J. C. Sturm, "Non-invasive Measurement of Charging in Plasmas using Microelectromechanical Charge-sensing Devices," *Applied Physics Letters*, vol. 69, no.10, pp. 1471-1473, September, 1996.

[9-7] J. H. Stathis, and D. J. Dimaria, "Reliability Projection for Ultra-thin Oxides at Low Voltages," Proceedings of the 1998 *International Electron Device Meeting*, pp. 167-170, 1998.

[9-8] M. Depas, T. Nigam, and M. M. Heyns, "Soft Breakdown of Ultra-Thin Gate Oxide Layers," *IEEE Transactions on Electron Devices*, vol. 43, no. 9, pp. 1499-1504, September, 1996.

[9-9] B. E. Weir, P. J. Silverman, D. Monroe, et. al., "Ultra-thin Gate Dielectrics: They Break Down, But Do They Fail?" 1997 *International Electron Device Meeting*, pp. 73-76, 1997.

[9-10] H. Shin, "Thin Oxide Damage by Plasma Processing," Ph.D. Thesis, University of California at Berkeley, pp. 78-79, 1993.

[9-11] E. H. Nicollian, and J. R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc., New York, pp. 332, 1982.

[9-12] C. N. Berglund, "Surface States at Steam Grown Silicon-Silicon Dioxide Interfaces", *IEEE Transactions on Electron Devices*, vol. 13, pp.701-710, 1966.

[9-13] E.H. Nicollian, and J.R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc, New York, pp. 319-372, 1982.

[9-14] S. Kamohara, D. G. Park, and C. Hu, "Deep-Trap SILC (Stress Induced Leakage Current) Model for Nominal and Weak Oxides," *36th Annual International Reliability Physics Symposium*, pp. 57-61, 1998.

[9-15] G. B. Alers, B. E. Weir, M. A. Alam, G. L. Timp, and T. Sorch, "Trap Assisted Tunnelling as a Mechanism of Degradation and Noise in 2-5 nm Oxides," *36th Annual International Reliability Physics Symposium*, pp. 76-79, 1998.

[9-16] D. G. Park, M. Kennard, Y. Melaku, N. Benjamin, T. J. King, and C. Hu, "Stress-Induced Leakage Current due to Charging Damage: Gate Oxide Thickness and Gate Poly-Si Condition Dependence," *3rd Plasma Process Induced Damage Conference*, pp. 56-59, 1998.

[9-17] C. Hu, "Gate Oxide Scaling Limits and Projection," 1996 *International Electron Device Meeting*, pp. 319-322, 1996.

10 Experimental Verification

10.1 Introduction

Chapter 8 introduced the generalized universal charging damage model. A load line analysis between the plasma impedance and gate conduction determines the stress condition during plasma processing. The model simulates plasma charging damage as a function of gate oxide thickness, antenna ratio, ion density, and electron temperature. The simulation predicts that damage peaks at a T_{Dmax} of 5 nm, dependent on the processing condition. Simulation results indicated that the antenna size affects T_{Dmax}, with larger antennas shifting T_{Dmax} to thinner oxides. This chapter details experimental verification of the model's major predictions. More specifically, the presence of a peak in damage with gate oxide scaling, and the shifting of T_{Dmax} with antenna ratio.

10.2 Experimental Design

Fully processed wafers were subjected to an Argon plasma exposure in the PIII machine. See Chapter 2 for a description of the PIII tool. Determining the shifted plasma impedance is the most difficult module of the simulation. To simplify this, the wafer holder was grounded. Therefore, a simple Langmuir probe measurement provided the plasma impedance. To change the ion density, the ECR microwave power was altered from 300 W to 1500W, while the wafer holder was moved fore and aft in the machine. The ion density ranged from $9{\cdot}10^{9}cm^{-3}$ to $9{\cdot}10^{10}cm^{-3}$. An argon gas flow rate of 20 sccm set the pressure of the chamber at 0.5 mtorr. For these conditions, the open circuit voltage

ranged from -5.1 to -4.9 volts. Stressing of the devices by plasma electrons (V_{oc} is negative), tripled the effective ion density (since plasma electron saturation currents are greater than plasma ion saturation currents). The electron temperature ranged from 3.2 eV to 3.4 eV. Increasing the pressure to 2 mtorr decreases T_e to 1.7 eV, but for this experiment the pressure was held constant.

The oxide thicknesses, measured optically, were 2.3 nm, 2.5 nm, 3.5 nm, 4.5 nm and 6.4 nm. Four dies of each t_{ox} were placed on the wafer holder simultaneously. The wafer holder was kept at a constant temperature of 25 °C with water cooling. Plasma non-uniformity was minimal over the small area exposure. To amplify damage, processing time was lengthened to a fairly long time of 300 seconds.

Damage was measured by stress induced leakage current in transistors of sizes: 5x1 μm, 20x5 μm, 200x10 μm, and 200x200 μm. The pad area of 40000 μm^2 yielded antenna ratios of 2, 40 400, and 8000, respectively.

10.3 Stress Induced Leakage Current Measurement

Comparing plasma damage in different oxide thicknesses is inherently difficult. *C-V* interface trap measurements are not effective for oxides thinner than 4.0 nm. SILC, though, is a viable technique, with a damage metric of

$$LCR = \frac{J_s}{J_o} \tag{10-1}$$

where *LCR* is the leakage current ratio, J_s is the current in the stressed sample, and J_o is the current in the control sample. SILC's main drawback is the arbitrary choice of the sense voltage, the single voltage at which the leakage is measured and compared for the *LCR*. Measuring *LCR* at a constant electric field is an option. But, thinner oxides have higher intrinsic leakage currents for the same electric field. This skews the comparison, with thinner oxides always having a lower *LCR* (since their J_o is higher). Another option

is to measure *LCR* at a constant intrinsic leakage current. By measuring *LCR* at the same J_o for all oxides, the role of J_o is minimized. Minimizing J_o maximizes the measured *LCR*. For this work, J_o was set at 10^{-14} A, except for the 2.3 nm and 2.5 nm oxides, which were measured at a constant 1 V_{sense} (with a J_0 above 10^{-14} A).

With this constant J_o technique, comparing different device areas (and hence AR's) becomes troublesome, since each device will be measured with a different V_{sense}. The large area control devices will reach 10^{-14} A at a lower voltage. Since lowering V_{sense} increases *LCR*, smaller antenna devices (larger gate area) may exhibit a larger *LCR* even if the damage trend is opposite. Therefore, when comparing damage for identical t_{ox}, the measurement voltage should be kept constant for all devices. For this work, when comparing damage for different t_{ox}, J_o is set at 10^{-14} A, while V_{sense} is constant when comparing damage for a single t_{ox} with different AR's.

10.4 Damage Results

Figure 10-1 plots *LCR* for the $9{\cdot}10^{10}\text{cm}^{-3}$ ion density condition ($3{\cdot}10^{11}\text{cm}^{-3}$ effective ion density) comparing damage as a function of t_{ox}. Figure 10-1 highlights the damage trends as a function of t_{ox}, with each *AR* measured at it's optimal J_o of 10^{-14} A. As explained in the previous section, comparisons of damage with *AR* for a single a t_{ox} are not valid in Figure 10-1.

The vertical dashed lines are the predicted T_{Dmax} for each antenna ratio. With an $AR = 2$, the experimental peak was 6.4 nm, while the predicted was at 5.2 nm. The disparity between the experimental and simulated T_{Dmax}'s is not the full 1.2 nm difference. The experimental T_{Dmax} is discrete, and the data only determine a range in which T_{Dmax} lies. For the $AR = 2$, the experimental data can only conclude that the T_{Dmax} is greater than 4.5 nm. The exact location is indeterminable.

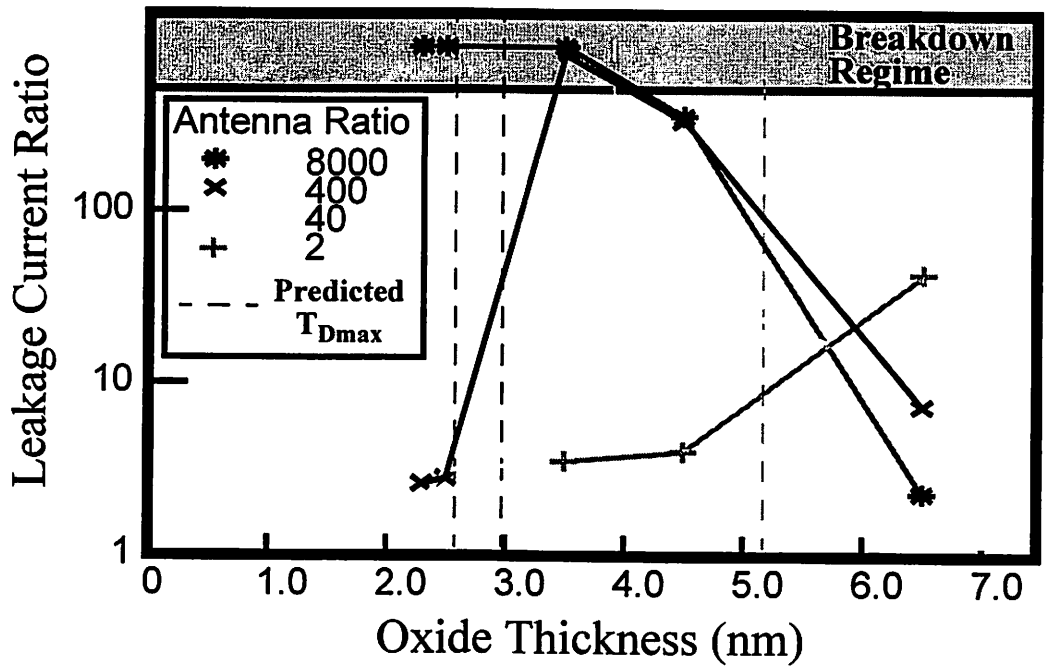

Figure 10-1 Experimental Verification of T_{Dmax}

Plasma charging damage peaks as gate oxides are scaled thinner. T_{Dmax} is a function of AR, with larger AR's shifting the peak to thinner oxides.

As the AR increases to 40 the simulation predicts that T_{Dmax} shifts to 3.9 nm. The data reveal a peak at 4.5 nm. With a AR of 400, the experimental peak is at 3.5 nm, while the prediction is for 2.9 nm. The 3.5 nm data point exhibited soft breakdown characteristics and is therefore placed in the breakdown region. All oxides that are broken down are placed at the same y-value. Finally, for AR = 8000, the simulation computes a T_{Dmax} of 2.7 nm, while the experimental results show breakdown for all the devices below 3.5 nm.

The data show definite peaks in damage with oxide scaling. This peak is not constant, but shifts with AR, as predicted. The simulation and experimental T_{Dmax} , within experimental determination, are consistent.

Figure 10-2 plots the damage trends as a function of AR. For this figure, all the LCR's for the same t_{ox} were measured at the V_{sense}, for optimal comparison. V_{sense} for

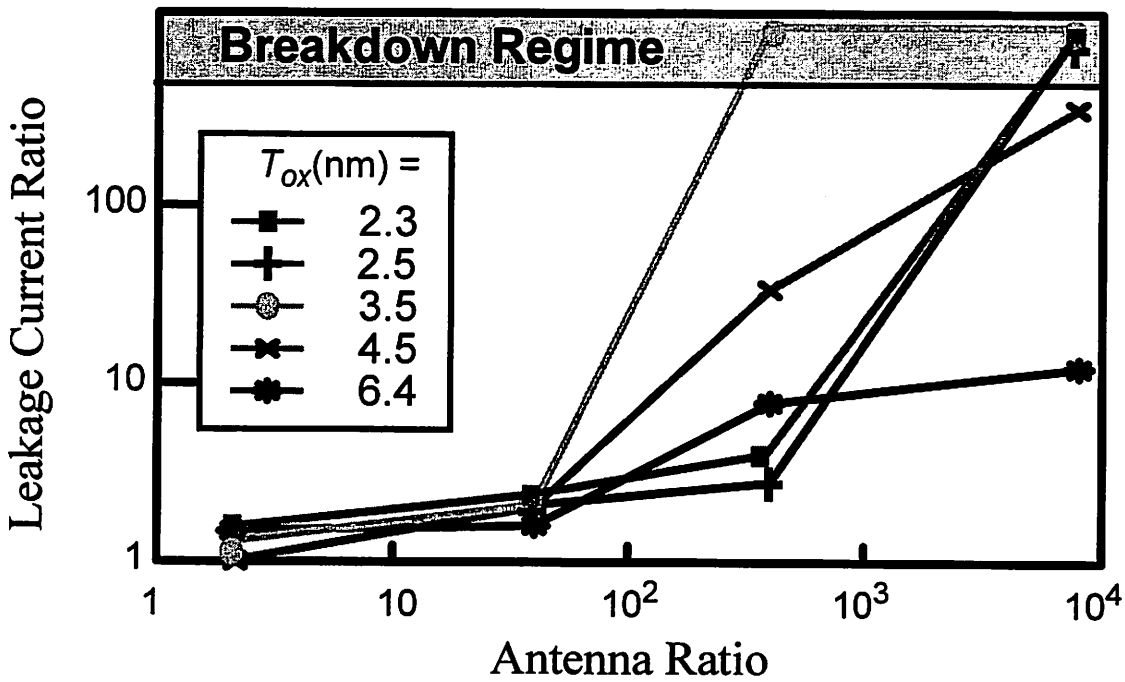

Figure 10-2 Experimental Verification of Antenna Effect

For all oxide thicknesses, damage increases with antenna ratio. This data is for an Argon plasma of ion density of $9x10^{10}/cm^3$, and a V_{oc} of -5.1 V.

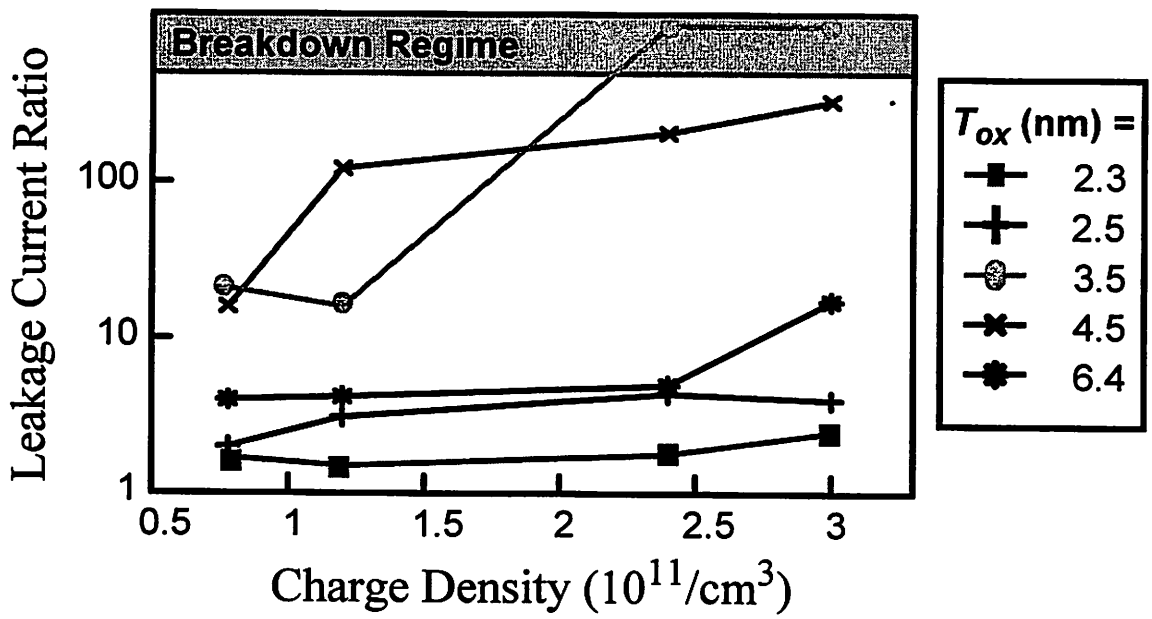

Figure 10-3 Experimental Verification of Ion Density Effect

For all oxide thicknesses, the general trend is a higher level of damage for higher ion densities. The data shown are for the $AR = 400$ devices.

each t_{ox} was set at the voltage corresponding to J_o = 10^{-14} A for the 5x1 μm control device. As expected, the data reveal a general trend of increasing damage with AR.

Figure 10-3 plots the damage trends as a function of ion density with the AR = 400 devices. For all conditions, both T_e and V_{oc} varied by less then 0.2 eV and 0.2 V, respectively. All oxide thicknesses show a trend of increasing damage with n_i. The n_i effect on damage is expected to be more pronounced with a larger variation in n_i across the wafer.

10.5 SILC Trends with Gate Oxide Thickness

The data in Figure 10-1 and Figure 10-2 reveal an interesting trend concerning SILC: the *LCR*'s for the oxides thinner than 3.5 nm are nearly bi-modal in distribution. The *LCR* is either close to 1 (e.g. no damage), or the oxides break down. This holds true for all data collected for the 2.3 and 2.5 nm oxides. For the data corresponding to t_{ox} = 3.5 nm, some of the collected *LCR*'s are between 5 and 10. This suggests that the maximum SILC before breakdown is dependent on t_{ox}, and that SILC might not be a sensitive predictor of damage in oxides thinner than 3.5 nm. From electrical stress data, the *LCR* before breakdown for t_{ox} less than 3 nm is less than 1.01 (e.g. a 1% increase in leakage) [10-1]. Therefore, extremely accurate measurements are necessary to extend the usefulness of SILC for detecting intermediate levels of damage thinner than 3 nm.

10.6 Conclusions

The major predictions of the universal charging damage model have been experimentally verified. Firstly, the peak in damage, T_{Dmax}, with gate oxide scaling has been verified. Secondly, T_{Dmax} has been found to be a function of antenna ratio, shifting to thinner oxides with larger antenna ratio. For this work, T_{Dmax} ranged from 6.4 down to below 3.5 nm. The data show that extreme stressing environments with high ion densities and large antennas can damage ultra-thin oxides. On the other hand with smaller antennas and lower stressing conditions, the thinner oxides appear more robust.

References

[10-1] J. H. Stathis, and D. J. Dimaria, "Reliability Projection for Ultra-thin Oxides at Low Voltages," Proceedings of the 1998 *International Electron Device Meeting*, pp. 167-170, 1998.

11 PIII Charging Damage

11.1 Introduction

The previous chapters developed and experimentally verified a generalized universal plasma charging damage model. This chapter extends the model for predicting and minimizing gate oxide charging damage for the Plasma Immersion Ion Implantation process.

All implantation processes are sensitive to plasma charging damage. Implantation deposits positive charge on the surface of the wafer. This charge accumulation may stress the gate oxides. With conventional beamline implantation, an electron flood gun showers the wafer surface with negatively charged electrons in an attempt to neutralize the positive charge. PIII does not utilize flood guns, but rather relies on bias pulsing for charge neutralization.

In the PIII process, high voltage microsecond negative pulses applied to the substrate, accelerate and implant the plasma ions. These implanted ions, and the secondary electrons that they eject, deposit positive charge on the surface of the wafer. An off time follows each implant pulse, allowing the plasma electrons to neutralize the positive surface charge. Minimizing charging damage requires the optimization of the pulse frequency (f_p) and the duty factor (DF).

The key to extending the universal damage model for the PIII process is to calculate the open circuit voltage as a function of f_p and DF. En [11.2] has successfully combined the equations governing the plasma and the gate oxide to calculate V_{oc} as a function

of pulsing condition. This model is modified to include the effects of sub-surface structures. More specifically, it is shown that wells and substrate type (n, p, or dielectric) affect V_{oc} and gate oxide charging damage.

11.2 Gate Oxide Damage Measurement

Oxide charging damage is assessed by extracting the interface trap density from *C-V* measurements. The ramp rate for the capacitance measurements in this chapter is 0.03 V/s. This is rapid enough to average out noise while reasonably maintaining thermal equilibrium. To further ensure equilibrium, the MOS capacitor is ramped from inversion to accumulation, eliminating minority carrier generation effects.

11.3 General PIII Oxide Charging Case

Before discussing the effect of sub-surface structures on gate oxide charging, the origin and dependencies of oxide charging in PIII needs to be understood [11.2].

For a floating surface, the net current from the plasma to the wafer must be equal to zero ($J_i = J_e$). To satisfy this condition, the surface voltage of the wafer is

$$V_s = V_p - T_e ln\left(\frac{M}{2\pi m}\right)^{1/2} \tag{11-1}$$

where T_e, m, M, and V_p are the electron temperature, the electron mass, the ion mass, and the plasma potential respectively. Because the gate oxide insulates the surface from the substrate, the initial equilibrium, before any applied bias, results in the surface voltage (V_s) being equal to the floating potential (V_f) of the plasma. With a grounded wafer holder, the voltage across the gate oxide is also V_f. Therefore, for simple plasma exposure V_{oc} equals the plasma V_f in the universal charging model.

The pulsed bias during PIII modulates the surface voltage and shifts V_{oc}. During the negative pulse, the impinging plasma ions and ejected secondary electrons make the

surface voltage more positive. Moreover, the plasma electrons can not surmount the sheath potential ($V_{pulse} >> T_e$), reducing J_e to zero. During the pulse-off stage, the incoming plasma electron current (determined by Equation (3-8)) will tend to return V_s back to V_f. However, if the pulse frequency is too rapid, the off pulse plasma electron current will not have enough time to reduce V_s back to the initial equilibrium before the next pulse begins. In this high frequency regime, some positive charge accumulates and V_s will be greater than V_f at the start of the second pulse. Additional pulses deposit more positive charge on the surface, making V_s even more positive, and consequently increasing the plasma electron current during the pulse off stage (by Equation (3-8)). This process repeats until an equilibrium is established, with the time-averaged plasma electron current balancing the plasma ion and secondary electron currents

$$\langle J_i \rangle + \langle J_{se} \rangle + \langle J_e \rangle = 0 \tag{11-2}$$

$$\langle J_i(1 + \gamma(V_i)) \rangle + \left\langle \left(\frac{1}{4} q n_o u_e \exp^{-\frac{(V_p - V_s(t))}{T_e}} \right) \right\rangle = 0 \tag{11-3}$$

Figure 11-1 depicts the transition from the initial equilibrium to the pulsing equilibrium.

11.3.1 Substrate Bias Frequency and Duty Factor Effects

The V_s required to balance the plasma electron current with the plasma ion and secondary electron currents depends on the pulsing frequency (f_p). In the limit of $f_p \rightarrow 0$, or $DF \rightarrow 0$, the pulsing becomes negligible, and the plasma electrons have plenty of time between pulses to neutralize the positive charge. For this case the equilibrium V_s and V_{oc} approach the V_f (usually a negative value). As the pulsing frequency or DF increases, the time available for J_e to satisfy Equation (11-2) becomes successively smaller. To offset this reduction in time, V_s and V_{oc} become more positive to draw more electron current from the plasma. At some critical frequency, V_{oc} actually equals zero, the minimum dam-

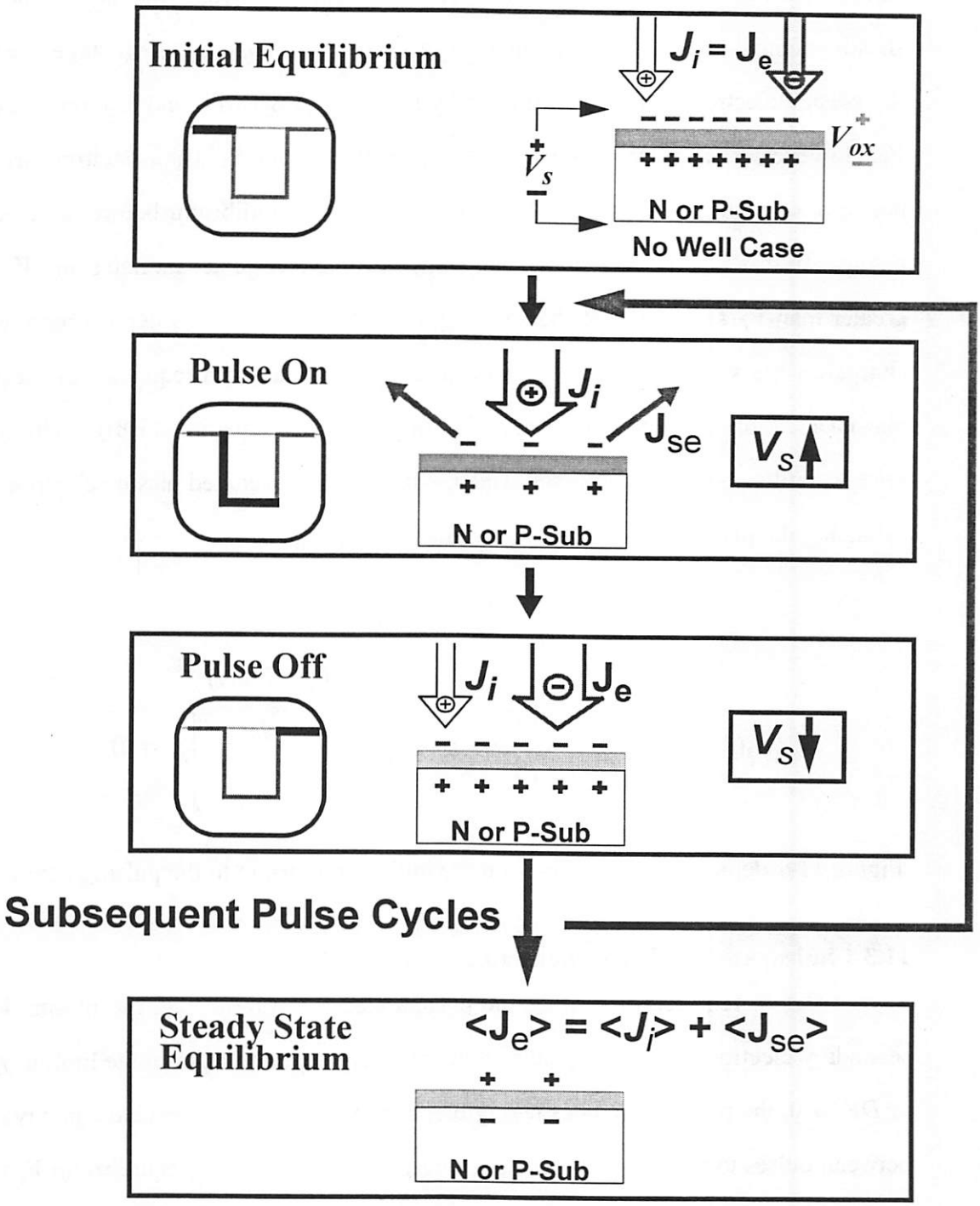

Figure 11-1 Transformation to Pulsing Equilibrium for PIII

Before pulsing, the surface voltage (V_s) is equal to the floating potential (usually negative). When the pulse is on, the large potential barrier repels the plasma electrons, while plasma ions bombard the surface ejecting secondary electrons, making V_s more positive. During the pulse-off stage, the plasma electrons return to the surface to neutralize the surface charge. After many pulse cycles, an equilibrium V_s is reached, which balances the time-averaged plasma electron current with the plasma ion and secondary electron currents.

age condition. Increasing f_p further switches V_{oc} from negative to positive, increasing damage again. Eventually, as $f_p \rightarrow \infty$, or the duty factor $\rightarrow 1$, the pulsing becomes DC like, and V_s rises uncontrollably, causing catastrophic oxide failure soon after implantation begins.

Using the PIII plasma model developed in Chapter 3, the frequency dependence of the pulsing equilibrium V_{oc} is plotted for the general wafer in Figure 11-2. Minimum damage results when $|V_{oc}| = 0$, which for the simulated 1μs, 5 kV ideal pulse, occurs at 150 kHz. This substrate bias effect on oxide charging has been experimentally confirmed [11.6].

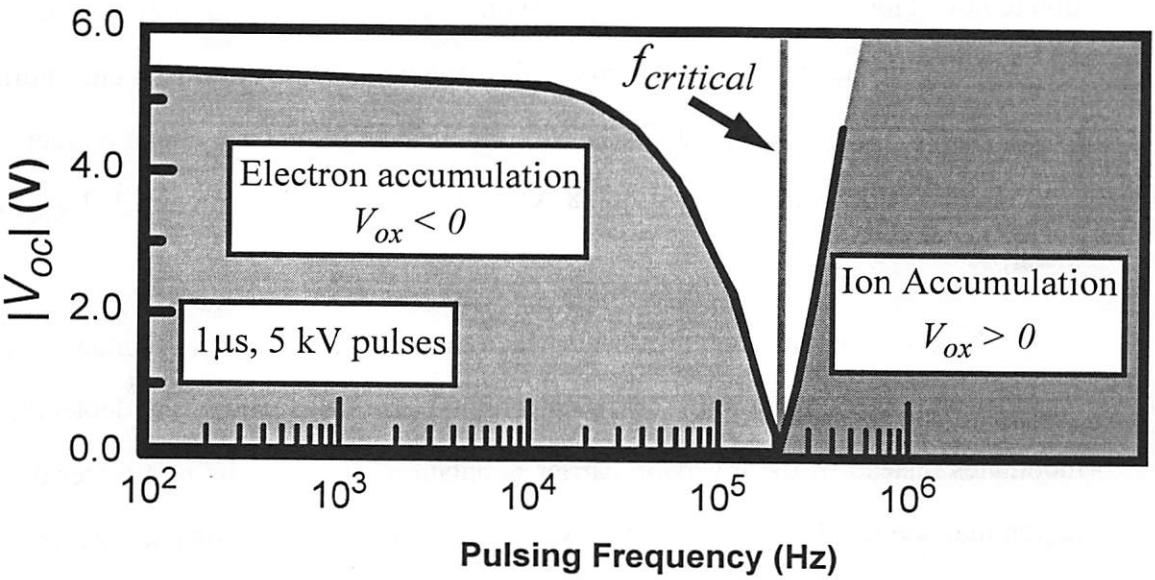

Figure 11-2 Substrate Pulsing Frequency Effect

Without any substrate bias, $|V_{oc}|$ equals to the plasma floating potential. As the frequency increases, the neutralizing time for plasma electrons decreases, requiring V_{oc} to rise in order to increase the plasma electron current. As the frequency increases, the V_{oc} required to balance all the currents changes sign from negative to positive, increasing rapidly. The frequency which yields a V_{oc} equal to zero is named $f_{critical}$. The simulation parameters are: 1μs, 5 kV pulses, 0.1μs rise and fall times, 10 nm gate oxide, $3.76 \cdot 10^{10}$ cm^{-3} Argon ion density, 4 eV electron temperature, and a 13.23 V plasma potential.

11.3.2 Substrate Effect

V_s equilibrates to approximately the same voltage, irregardless of the substrate type. If a depletion region exists beneath the oxide, the substrate will drop a portion of the V_s. This reduces V_{oc} across the oxide, lowering the damage. The maximum steady-state voltage dropped in a depletion region for an inverted channel (i.e. $V_s > V_{threshold}$) is:

$$V_{depl0} = 2\frac{kT}{q}ln\left(\frac{N_{ch}}{n_i}\right) \tag{11-4}$$

where V_{depl0}, T, N_{ch}, and n_i are the thermal equilibrium depletion voltage, the substrate temperature, channel doping, and the intrinsic carrier level, respectively. Since the depletion region is formed underneath the gate oxide, the doping concentration directly beneath the oxide in the channel region determines the depletion width. For a 10^{17} cm^{-3} uniformly doped channel, V_{depl0} equals -0.82V. Therefore, in steady-state, the depletion region lowers the gate oxide voltage stress by 0.82V. If the channel is not inverted, V_{depl} will be lower.

Non-steady state situations occur when the voltage on the gate changes more rapidly than the inversion carriers form or recombine. In this situation, the depletion width modulates instead of the inversion carrier population. This results in a wider depletion region than the steady-state, which decays to the steady-state value on the order of the carrier generation/recombination rate, which typically ranges from microseconds to milliseconds. The larger depletion widths occurring in transient situations provide extra protection for the gate oxide, with voltage drops in the depletion region exceeding 1 volt.

The substrate effect occurs for positive stressing for P-substrates, and negative stressing for N-substrates. With plasma exposure, the surface voltage is usually negative. Therefore N-substrates will contain a depletion region and should show less exposure damage than P-Substrates. With PIII pulsing, the voltage stress is negative for slow f_p and

positive for faster f_p. Therefore the substrate type that shows less damage will switch with increasing pulsing frequencies.

Figure 11-3 shows PIII damage for 11 nm gate oxides on both N and P substrates under varying pulsing frequencies. Plasma exposure time was kept constant at 5 minutes. In this experiment, the pulsing frequency was never high enough to switch the surface voltage from negative to positive, and therefore the P-substrates oxides show more damage for the entire frequency range. It is predicted that if the pulsing frequency could be raised further, the N-substrate would eventually exhibit higher damage than the P-substrate.

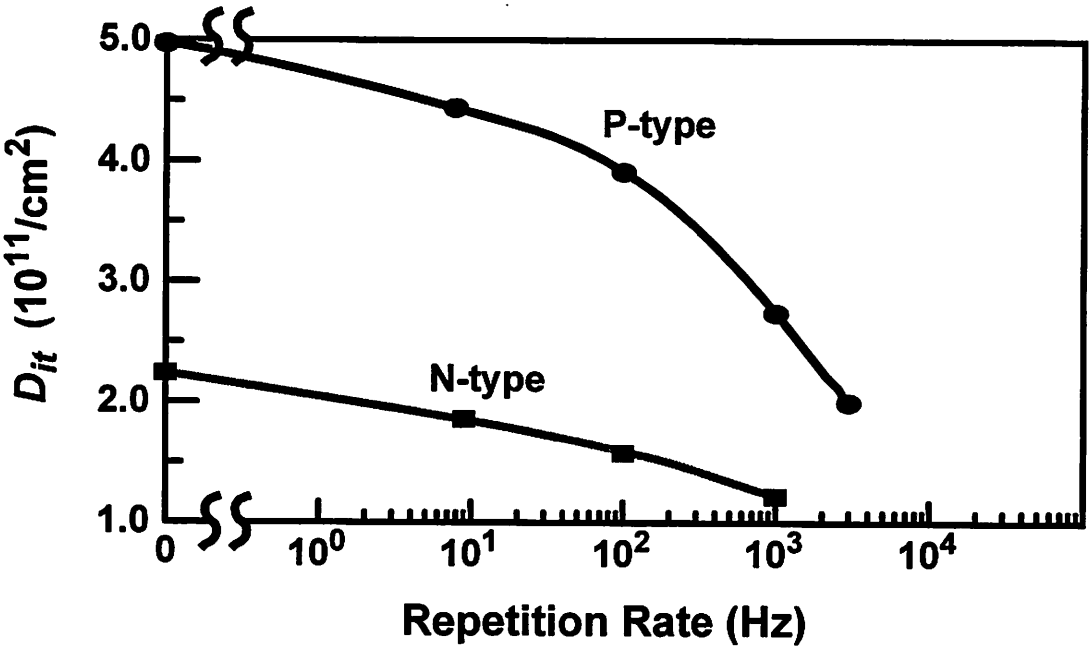

Figure 11-3 Damage Comparing N-Substrates and P-Substrates

The N-substrate suffers lower damage for negative stress voltages because of the depletion region in the substrate. P-substrates suffer lower damage for positive stressing. In this experiment, the relatively low pulsing frequencies resulted in negative stress for all pulsing conditions, with the N-substrates exhibiting less damage. Implantation was a constant 5 minutes for all devices.

11.4 Well Structure Effects

Well structures are essentially p-n diodes, which can either be forward biased or reverse biased (Figure 11-4). When forward biased, the well drops little voltage and is like a short. When reverse biased, the well acts like a capacitor, and can support a significant voltage. The capacitance from the well-bulk junction is determined by the lower doped region, which is usually the bulk. The voltage across the well-bulk junction modulate V_s, altering the oxide charging damage. In the simulations two different well structures are compared, P-Well and N-Well. For each case, the substrate doping is $5 \cdot 10^{15} cm^{-3}$.

11.4.1 N-well

An N-well beneath the gate oxide effectively adds a diode in series with the gate oxide (Figure 11-4). Assuming that all the charge leaks out of the well before pulsing begins, the initial equilibrium is the same as the no well case with $V_{well} = 0$, and $V_{oc} = V_{ox} = V_s = V_f$. During pulsing, the charge in the well does not necessarily have time to leak out, producing a voltage drop across the well junction. With positive charge deposition the N-well is reverse biased. From Poisson's equation, with an abrupt, one-sided junction the voltage dropped by the well is:

$$V_{well} = \frac{(Q_{well})^2}{2q\varepsilon_s N_{sub}} \tag{11-5}$$

where N_{sub}, V_{well}, Q_{well}, and ε_s are the bulk doping concentration, the well voltage, well stored depletion charge, and silicon permittivity, respectively. As stated previously, the equilibrium V_s is nearly independent of substrate type. Therefore, in order to achieve the necessary surface potential rise to the new equilibrium, V_{ox} does not have to increase much because the well capacitance supports the extra voltage. Moreover, because the reverse-biased capacitance of the low doped well junction is much less than the oxide

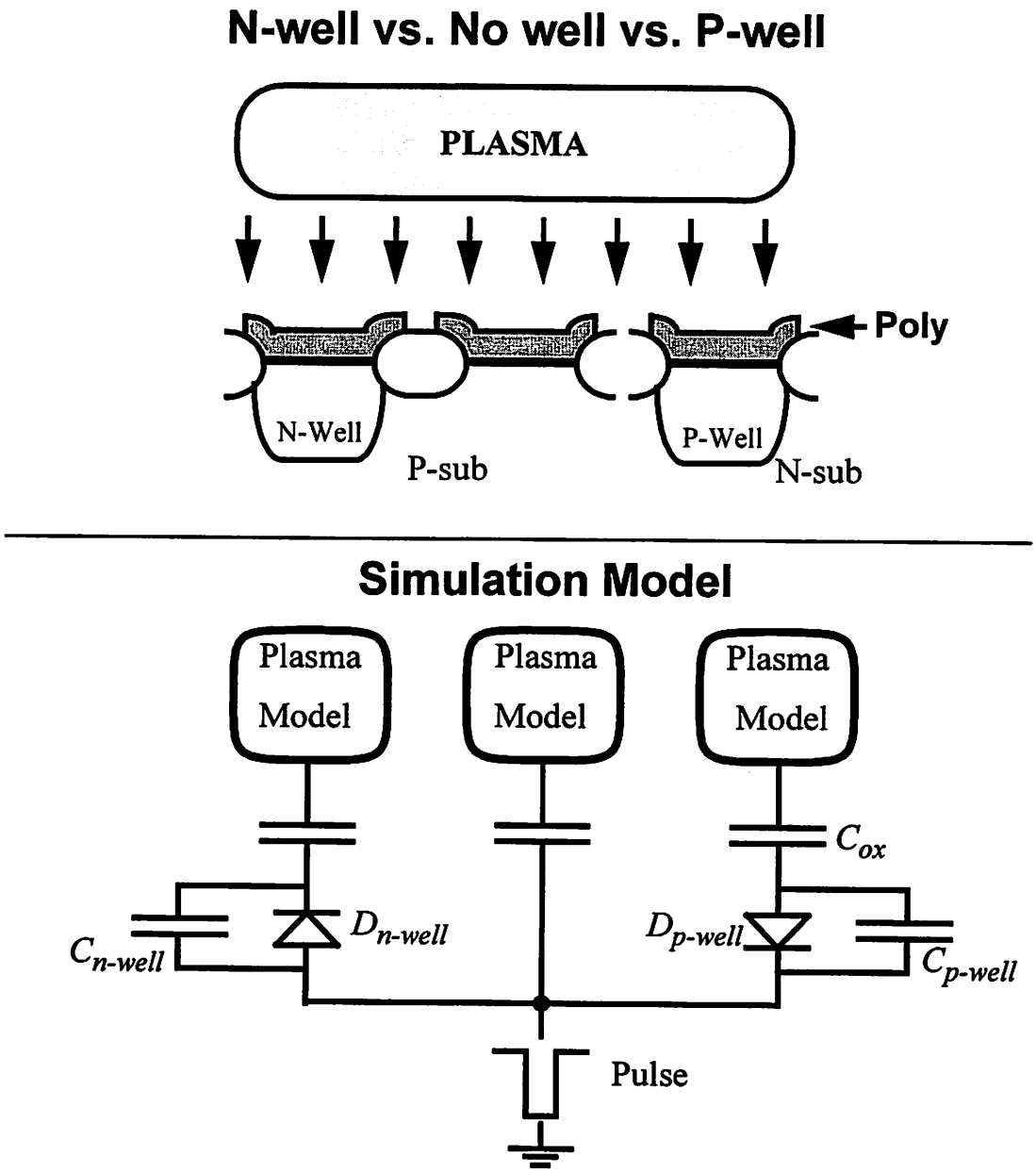

Figure 11-4 Well Simulation Model

Simulation model for wells. The difference between a P-well and an N-well is the polarity of the diode. The parallel capacitor includes junction and transit time capacitances, while generation in the space charge region is included as the leakage mechanism.

capacitance, the well supports the majority of the rise in V_s. This translates to a small ΔV_{ox} during pulsing, as compared to the no well case.

11.4.2 P-well

The initial equilibrium is the same the as two previous cases, with $V_{oc} = V_{ox} = V_s = V_f$, and $V_{well} = 0$. During the pulse, the positive charge deposition forward biases the P-well, which supports a small forward voltage, and stores a correspondingly small amount of injected minority carriers (Figure 11-5). When the pulse is turned off, the electron current deposits negative charge on the surface. The electrons neutralize the charge stored in the well, eventually reverse biasing it. In contrast to the N-Well case, the P-well voltage subtracts from V_{ox}, rather than adding to it. Therefore, to obtain the required V_s dictated by the pulsing conditions, V_{ox} must be more positive to compensate for the negative voltage stored in the well. This results in a larger ΔV_{ox} for the P-well case.

11.4.3 Well Simulation Results

Table 11-1 summarizes the well effect during the initial equilibrium, pulse on, and pulse off stages. Simulated transient results for ΔV_{ox} for the different well cases are shown in Figure 11-6. For each case V_{ox} begins at V_f, and then adjusts to a new equilibrium based on f_p, duty factor, and pulse voltage. The response variable is ΔV_{ox}. Compared to the no well case, a P-well results in a larger ΔV_{ox}, while an N-well results in a smaller ΔV_{ox}.

11.4.4 Leakage Current

Figure 11-7 shows the well potential during pulsing. The N-Well is always reverse biased with an offset from zero. This is the main reason why ΔV_{ox} is so much smaller for the N-Well case (Figure 11-6). If this offset charge leaks out over time, the effect of the N-Well will be diminished. In contrast, since there is no permanent stored charge in the P-Well (it switches from forward to reverse biased with each half cycle), the P-Well effect

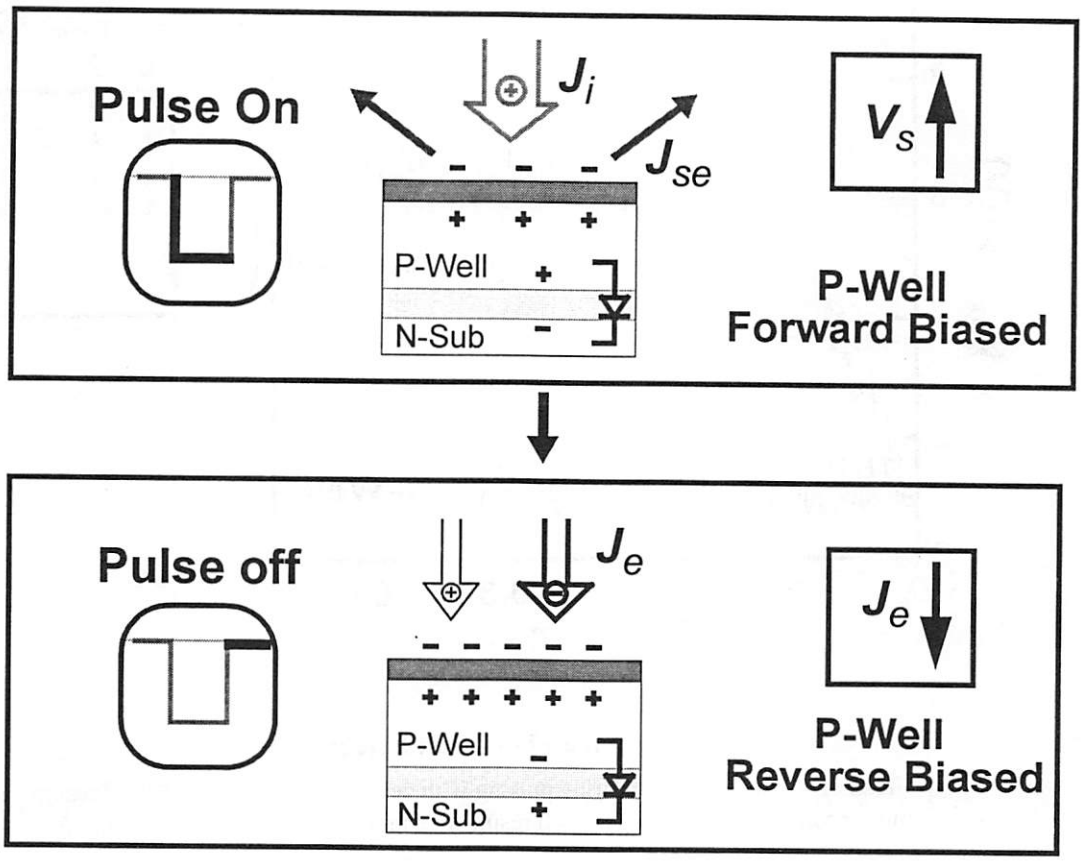

Figure 11-5 PIII Pulsing with a P-Well

During the pulse on time, the P-Well is forward biased, and conducts the implanted charge through to the back contact. Then, during the pulse off time, the plasma electron current reverse-biases the well, and creates a negative voltage that repels additional plasma electrons, reducing J_e. The net effect of the well, is to eventually make the V_{ox} more positive to compensate for the negative well voltage.

Table 11-1 Well Effect for the Three Stages of PIII

Pulse Condition	N-Well	P-Well
Initial Equilibrium	no charge	no charge
Pulse On	Reverse-biased	Forward Biased
Pulse Off	Reverse-biased	Reverse-biased

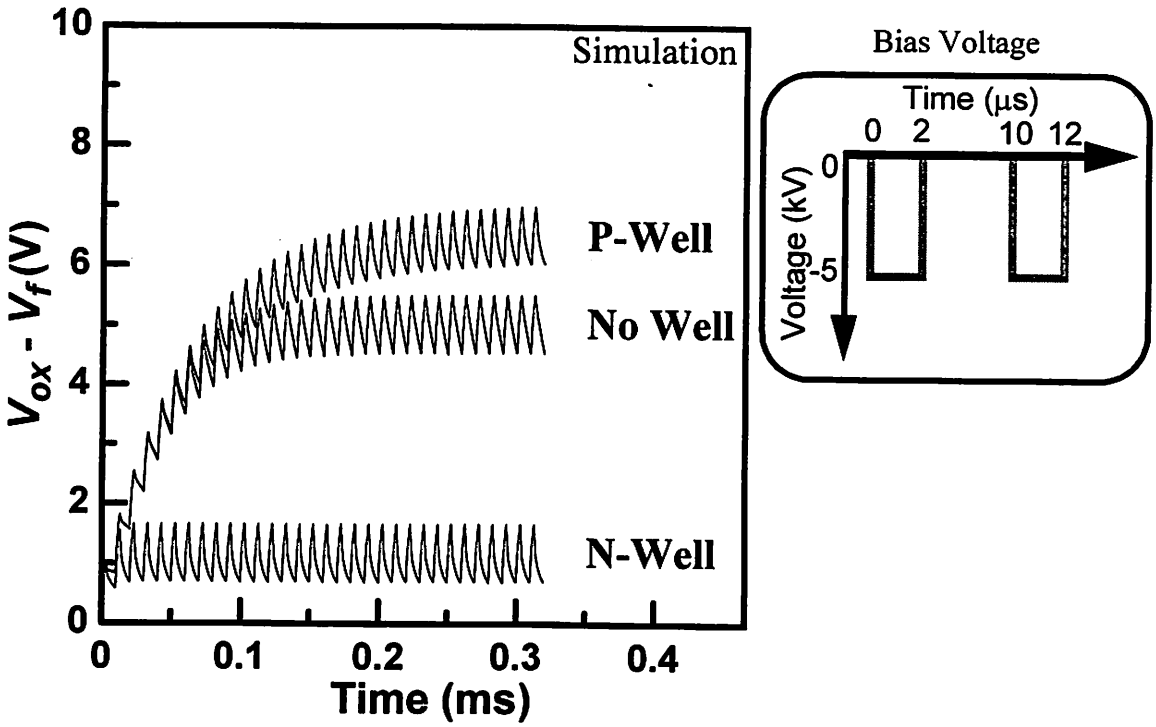

Figure 11-6 Well Effect on V_{ox}

Transient analysis of V_{ox} during pulsing with devices in wells. The P-Well results in a more positive V_{ox}, while an N-well results in a more negative V_{ox}. V_f equals -6 V.

will not be diminished by leakage as long as the carrier generation rate is less than the pulsing frequency.

The leakage rate for the reverse biased wells depends on the light intensity during the plasma processing. Without light, leakage is low; for the wells fabricated the leakage was less than 1μA/cm^2 at -5V. With unobstructed illumination, the leakage jumps by many orders of magnitude, to over 1mA/cm^2. The leakage rate depends on how much light reaches the underlying silicon, and would be reduced by absorption or reflection by surface layers, such as the poly gate, field oxide, metal layers, and inter-level dielectric. For PIII processing, charge deposition rates typically range around 1mA/cm^2. Therefore, the well effect may be diminished by leakage under high wafer illumination and low absorption by overlaying layers.

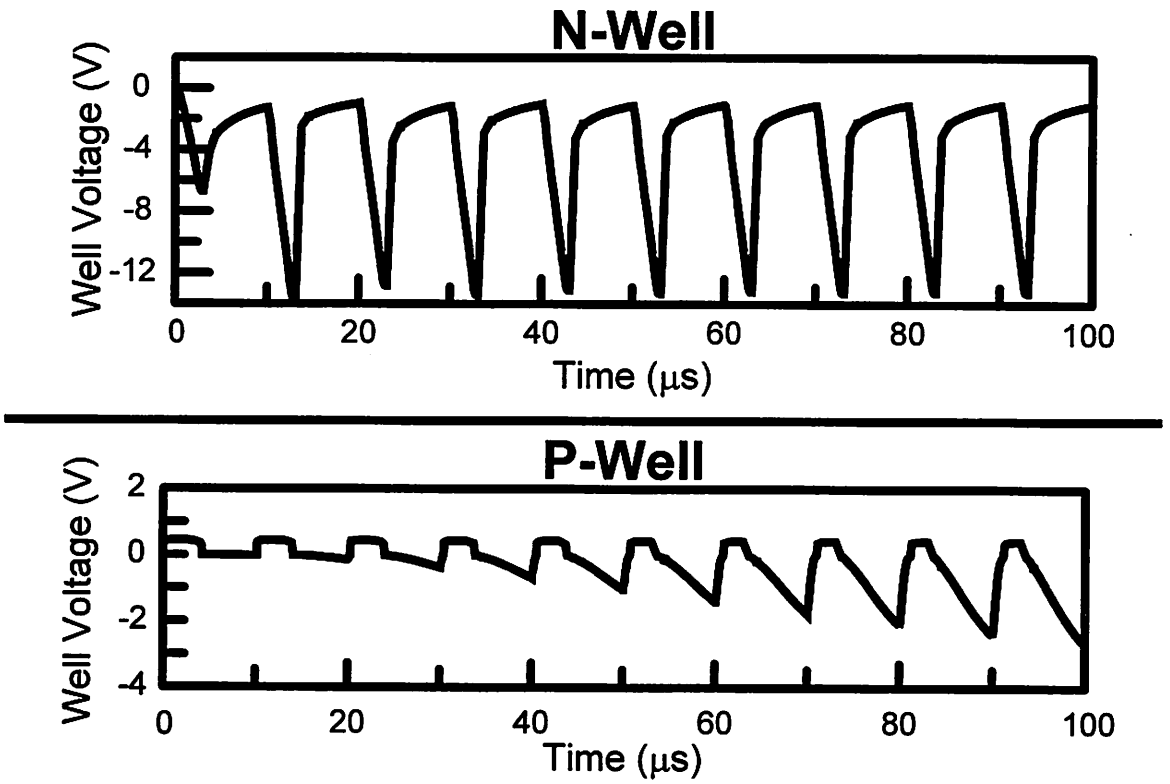

Figure 11-7 Well Voltage during Processing

The simulated well voltages for a $5 \cdot 10^{15}$ doped substrate. The N-well has a DC offset which, over time, may be reduced by leakage. If the well losses its charge offset due to leakage currents, the effect of the well on gate oxide charging is reduced. The P-Well changes from forward to reverse biased with each pulse, and therefore, leakage is only important if it is significant within one pulse.

11.4.5 Well and Substrate Effect

In the previous sections, the well and substrate effects are de-coupled, but to form an accurate model they must be combined. In an N-well, the channel is doped n-type. If the stressing voltage is negative, then both the substrate and well effect will affect gate oxide damage. With the same analysis, the P-well, with p-type channel doping, will have a depletion region when $V_{ox} > 0$. Therefore, both the substrate and well effects will occur during positive stressing, at very high f_p's

11.4.6 Experimental Verification

To verify the well and substrate effects, two different wafers, an N-substrate with a P-well, and a P-substrate with an N-well, with 11 nm gate oxides were implanted at vary-

ing frequencies. The microwave power for ECR plasma generation was 900W, the pulse voltage was 2.5 kV, the pulse width 1μs, the pulse fall time ~35 μs, and the pulsing frequencies varied from 100 Hz - 22 kHz. Damage was monitored by comparing interface trap densities extracted from *C-V* measurements before and after processing (Figure 11-8). All four curves follow the same trend, initially showing some increase in damage with pulsing frequency, until at high frequency the damage is reduced. The initial rise in damage can be attributed to the increase in wafer temperature as the pulse frequency increases [11.8]. This experiment utilized simple sample clamping without silver paste or water cooling. As predicted by simulation (Figure 11-2), further increases in f_p create less damage. The N-Well, and N-substrate show less damage then the P-regions on the same wafer. This is due to the depletion region underneath the gate oxide reducing V_{ox}. The pulsing frequency was never fast enough to change V_s and V_{oc} from negative to positive. Therefore the P's never showed less damage, as was predicted for very high frequencies (Figure 11-2). We expect the substrate and well effects to be more prominent when combined with antennas.

11.4.7 The Effect of Different Well Structures

The well effects shown in the simulations are highly sensitive to the well capacitance. For the well effect to be significant $C_{well} << C_{ox}$. Various well structures are qualitatively ranked by the degree of the simulated well effect (Figure 11-9). The high doping on both sides of the triple well junction results in the highest capacitance, and the least amount of well effect.

11.5 Charging Damage and Dielectric Substrates

Using the fully coupled SPICE model (Section 3.3), PIII charging damage may be simulated for dielectric substrates. Thin film transistors and silicon on insulator technologies contain dielectric substrates that modify charging damage during PIII. The model assumes a fully insulating substrate with negligible leakage currents. As in the well and

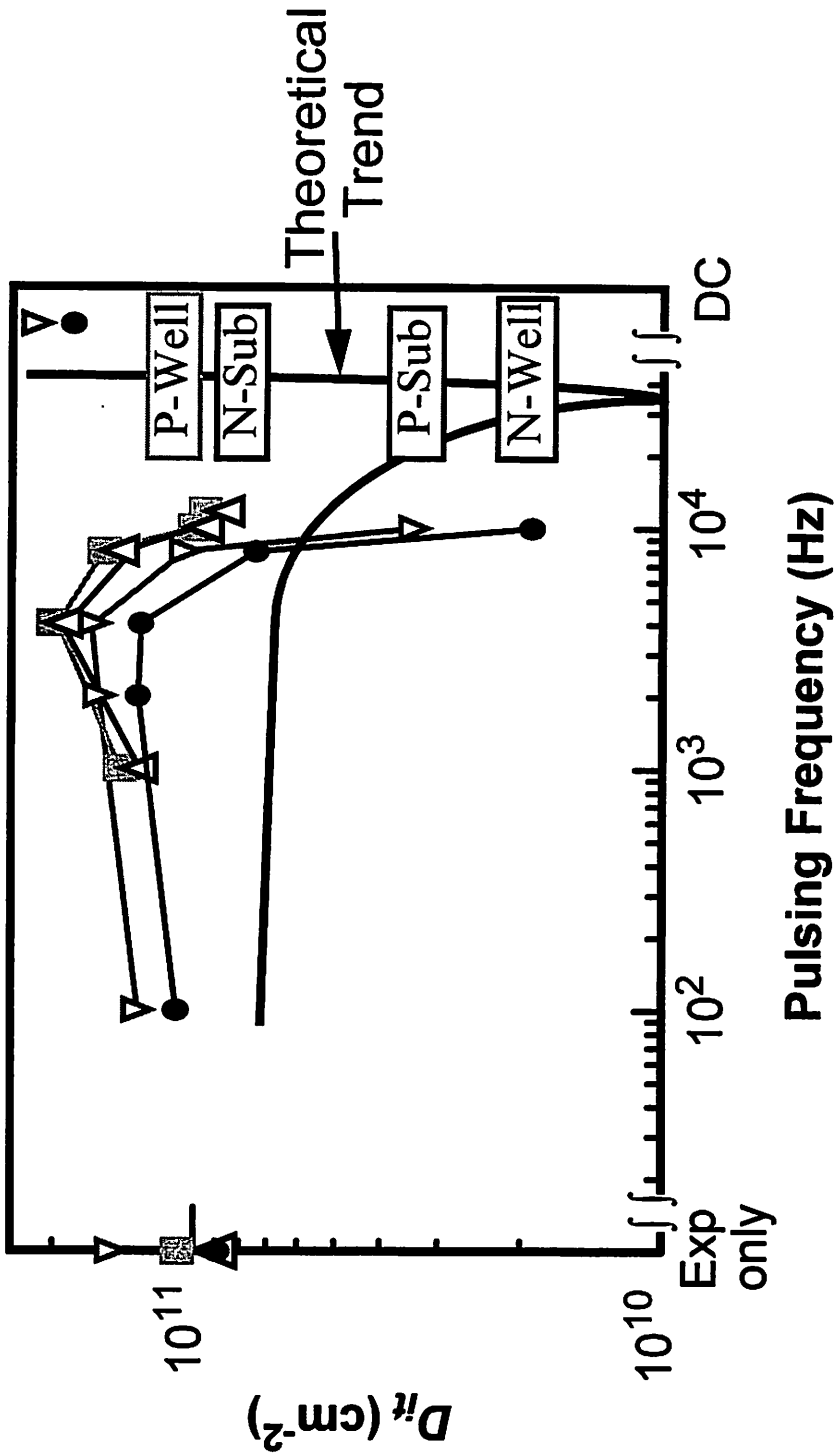

Figure 11-8 Experimental Data Demonstrating the Well Effect

Generated interface traps for 4 different structures: N-Well, P-Substrate, P-Well, and N-Substrate. All 4 curves follow the same general trend predicted by the model. Because some voltage drops across the depletion region, the n-doped channel region devices exhibit less damage. The pulse width for the experiment was 1μs.

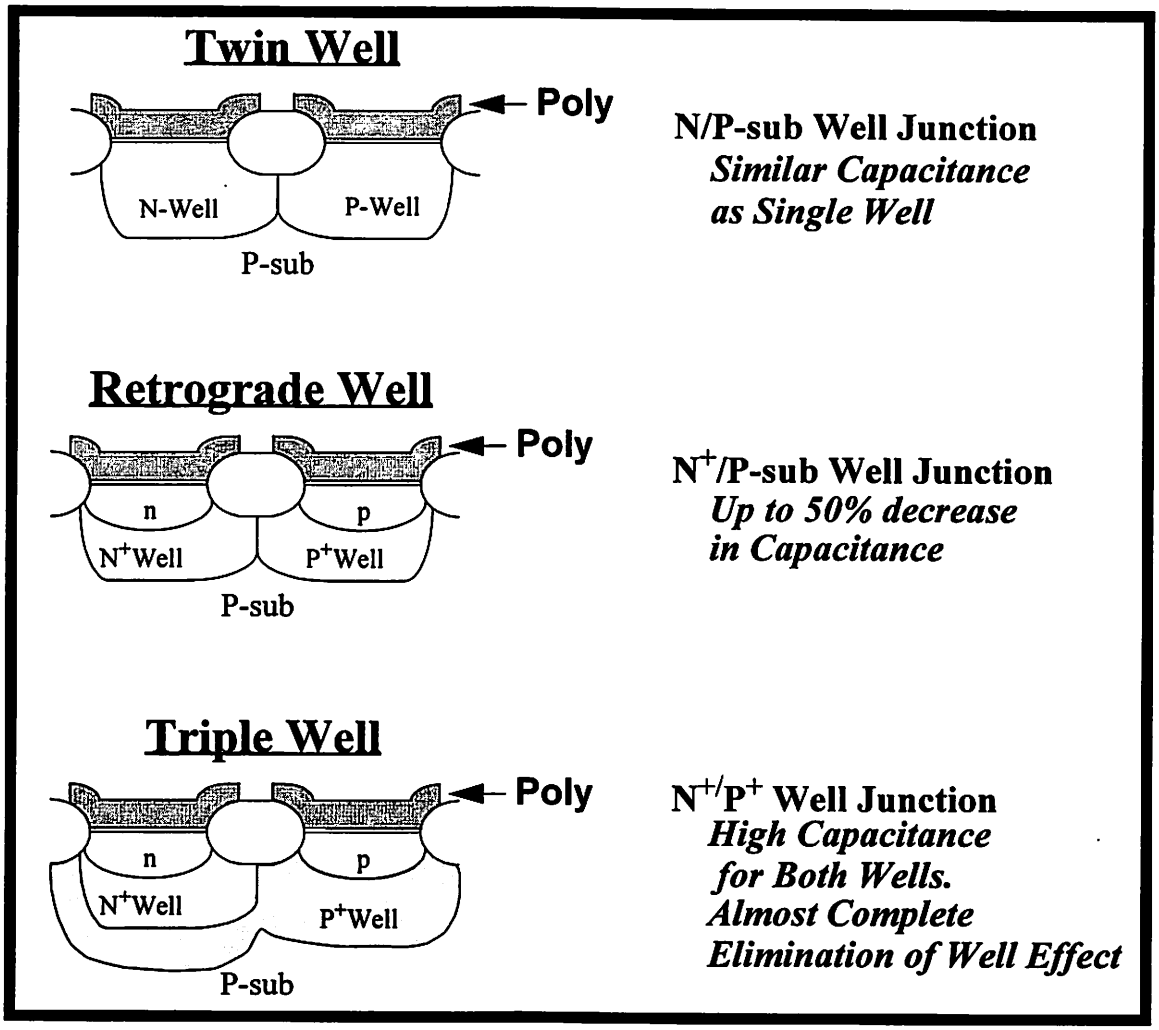

Figure 11-9 Well Effect Comparisons

The choice of well structure determines the degree of well effect, with higher capacitance well structures exhibiting less effect on charging.

substrate simulations, the time-average equilibrium surface potential is determined by the pulsing conditions. Sub-surface structures simply alter the percentage of V_s that drops across the gate oxide. With perfectly insulating substrates, a simple capacitor divider model is appropriate. Since the insulating substrate capacitance will usually be much smaller than the gate oxide capacitance, the majority of the surface potential drops across the substrate and not the gate oxide. Therefore, gate oxides should show little charging damage during processing with insulating substrates.

11.6 Single Pulse AC Damage from Antennas

With actual device layouts, gates are not isolated from one another, but are connected together with either metal or poly lines. These conducting paths usually run over thick dielectric isolating material such as field oxide. The capacitance of the field oxide is much less than the gate oxide, leading to a varying surface voltages across the wafer with uniform charge deposition. Charge flows from the interconnect to the gate to equalize the voltages. If the electric field across the gate oxide yield significant tunneling currents, stress and damage result. This funneling of charge from a large collecting area (the antenna) to the gate oxide is called the antenna effect.

11.6.1 Conventional Antenna Effect

For a given charge deposition, the voltage generated across the gate oxide is a function of the ratio of capacitance of the gate oxide and the field oxide, and the ratio of the areas of the gate and the interconnect:

$$V_{ox} = \frac{(AR+1)Q_{dep}}{AR \cdot C_{ant} + C_{ox}} C_{ox} \tag{11-6}$$

where V_{ox}, Q_{dep}, C_{ox}, C_{ant}, and AR are the oxide voltage, charge deposited per unit area, the oxide capacitance per unit area, the antenna capacitance per unit area, and the antenna ratio, respectively. The antenna effect is not the result of charge build-up over many pulse cycles, as is the case in the previous sections, but rather from the charge deposited from single pulses. Large antennas (greater than 100) generally act as like voltage sources, since the charge deposited on the antenna exceeds the charge tunneling through the gate oxide by an order of magnitude. This is shown in Figure 11-10, which shows a simulation of tunneling current from an instantaneous $10^{12}/cm^2$ deposition across the wafer, with a gate oxide of 5 nm, field oxide thickness of 200 nm, and an antenna ratio of 100. The peak electric field is 13.5 MV/cm, which decays to 10.5 MV/cm in 10μs. The $10^{12}/cm^2$ charge deposition consists of mostly secondary electron ejection. The total charge con-

ducted through the oxide in a single pulse is minuscule. But, considering that a typical PIII implant contains at least 10^5 pulses and potentially more than 10^8 pulses, the integrated tunneling current over the entire implant is enough to damage the oxide. From simulation, it is determined that the charge per pulse must be less than $10^{11}/cm^3$ with large antennas (AR > 100) to avoid gate oxide charging damage. In addition, the maximum charge per pulse scales with gate oxide thickness, with thinner oxides having a smaller maximum dose per pulse. The PIII antenna effect has been experimentally confirmed [11.7].

Since the single pulse antenna effect occurs during individual pulses, the only ways to eliminate the effect is to limit the amount of charge deposited per pulse, limit the antenna size, or provide a leakage path for the antenna through a connection to the substrate.

11.6.2 Dielectric Substrate Antenna Effect

The thick buried oxide (BOX) in SOI devices profoundly affects gate oxide charging and the antenna effect. As before, the capacitance to ground is lower over the field regions than the gate oxide regions, generating larger voltages for uniform charge deposition. Charge then flows from the field regions to the gate regions to equalize the voltages. The difference between SOI and bulk devices is that the gate oxide capacitance is now in series with the BOX. The buried layer will usually be at least a magnitude thicker than the gate oxide, reducing the capacitance by a similar value. Therefore, little charge needs to flow to build-up enough voltage in the BOX layer to impede further charge conduction. Therefore, antennas should not increase gate oxide damage significantly, since the BOX layer supports the extra voltage generated by the antenna, and not the gate oxide. These predictions have been confirmed experimentally [11.9, 11.10].

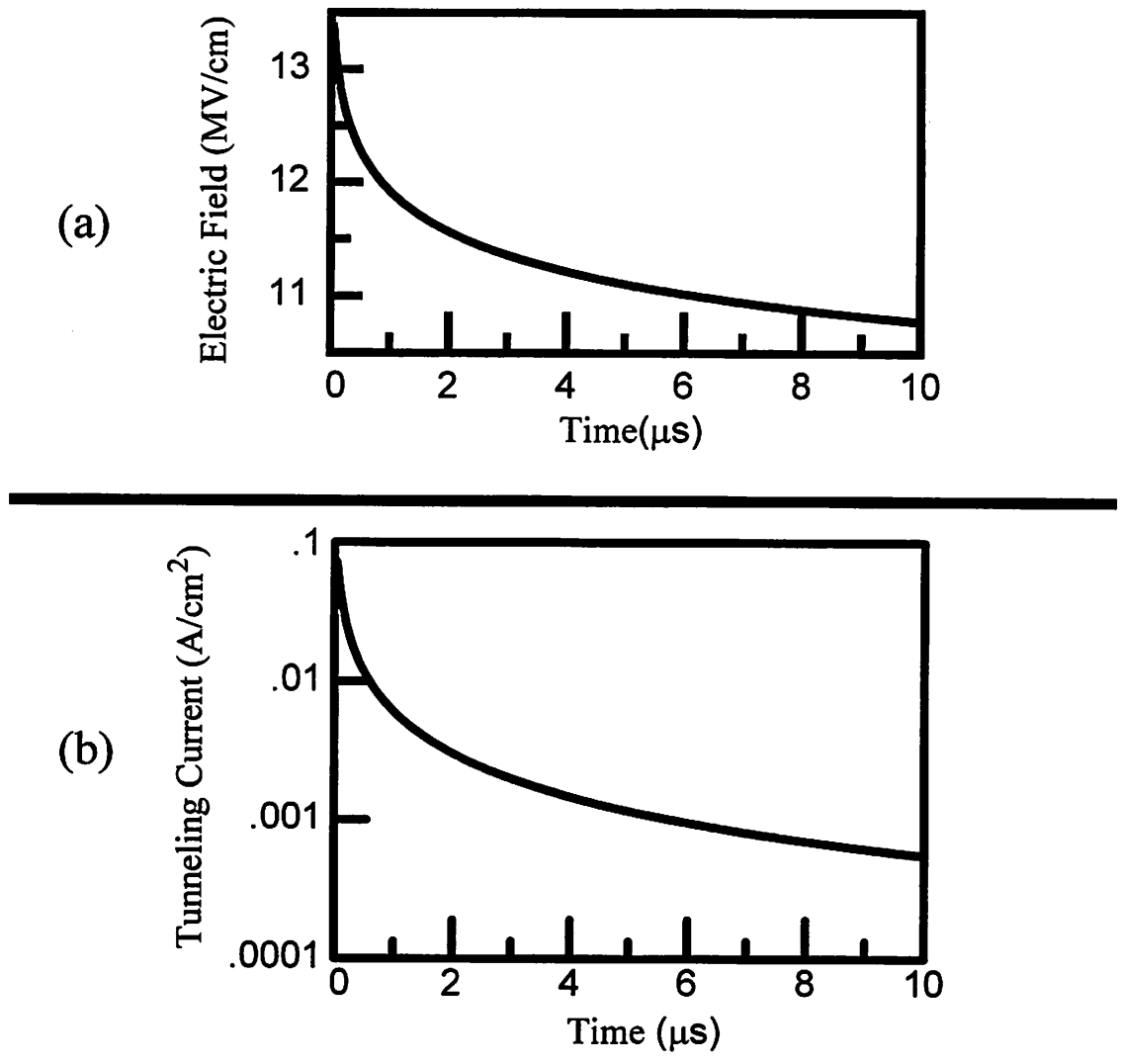

Figure 11-10 Simulated Tunneling Current and Gate Voltage with Antennas

Simulation for an instantaneous charge deposition of $10^{12}/cm^2$ across an antenna with a field oxide thickness of 200 nm and an antenna ratio of 100. The antenna is connected to a 5 nm gate oxide. (a) shows the electric field across the gate oxide, while (b) shows the tunneling current through the gate oxide. This tunneling current integrated over millions of pulses may cause charging damage.

11.6.3 Well Single Pulse Antenna Effect

In general, the single pulse antenna effect arises from surface voltage variations from varying capacitances across the wafer. In the previous sections, the capacitance variations were due to the different thicknesses of the field and gate oxide. A spatially varying V_s results in charge transfer from the low capacitive region (field) to the high capacitive

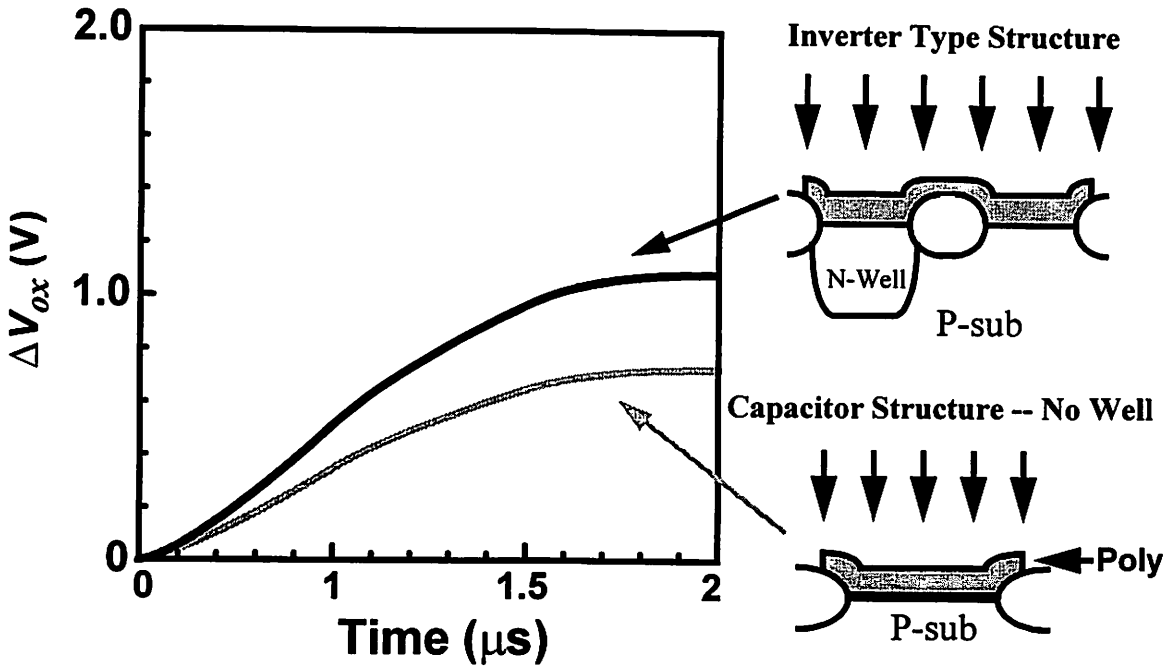

Figure 11-11 Well Antenna Effect

Change in V_{ox} during a single 5kV, 2μs pulse for a simple MOS capacitor and the NMOS part of an inverter. The well capacitance of the inverter generates an antenna like effect that enhances the NMOS gate oxide voltage, as compared to the simple capacitor. The simulation conditions are a with 100 nm gate oxides.

region (gate oxide) equalizing the voltage. These currents increase the voltage stress for the gate oxide. The well in the substrate also adds a capacitor in series with the gate oxide, with an effective capacitance:

$$C_{eff} = \frac{1}{\frac{1}{C_{ox}} + \frac{1}{C_{well}}} \tag{11-7}$$

where C_{ox}, C_{well}, and C_{eff} are the gate oxide capacitance, well junction capacitance, and the effective total capacitance, respectively. The difference in capacitance across the wafer due to wells is analogous to the spatial capacitance variation due to the field and gate oxide regions that generates the conventional antenna effect.

In the case of an N-well CMOS inverter, where two gates are connected together, charge flows from the PMOS/N-well gate (lower capacitive region) to the NMOS gate (higher capacitive region), increasing the stress for the NMOS oxide. In Figure 11-11, the well antenna effect nearly doubles V_{ox} for the NMOS gate as compared to an equivalent capacitor, significantly increasing the stressing voltage. Since the increased V_{ox} occurs over just one pulse, the well antenna effect is significant for generation rates slower than the pulse, nominally 2μs.

11.7 Conclusion

Gate oxide charging must be controlled during all implantation processes, including PIII. With plasma exposure, the wafer surface potential and open circuit voltage equals the plasma floating potential, usually a negative value. If V_f is large enough, oxide damage may occur with a simple plasma exposure. During PIII, the voltage on the surface of the wafer adjusts until the plasma electron current during the pulse off time balances the plasma ion and secondary electron currents during the pulse on-time. The faster the pulsing frequency, the more positive the equilibrium surface potential must be to attract enough plasma electrons. Since the initial equilibrium voltage is negative (it is V_f), as V_{oc} becomes more positive it must go through zero at some pulsing frequency. At this frequency, damage is minimized. This frequency is usually quite high, at above 25 kHz.

It has been shown through simulation that wells and the substrate type have a significant impact on the overall induced gate oxide stress. A depletion region protects an n-type doped channel oxide when $V_{ox} < 0$, and a p-type doped channel oxide when $V_{ox} > 0$. Compared to a structure without a well, an N-well oxide charges more negatively, while a P-well oxide charges more positively. The well effect depends on the leakage rate, but as long as the rate is slower than the pulse width, the wells will have some effect. The well effect is extremely sensitive to the well junction capacitance, and becomes smaller as the well capacitance increases. Experiments confirm the charging trend with

frequency and that n-doped channel devices exhibit less damage than their p-channel doped counterparts.

Damage may arise during single pulses when the gate oxide is connected to antennas. These antennas could be conventional metal or poly antennas, or antennas from well structures. The only ways to avoid single pulse antenna damage is to limit the amount of charge deposited per pulse, limit the antenna size, or provide a leakage path for the antenna through a connection to the substrate

Devices on dielectric or SOI substrates are expected to be immune to gate oxide charging damage during PIII. The dielectric protects the device by altering the capacitance of the system and preventing DC current flow.

Wells and substrate type can have an impact on oxide charging, and must be considered in the formulation of a global charging model. Through simulation and experiment gate oxide charging during PIII has been thoroughly investigated, with optimal pulsing conditions identified.

References

[11.1] S. Fang and J. McVittie, "Thin Oxide Damage from Gate Charging during Plasma Processing", *IEEE Electron Device Letters*, vol. 13, no. 5, pp. 288-290, May 1992.

[11.2] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *Journal of Vacuum Science and Technology B*, vol. 12, no. 2, pp. 833-837, 1994.

[11.3] E.H. Nicollian, and J.R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc., New York, pp. 332, 1982.

[11.4] C.N. Berglund, "Surface states at steam grown silicon-silicon dioxide interfaces", *IEEE Transactions on Electron Devices*, vol. 13, pp.701-710, 1966.

[11.5] E.H. Nicollian, and J.R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc., New York, pp. 319-372, 1982.

[11.6] W. En. B.P. Linder, and N. W. Cheung, "Modeling of Oxide Charging effects in Plasma Processing," *Journal of Vacuum Science and Technology B*, vol. 14, no. 1, pp. 552-559, 1996.

[11.7] W. En, S. Bell, B. P. Linder, and N. W. Cheung, "Effect of Gate Oxide Thickness on Charging Damage in PIII," *2nd International Symposium on Plasma Process-Induced Damage,* pp. 161-164, 1997.

[11.8] K. Scheugraf and C. Hu, "Effects of Temperature and Defects on Breakdown Lifetime of Thin $Si0_2$ at Very Low Voltages," *IEEE Transactions on Electron Devices*, Vol. 41, no. 7, pp. 1227-1232, July, 1994.

[11.9] M. J. Sherony, A.J. Chen, K.R. Mistry, D.A. Antoniadis, et. al., "Comparison of Plasma-Induced Charging Damage in Bulk and SOI MOSFETs", *1995 IEEE SOI Conference Proceedings*, pp. 20-21, 1995.

[11.10] S. Krishnan, S. Aur, G. Wilhit, and R. Rajgopal, "High Density Plasma Etch Induced Damage to Thin Gate Oxide", *1995 International Electron Device Meeting*, pp. 315-318, 1995.

12 Conclusion

12.1 Introduction

This work has developed two different models for plasma processing The first model, the Coupled Plasma Model, predicts the implant energy distribution for the Plasma Immersion Ion Implantation process. The second model, the Universal Charging Damage model, predicts gate oxide charging damage during plasma processing. Combining both models predicts and allows minimization of charging damage during the PIII process.

12.2 Coupled Plasma Model

The Coupled Plasma Model (CPM) has a wide array of plasma processing applications, but in this work is applied to PIII. The basic framework merges three different subsections: the plasma model, the wafer structure model, and the substrate bias model. The plasma model encompasses the plasma ion current density (J_{ion}), plasma electron current density (J_e), secondary electron current density (J_{se}), and displacement current density (J_{disp}). The plasma model consists of a set of physically derived differential equations with no fitting parameters; a Langmuir probe measurement provides all the necessary variables: plasma potential, plasma floating potential, electron temperature, and ion density. The wafer structure model is a translation of the device structures into circuit equivalent devices. Finally, the substrate bias is modeled as a voltage or current source with parallel and series non-ideal elements.

12.2.1 Fully-Coupled and De-Coupled Models

The model may be solved in a fully-coupled mode, with all three sections solved simultaneously. Conducting substrates (e.g. a bulk silicon wafer), allow a de-coupled mode, where the plasma ion currents and plasma sheath thicknesses are solved independent of the wafer structure model. The de-coupled mode affords a magnitude increase in computational speed. Two platforms have been used for the simulations, MATLAB and SPICE. SPICE allows easy incorporation of extra circuit elements, and is effective in the fully coupled mode. On the other hand, MATLAB contains more flexible differential equation solvers, and a better storage interface, but lacks the built in circuit models of SPICE.

12.2.2 Implant Energy Distribution Prediction

A pulsed PIII implant is not mono-energetic, but rather contains a significant energy spread. The CPM can predict pulsed PIII implant profiles. Implantation of the matrix sheath ions fundamentally limits the energy integrity of the implant. On top of that, the rise and fall times further increase the energy spread. Accounting for all the sources of energy spread, and understanding the limitations and the scaling trends with the implant variables (pulse width, pulse frequency, rise time, hold time, plasma ion density, and implant bias), allows the optimization of the implant.

12.3 Model Extensions

Adding modules to the CPM enables modeling of dielectric substrates, multiple species, and sheath collisions.

12.3.1 Dielectric Substrates

SOI structures and thin film transistor are fabricated on dielectric substrates. Implanting into dielectric substrates affects the implant in two ways. First, a portion of the bias voltage couples directly to the substrate, thus reducing the effective implant voltage.

Second, charge accumulates on the substrate during the implant, further reducing the implant voltage while introducing energy spread. The extended CPM allows optimization of the pulse width, pulse frequency, bias voltage, and plasma ion density to control substrate charging with an acceptable dose rate.

12.3.2 Multiple Species

Almost all plasmas sources for PIII contain multiple ion species. The formulation of an effective mass and an effective Bohm velocity enables the single species model to handle multiple ion species. This model is valid for short pulses up to infinite pulses (DC implantation). The model predicts that the ion implant ratios are a function of the pulse width. For short pulses, the ratios are equal to the ion density ratios. For longer pulses (and DC implants), the implant ratios are modified by the ion mass. Therefore, the implant ratios for the lighter ions are boosted at longer pulse widths.

12.3.3 Sheath Collisions

PIII implantation occurs at the same pressure as plasma generation (in the mtorr range). Typical PIII implants operate in a slightly collisional regime, which affects the implant energy profile, dose rate, and target current. Even implants at 1 mtorr and below suffer from ion sheath collisions. The effect of charge-exchange collisions is calculated with Monte Carlo simulation.

Starting with the basic plasma, wafer structure, and bias models, and adding the dielectric, multiple species, and collision modules forms a fairly comprehensive one dimensional PIII dose and implant simulator.

12.4 Universal Charging Damage Model

An undesirable by-product of all plasma processes is gate oxide plasma charging damage. The scaling of gate oxides to 3 nm and below has brought plasma charging damage to the forefront. A Universal Charging Damage model has been formulated to predict

plasma damage as a function of the oxide thickness, the antenna ratio, and the plasma condition. A load line analysis between the plasma impedance and gate conduction establishes the stress condition during the plasma process, while an oxide reliability model correlates the stress condition with oxide damage.

The simulation reveals three distinct oxide scaling regimes: constant voltage stressing, constant current stressing, and direct tunneling. In the constant current regime, plasma damage peaks, with thinner oxides suffering less damage. This peak, T_{Dmax}, is around 5 nm, but depends on the exact processing conditions. T_{Dmax} is a steep function of antenna ratio. Large ARs shift T_{Dmax} to thinner oxides, making large antenna data difficult to extrapolate to smaller, and more realistic antenna sizes. Reducing the electron temperature, open circuit voltage, and ion density all are effective ways of diminishing charging damage.

The simulation reveals two explanations for the divergent reports in the literature concerning oxide scaling and plasma damage. Many groups report decreasing damage with thinner oxides, while others report the opposite. First, the use of different ARs shifts T_{Dmax}. In addition, different plasma impedances alter the threshold thickness for damage. Plasma damage is much less significant if $t_{ox} > 10\frac{MV}{cm} \cdot V_{oc}$. Therefore, groups with different processing conditions will operate with a different damage threshold and T_{Dmax}.

The major predictions of the plasma charging damage model, the damage peak with oxide scaling, and the shifting of the peak have been experimentally confirmed.

12.5 PIII Charging Damage

Combining both the Coupled Plasma Model and the Universal Charging Damage Model allows the predication of charging damage in PIII. During PIII, the voltage on the surface of the wafer adjusts until the plasma electron current during the pulse off time balances the plasma ion and secondary electron currents during the pulse. If this time-aver-

aged DC equilibrium voltage is large enough, charging damage may occur. The faster the pulsing frequency, the more positive the equilibrium surface potential must be to attract enough plasma electrons. Since the initial equilibrium voltage is usually negative, as the oxide voltage becomes more positive it must go through zero at some pulsing frequency. At this frequency, damage is minimized. This frequency is usually quite high, above 25 kHz.

Circuit structures, such as wells, channel doping, circuit antennas, and dielectric substrates affect PIII charging damage. The surface equilibrates to nearly the same voltage for all conditions. Depending on the sign of the surface voltage a depletion region underneath the oxide may drop some of the voltage, protecting the oxide. Wells also affect charging, with oxides in N-wells charging more negatively, while oxides in P-wells charge more positively. The well effect depends on the well leakage rate, and may be negligible with high leakage.

Large area surface conductors, called antennas, can create AC, single pulse charging damage. The antennas collect and funnel charge down to the oxide amplifying the effective charge deposition density. With large antennas, enough charge can build-up during a single pulse to create charging damage. This fundamentally limits the total charge per pulse and throughput, especially at higher implant voltages. Simulations show that devices on dielectric or SOI substrates are generally immune to gate oxide charging damage during PIII.

Experiments confirm the effect of bias frequency on charging damage, that n-doped channel devices suffer less damage than their p-channel doped counterparts. The single pulse antenna effect has also been experimentally verified.

12.6 Future Work

The work in this thesis may be extended in two main directions. First, many of the PIII implant profile predictions based on the energy distribution need to be experimentally verified with Secondary Ion Mass Spectrometry (SIMS) profiles. The broadening effect of ion sheath collisions, and the relationship between the rise, hold, and fall times on the final profile need to be corroborated with experimental data. Equally important, the predicted implant ratios for multiple ion plasmas need verification. This is especially important for commercial applications in which almost every implant of interest will contain multiple species.

The second thrust concerns the integration of alternative gate dielectrics. Current literature predicts SiO_2 will reach it's practical scaling limit near 2 nm. Much research now seeks to identify alternative higher dielectric constant substitutes for SiO_2. The front runners are Si_3N_4, Al_2O_3, Ta_2O_5, and TiO_2. Plasma charging damage has to be re-evaluated for these different materials. This requires the replacement of two modules in the Universal Charging Damage model, the gate tunneling current module, and the oxide reliability module. The newer materials have different *I-V* relationships, but will definitely have lower leakage currents than thermal SiO_2. Currently, there is little or no reliability information, except for Si_3N_4 which shows promising trends.

With the current IC processing trends, the newer dielectrics may be more susceptible to plasma damage. The high leakage currents with ultra-thin SiO_2 oxides, reduce the stress voltage during plasma processing, the fundamental force behind the drop in damage with oxide scaling. The lower leakage currents with the replacement dielectrics will translate into higher voltage stressing with a minor decrease in the current stress (since the oxides are stressed in the constant current regime). The new dielectrics will be similar to a thicker SiO_2. In assessing plasma damage with the new materials, the increased voltage stress must be weighed against the change in oxide reliability. As more information

becomes available, the Universal Charging Damage Model will provide a framework for analyzing the new dielectrics.

Appendix A: Symbol Page

A	Area.
AR	Antenna Ratio which is the Antenna area divided by the gate area.
BOX	Buried oxide layer.
C	Generalized capacitance.
C	Percentage of ions that undergo a sheath collision before implantation.
C_{BOX}	Capacitance of the buried oxide found in SOI devices.
CDF	Cumulative Distribution Function.
C_{FOX}	Capacitance of the field oxide.
CPM	Coupled PIII Model.
CPM	Contact Potential Measurement.
CV	Capacitance Voltage curve.
C_{ant}	Capacitance of an antenna.
C_{depl}	Capacitance of the silicon depletion region beneath the gate oxide.
C_{eff}	Effective capacitance for a combination of parallel and series capacitances.
C_{hf}	High Frequency capacitance of a MOS system.
C_{it}	Interface trap capacitance.
C_{ox}	Capacitance of gate oxide.
C_{plasma}	Capacitance of plasma sheath.
C_{qs}	The Quasi-Static capacitance of a MOS system.
C_{qs1}	Undamaged Quasi-Static capacitance of a MOS capacitor.
C_{qs2}	Quasi-Static capacitance of the stressed MOS capacitor.
C_s	Plasma sheath capacitance.
C_{sub}	Capacitance of substrate. This is significant for dielectric substrates.

C_{well}	Instantaneous capacitance of the well junction, which is part of the well model.
DF	Duty factor.
DT	Direct Tunneling.
DTA	Device Transient Analysis.
Δ_{Vox}	Change in the gate oxide voltage from intial equilibrium.
D_{it}	Interface face trap density. The units are $cm^{-2}ev^{-1}$ or cm^{-2} depending on the context.
D_{well}	Name for diode in well model.
ECR	Electron Cyclotron Resonance
EEDF	Electron Energy Distribution Function.
E_{gain}	The energy of an electron after tunneling through the gate oxide.
E_k	Implant energy of particle k.
E_o	Gate oxide stress electric field during plasma processing.
E_{ox}	Electric field in the gate oxide.
F-N	Fowler Nordheim.
HF	High Frequency.
I	Generalized current.
I_{DT}	Direct tunneling current through the gate oxide.
I_{FN}	Fowler Nordheim tunneling current through the gate oxide.
I_{disp}	Plasma displacement current.
I_e	Plasma electron current.
I_{esat}	Plasma electron saturation current.
I_i	Plasma ion current.
I_o	Gate oxide stress electric field during plasma processing.
I_{sc}	Short circuit stress current to the gate oxide during plasma processing.
I_{se}	Secondary electron current.
I_{sh}	Current sinked by a shunt resistor in the matching network. This current drains power from the pulser, reducing the maximum amount of implant power.
I_{total}	Total current during a pulse

J_+	Total positive current density. This is the sum of the secondary electron and plasma ion densities.
J_{disp}	Plasma displacement current density.
J_e	Plasma electron current density.
J_i	Plasma ion current density.
J_n	Plasma neutral flux density.
J_o	Gate oxide current density in an undamaged oxide.
J_s	Gate oxide current density for the sample under test.
J_{se}	Secondary electron current density.
K	Secondary electron yield constant relating implant voltage and yield.
K_1	Materials constant for direct tunneling current calculations.
K_2	Materials constant for Fowler-Nordheim tunneling current calculations.
LCR	Leakage Current ratio for SILC measurements.
M	Ion Mass.
M_{eff}	Effective ion mass for a multiple species plasma
M_k	Ion Mass of ion species k.
N_{ch}	Channel doping concentration.
N_{sub}	Doping concentration of the substrate.
P2ID	Plasma Process Induced Damage.
PECVD	Plasma Enhanced Chemical Vapor Deposition.
PICMIC	Plasma Implantation Monte Carlo Analysis of Collisions module.
PIII	Plasma Immersion Ion Implantation.
P_{low}	Percentage of ions that implant with less then V_{pulse} during the pulse.
Φ_s	Potential of the Silicon/Oxide interface.
Φ_p	Barrier for hole tunneling.
QS	Quasi-Static.
Q_{bd}	Electron charge to breakdown of a gate oxide.
Q_{dep}	Total charge deposited per pulse.
Q_p	Hole charge to breakdown of a gate oxide.
Q_{well}	Charge in the well junction depletion region.
RIE	Reactive Ion Etching.

RTS	Random Telegraph Signal.
R_{sh}	Shunt Resistance Value in the pulse supply matching network.
SILC	Stress Induced Leakage Current.
SIMOX	Separation by IMplantation of OXygen.
SOI	Silicon on Insulator.
SPM	Surface Potential Measurement.
STA	Sheath Transient Analysis.
T	Temperature.
TDDB	Time Dependent Dielectric Breakdown.
T_{dmax}	The oxide thickness that sufferes the most plasma charging damage.
TFT	Thin Film Transisistor.
T_e	Electron temperature in electron volts.
V_{BD}	Voltage at which the gate oxide breaks down.
V_{depl}	Instantaneous voltage dropped across the silicon depletion layer.
V_{deplo}	The thermal equilibrium voltage maintained by the silicon depletion layer.
V_f	Plasma floating potential.
V_{fb}	Flatband of the MOS system.
V_g	Voltage applied to the gate in a CV sweep.
V_i	Energy of implanted ions.
V_{max}	Maximum voltage dropped across the sheath.
V_o	Instantaneous applied voltage, gate oxide stress voltage during plasma processing.
V_{oc}	Open circuit voltage stress to the gate oxide during plasma processing.
V_{ox}	Gate oxide voltage.
V_p	Plasma potential.
V_{pulse}	The maximum magnitude of the voltage pulse. This corresponds to the value of the pulse during the hold time of the pulse.
V_s	Instantaneous sheath voltage.
V_s	Substrate surface potential.
V_{sense}	Votlage at which the SILC is measured.
V_{shift}	Voltage shift of the Plasma Impedance.

V_{sub}	Voltage dropped across dielectric substrate.
V_{well}	Voltage dropped across the well junction.
a	Langmuir probe radius.
α	Hot hole production efficiency from tunneling electrons.
α_k	Percentage of ion k to total ion density.
b	Pre-factor constant in exponential fit.
d	Langmuir probe length.
δ	Implant energy spread. Defined as the differences between the implant energy at the onset of the hold time and the end of the hold time.
ε_0	Permittivity of free space.
ε_s	Permittivity of silicon.
f_p	Pulse frequency.
ϕ	Voltage in sheath as a function of distance from the wafer.
γ	Secondary electron yield per impinging ion.
i	Number of ions implanted.
j_k	Flux density of ion speices k
k	Boltzmann's constant.
k_e	Electron shadowing factor.
k_i	Ion shadowing factor.
λ	Electron mean free path in the oxide conduction band.
λ_i	Ion mean free path.
λ_{DE}	Electron debye length.
m	Electron mass.
m	Number of ion species in the plasma.
$m_{p,ox}$	Effective electron mass in silicon dioxide.
n	Number of neutrals implanted.
n_{diff}	Number of ions that diffuse across the sheath boundary.
n_e	Plasma electron density.
n_{expand}	Number of ions that are uncovered by the expanding sheath.
n_f	Number of ions that implant during the fall time. This is the sum of n_{fdiff} and n_{smax}.

n_{fdiff}	Number of ions that diffuse across the substrate during the fall time.
n_k	Density of ion species k.
n_i	Plasma Ion Density, or intrinsic carrier density in Silicon.
n_{matrix}	Number of ions in the matrix sheath.
n_o	Total bulk plasma ion density.
n_{s1}	Maximum spatial plasma ion density.
n_{s2}	Miminum spatial plasma ion density.
n_{smax}	Number of ions in the sheath when the sheath thickness is s_{max}.
Φ_b	Silicon to Silicon Dioxide barrier. Usually assumed to be 3.2eV.
q	Unsigned charge of an electron or ion.
s	Sheath width.
s_c	Steady state Child Law sheath thickness.
s_m	Matrix sheath thickness.
s_{max}	Peak sheath thickness during the pulse.
t_{bd}	Time to breakdown of a gate oxide.
t_c	Ion transit time across a steady state Child Law sheath.
t_{c2}	Ion transit time across a steady state Child Law sheath assuming no further acceleration.
t_f	Fall time of pulse.
t_{on}	Hold time of pulse.
t_{ox}	Thickness of gate oxide.
t_r	Rise time of pulse.
t_w	Pulse width.
u_b	Bohm velocity.
u_{bk}	Bohm velocity of ion species k.
u_{eff}	Effective Bohm velocity for a multiple species plasma.
v	Ion velocity.
v_e	Electron velocity.
v_s	Distributed sheath velocity. This is used instead of Bohm velocity for Electron Cyclotron Resonance plasmas.
x	Distance from substrate.

η	Coupling Efficiency. Defined as the percentage of the applied bias that couples to sheath as compared to the substrate.

Appendix B: Library Examples with De-Coupled Plasma Model

B.1 Introduction

As stated in Chapter 3, there are two methods of solving the plasma/wafer structure system under an applied bias, either fully coupled or de-coupled. The fully coupled method solves all currents and voltages simultaneously, while the de-coupled mode solves the differential equation for the plasma solution independent of the currents and voltages on the wafer. De-coupling the plasma and the wafer structures increases the calculation speed, since there are fewer simultaneous equations to solve self-consistently. The only assumption necessary with de-coupled method is that the surface voltage of the substrate is nearly equal to the applied bias. Stated another way, the substrate must be conducting, and therefore the de-coupled mode is not sufficient for dielectric substrates such as thin film transistors and silicon-on-insulator technologies. Since the plasma electron current is sensitive to fractions of a volt differences in the surface voltage of the wafer, it must always be solved simultaneously with the wafer structures.

Besides the immediate decrease in computational complexity, the de-coupled method allows for the storage of plasma solutions to be recycled many times with different wafer structures, further reducing CPU time. This process of storing plasma solutions is diagramed in Figure B-1.

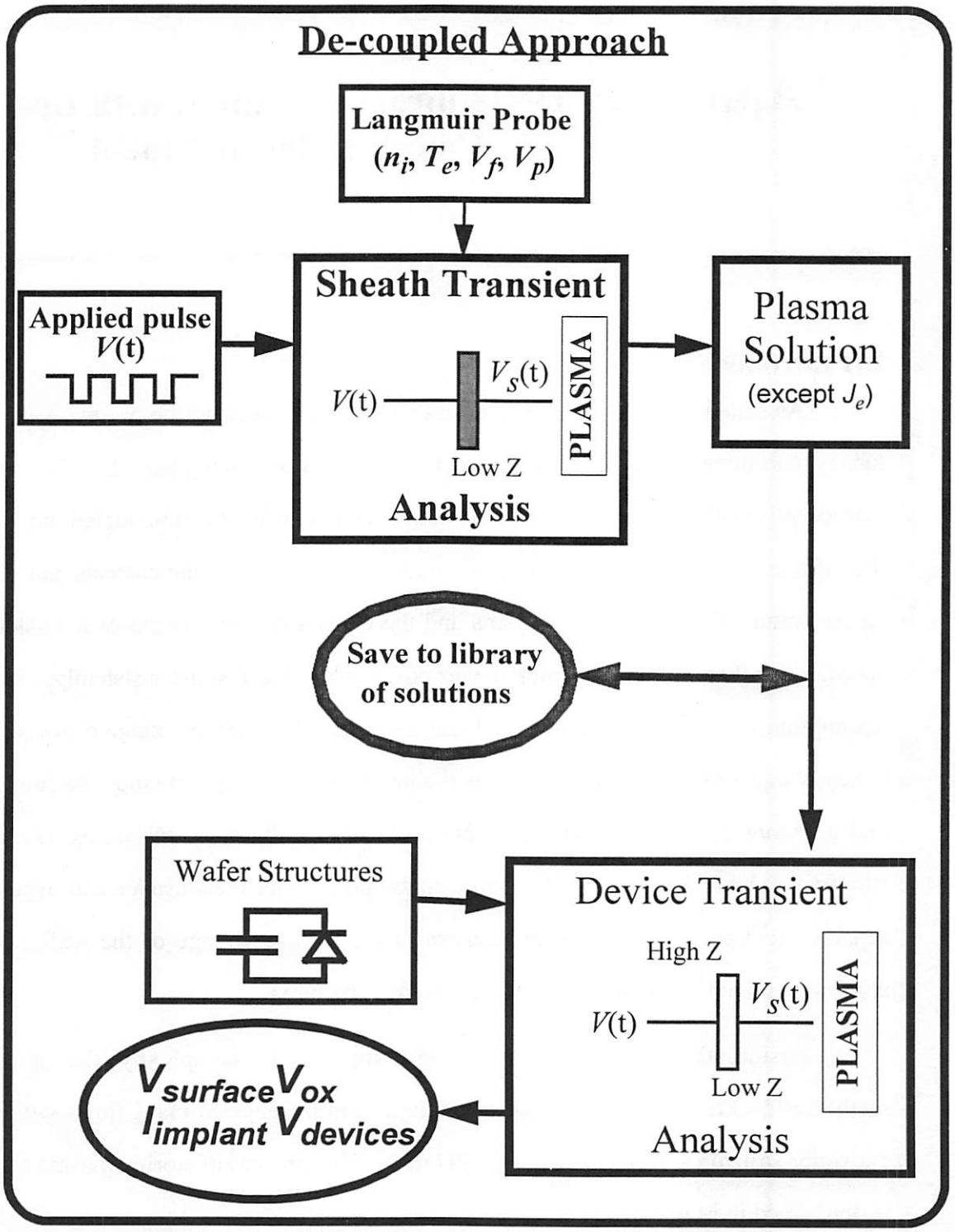

Figure B-1 De-Coupled Modular PIII Model

In the de-coupled approach, the sheath thickness and the plasma currents, except J_e, are solved independently of the wafer structures, allowing a magnitude increase in computational speed.

B.2 Library Examples

In order to illustrate the benefit of the library of solutions, I will step through an example for determining the effect of wells on gate oxide charging. The first step is to solve the plasma and sheath for each situation. The Sheath Transient Analyzer is fed the plasma characteristics and an applied bias. For this example $n_i = 5 \cdot 10^{10}/\text{cm}^3$, $V_f = -5.5$ V, $V_p = 13.23$ V, $T_e = 4$ eV, and the gas is argon. The applied bias is a -2 kV, 100 kHz pulse train. The plasma sheath solver output is the sheath thickness, plasma ion current, plasma displacement current, and secondary electron current as a function of time (Figure B-1). This solution is stored for later retrieval.

After the sheath portion has been calculated, the current and voltages on the wafer need to be computed. The inputs for the Device Transient Analyzer are the name of the file with the saved sheath solution, and the wafer structure models. For this simple example, the model will be a substrate with a gate oxide. The Device Transient Analyzer solves the current and voltages for the gate oxide, substrate, and the plasma electron current. The plasma electron current must be solved in union with the wafer structures, since it is extremely sensitive to small changes in the surface voltage, such as the voltage drop across a gate oxide. Figure B-2 plots V_{ox} for this system. For a full explanation of the time response of V_{ox} see Chapter 4. To compute the effect of a well structure on V_{ox}, the Device Transient Analyzer is given the name of the file with the stored sheath solution (the same filename as before), and the new wafer structure model including the P-well. With the results of the Device Transient Analyzer, the effect of the P-Well on gate oxide charging is shown in Figure B-2. As can be seen, the P-Well results in a larger change in voltage across the oxide than without a well. This effect is described in detail in Chapter 4.

In order to solve the effect of the well structure, the Sheath Transient Analyzer is only executed once. It is not necessary to solve the Sheath Transient Analyzer every time, which speeds up the total computational time. The effect of other device structures

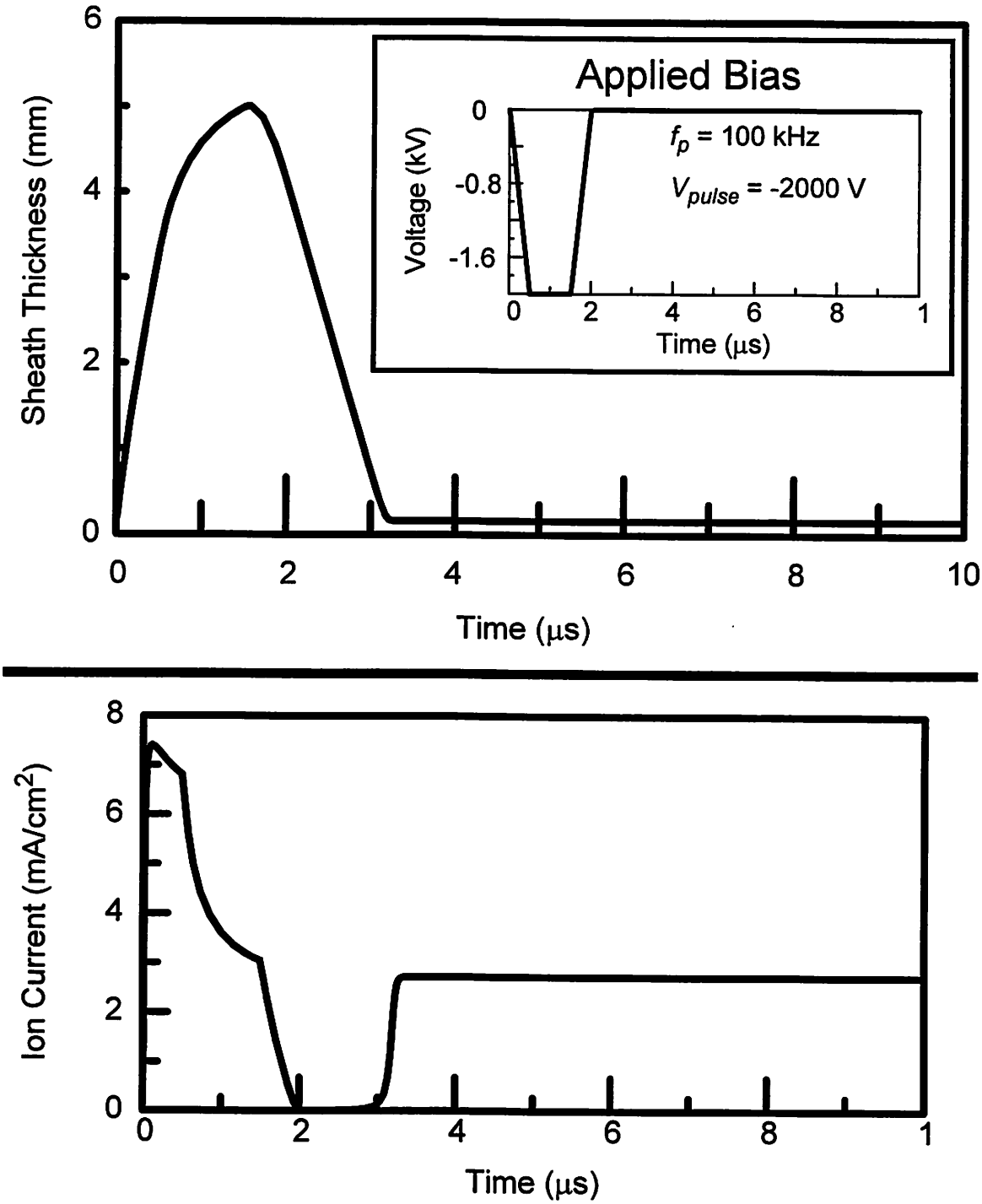

Figure B-1 Sheath Transient Analyzer Output

The Sheath Transient Analyzer output for a typical plasma condition (n_i= 5•10^{10}/cm^3, V_f= -5.5 V, V_p = 13.23 V, T_e = 4 eV) and a -2 kV, 100 kHz applied bias. The sheath solution shows sheath expansion to about 5 mm before the pulse ends. The ion current has a sharp initial peak, followed by a decay, reaching zero while the sheath is collapsing. This output is saved in a file for future use by the Device Transient Analyzer.

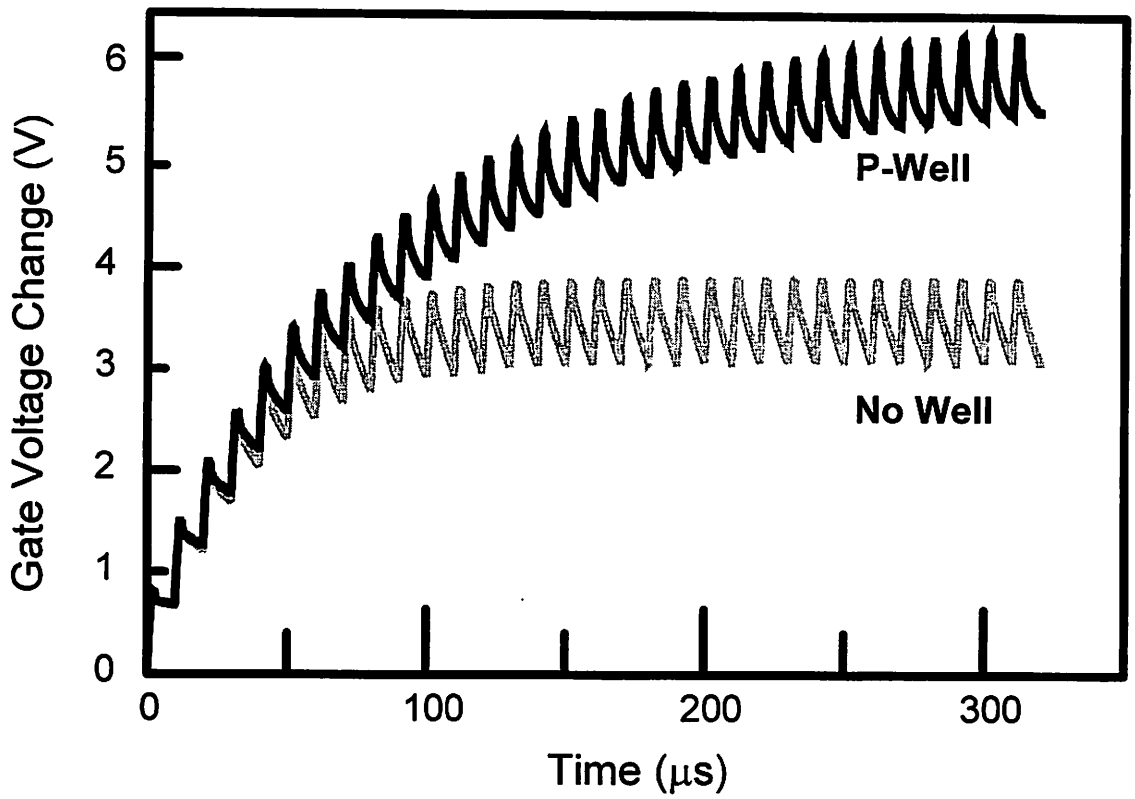

Figure B-2 Device Transient Analyzer Output

Some of the output information from the Device Transient Analyzer. This figure compares the change in gate oxide voltage during pulsing. The Sheath Transient Analyzer library solution and the device models are the input for the Device Transient Analyzer. The sheath is solved only once, and then referenced by the Device Transient Analyzer twice. This translates to a savings in computation.

could be investigated simply by inputting the new device structure models and the saved sheath solution into the Device Transient Analyzer.

It is possible, to create a library of solutions for different applied biases and plasma conditions solving the Sheath Transient Analyzer for each condition and saving the output in a file. Once the library is created, investigating the effect of different plasma conditions with different wafer structures is as easy as remembering the name of the saved sheath solution and inputting it into the Device Transient Analyzer.

B.3 Conclusions

De-coupling the computation of the sheath solution and the device transients saves considerable CPU time. The solution itself is simpler to calculate, since fewer equations are solved simultaneously, reducing the complexity of the problem. Secondly, by referring to the library of sheath solutions, the sheath only needs to be solved once, and then input into the Device Transient Analyzer. The combination of these two benefits reduces computation time, allowing investigation of more complex situations and the inclusion of more wafer structures, leading to a more complete picture of gate oxide charging.

Appendix C: Simulation Source Code

C.1 Source Code for Wafer Temperature During Implanatation

Main Program

```
%This program simulates the temperature rise during processing

% Need to calculate thermal mass for holder and wafer.
radius      = 2.54*2;;              %4 cm wafer
areaw       = pi*radius^2;          %Wafer area in cm2
volw        = 450e-4 * areaw;       %Wafer vol. Thickness*area
massw       = 2.32 * volw;          %Si density 2.32g/cm3
cpw         = .168;                 %cal/g/K, thermal constant
tmw         = cpw*massw*4.185;      %Thermal mass Si change to Joule.
width       = 1;                    %1cm thick holder
areah  = areaw + 2*pi*radius*width; %Holder area
volh        = width * areaw;        %Holder vol. Thickness*area
massh       = 2.7 * volh;           %Al density 2.7g/cm3
cph         = .215;                 %thermal constant Al, cal/g/K
tmh         = cph*massh*4.185;      %thermal mass of holder in  Joules
tottm       = tmh + tmw;            %thermal mass waf. and hold.

%Now calculate effective emissivity
%For silicon use data and interpolate
temp = [10 280 340 400 450 500 550 600 700 800]+273;
emsi = [.10 .12 .15 .22 .34 .52 .61 .68 .72 .72];

% Emmisivity of Al ranges from .06 to .2
eh          = .2;

%Simulation of temperature rise
powerin     = 20;                   %Implant Power (Watts)
sigma = 5.68e-8*1e-4;               % W/cm2 K4, radiation constant

timestep = 2;                       % step in between calculations (s)
endtime     = 900;                  %Total Implant timer
T(1)        = 302;                  %Inital temp in C
```

```
dT(1)          = 0;                          %initial temperature deriviative;

time = [0:timestep:endtime];

for t = 2:length(time)

 %Calculate effective emissivity

   ew = interp1(temp,emsi,T(t-1));          %interpolate for Silicon
   etot = eh*areah + ew*areaw;              %Total system emmisivity

  dT(t) = simriseeq(powerin,tottm,etot,sigma,T(t-1));
  T(t)  = T(t-1) + timestep*dT(t);

end;

T2 = T-273;    %convert to Celsius
plot(time,T2,'b')
```

Subroutine simriseeq

```
%This is the thermal balance equation used by simrise
%This gives the rate of change of temperature;

function m = simriseeq(powerin,tottm,etot,sigma,T)

                T0 = 300;
                m = ( powerin -  etot*sigma*(T)^4 + etot*sigma*T0^4)/tottm;
end
```

C.2 Source Code for Monte Carlo Collision Analysis

```
%This program sends in x particles.
%Accelerates them and then calculates nuetral flux and ion flux.

%This program assumes no neutral/neutral collsions
%Lambda is a function of ion energy.

p       = .5e-3;                  % p in mtorr
V0      = 10000;                  % Implant Voltage
s       = 10e-2;                  % Sheath thickness
ng1     = 1/(3.3e19 * p* 1000);   % neutral density (Argon)


%Need to caluculate size of neutral array.  Estimate using a fixed lambda.
```

```
lambda      = .05/(p*1000);

dist        = 100;
particles = 10000;
nb = 50;                                        % bin size in Volts for graphing


%Set the voltage distribution in the sheath
%Use either collisionless or fully collisionsal.
%Or calculate for moderatelly collisional

x           = [0:s/dist:s];
%V          = V0*(x/s).^(4/3);          %collisionless child law
V           = V0*(x/s).^2;              %fully collisional child law


volt        = ones(1,particles)*0;          %Energy in (Volts) for ions
neutralv    = zeros(1,particles*1.1*s/lambda); %Energy in (Volts of neutrals)
neutralni   = 0;                            %number of neutrals
zeropart    = zeros(1,particles);

% Simulation part

for step=2:dist+1;

 %1) Increase energies of ions
 %2) Calculate probability of colliding for each particle
 %3) Move collided particles to neutral array
 %4) Zero out energy of ions that collided

% step 1:  increase energy of ions

 volt       = volt + (V(step) - V(step-1));

% step 2: collision probability

 %1) cross-section equals
 %   = -8.7742e-22*(log(V).^2) - 4.0613e-20*log(V) + 5.8779e-18;
 %  This is a quadratic fit of cross section from Phelps, 1991.
 %2) calculate mean free path (mfp) = 1/ng*sigma;
 %3) calculate probability of collisions

 sigma      =  -8.7742e-22*(log(volt).^2) - 4.0613e-20*log(volt) + 5.8779e-19;
mfp         = ng1./sigma;               %ng1 is 1/ng
probcollide = 1 - exp(-s./(dist*mfp));
```

```
 prob          = rand(1,particles);
 collides      = find( prob < probcollide);

%Step 3: move collided ions to neutral matrix

 neutralv      = [neutralv(1:neutralni) volt(collides)];
 neutralni     = neutralni + length(collides);

%Step 4: zero out energy of ions that collided

 temp          = (max([(prob - probcollide); zeropart]) & 1);
 volt          = temp.*[volt];

end

%Simulation Complete, now calculate cumulative energy distribution function

%Seperate particle energy into bins

 %Ions

 voltbin = ceil(volt./(V0/nb));
 for x=1:nb
  energyion(x) = length(find(voltbin == x));
 end;

energyion = energyion./(particles*V0/nb);

%Neutrals

voltbinn = ceil(neutralv./(V0/nb));
for x=1:nb
 energyneutral(x) = length(find(voltbinn == x));
end;

energyneutral = energyneutral./(neutralni*V0/nb);

Vstep = [V0/nb:V0/nb:1000];

% Cdf is simply the volt array sorted,
% The y's are the y axis for the graphs
%Maxpart is number of ions that survive w/o collisions

ioncdf = sort(volt);
maxpart = min(find(ioncdf == max(ioncdf)));
ioncdf = ioncdf(1:maxpart);
```

```
iony = [1:maxpart]/particles;%generates values from 0 to 1 for cdf

neutralcdf = sort(neutralv);
neutraly = [1:neutralni]/neutralni;

totalcdf = sort([ioncdf neutralcdf]);
totaly   = [1:maxpart+neutralni]/(particles+neutralni);

%plot cdf's

plot(ioncdf,iony,'b',neutralcdf,neutraly,'r',totalcdf,totaly,'g');
title('B: ioncdf, R: neutralcdf. G:totalcdf')

%Final Statistics
%Aless is percent that don't collide
%Amore is percent that collide
%Amore2 is what Amore would equal if lambda was constant

Aless        = iony(maxpart)
Amore        = (particles-maxpart)/particles
Amore2 = exp(-s/lambda)  %theoretical Amore2
```

Appendix D: Secondary Electron Emission in a Collisional Plasma

The effect of ion-neutral sheath collisions is the topic of Chapter 7. The experimental measurement of ion-neutral sheath collisions is based on the fact that ion-neutral sheath collisions increase the total substrate current. Collisions create fast neutrals that implant into the wafer. The fast neutrals themselves do not contribute to the current themselves, but they do eject secondary electrons. Splitting the implant voltage among a number of neutrals and a single ion yields more secondary electrons, then implanting a single particle (ion) with the entire implant energy. This is shown below.

Since only ion-sheath collisions are considered, energy flux conservation insists that for a single ion crossing the sheath

$$\sum_{k=1}^{n+1} E_k = q \cdot V_{pulse} \tag{D-1}$$

where n, and E_k are the number of generated fast neutrals and implant energy of each particle, respectively.

The total number of secondary electrons ejected in the collisional case is proportional to

$$\sum_{k=1}^{n+1} \sqrt{E_k} \tag{D-2}$$

Squaring and expanding yields

$$\left(\sum_{k=1}^{n+1} \sqrt{E_k} \right)^2 = \sum_{k=1}^{n+1} E_k + 2 \sum_{k=1}^{n+1} \left(\sum_{l=k+1}^{n+1} \sqrt{E_k E_l} \right) \tag{D-3}$$

and

$$\sum_{k=1}^{n+1} E_k + 2 \sum_{k=1}^{n+1} \left(\sum_{l=k+1}^{n+1} \sqrt{E_k E_l} \right) > \sum_{k=1}^{n+1} E_k \tag{D-4}$$

Combining, Equation D-3, and Equation D-4 yields

$$\left(\sum_{k=1}^{n+1} \sqrt{E_k} \right)^2 > \sum_{k=1}^{n+1} E_k \tag{D-5}$$

Combining Equation D-1and Equation D-5 and taking the square root yields

$$\sum_{k=1}^{n+1} \sqrt{E_k} > \sqrt{q \cdot V_{implant}} \tag{D-6}$$

Equation D-6 states that the secondary electrons ejected by the implantation of an ion undergoing ion-neutral charge exchange collisions is greater than that ejected by an ion undergoing no collisions.

Equation D-6 may be summed over all the ions entering the sheath

$$\sum_{k=1}^{n+i} \sqrt{E_k} \geq i \cdot \sqrt{q \cdot V_{implant}} \qquad \text{(D-7)}$$

where n and i are the total number of generated fast neutrals, and the total number of implanting ions, respectively.

The crux of the above derivation, is the fact that secondary electron production is proportional to the square root of the particle energy. Equation D-7 may be generalized to a power law relationship

$$\sum_{k=1}^{n+i} (E_k)^y \geq i \cdot (q \cdot V_{implant})^y \qquad \text{(D-8)}$$

where y is the power law relationship between particle energy and secondary electron production. Equation D-8 is true as long as $y < 1$.