APPLICATION OF THE PLASMA IMMERSION ION IMPLANTATION COUPLED PLASMA MODEL FOR ENERGY SPREAD AND OXIDE CHARGING

by

Barry P. Linder

Memorandum No. UCB/ERL M96/98

19 December 1996

APPLICATION OF THE PLASMA IMMERSION ION IMPLANTATION COUPLED PLASMA MODEL FOR ENERGY SPREAD AND OXIDE CHARGING

by

Barry P. Linder

Memorandum No. UCB/ERL M96/98

19 December 1996

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

List of Figures

1 Introduction

Ion Implantation is one of the crucial steps in a semiconductor process flow. In current CMOS technology, there are at least five fabrication steps which utilize implantation: well formation, channel stop, threshold shift, dual poly implant, and source/drain implant. Thin film transistor technology adds two more implantation steps: hydrogenation and poly grain size control. Conventional beamline implantation excels at dose control and uniformity, but has low implant currents especially at low implant energies. Current trends of larger wafer sizes and large arrays of thin film transistors exacerbate the dose rate requirement. For instance, a $10^{16}/cm^2$ doping of a 500mm x 500mm substrate with a 10mA conventional implanter demands over 400 seconds/substrate, yielding a woefully low throughput. Plasma Immersion Ion Implantation (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 which immerses the substrate 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 is pulsed, and an off time follows each implant pulse, allowing the plasma electrons to neutralize the deposited positive charge.

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. The implant time for a 300 mm (12 inch) diameter wafer is the same as an 200mm (8 inch) wafer. This contrasts sharply with conventional

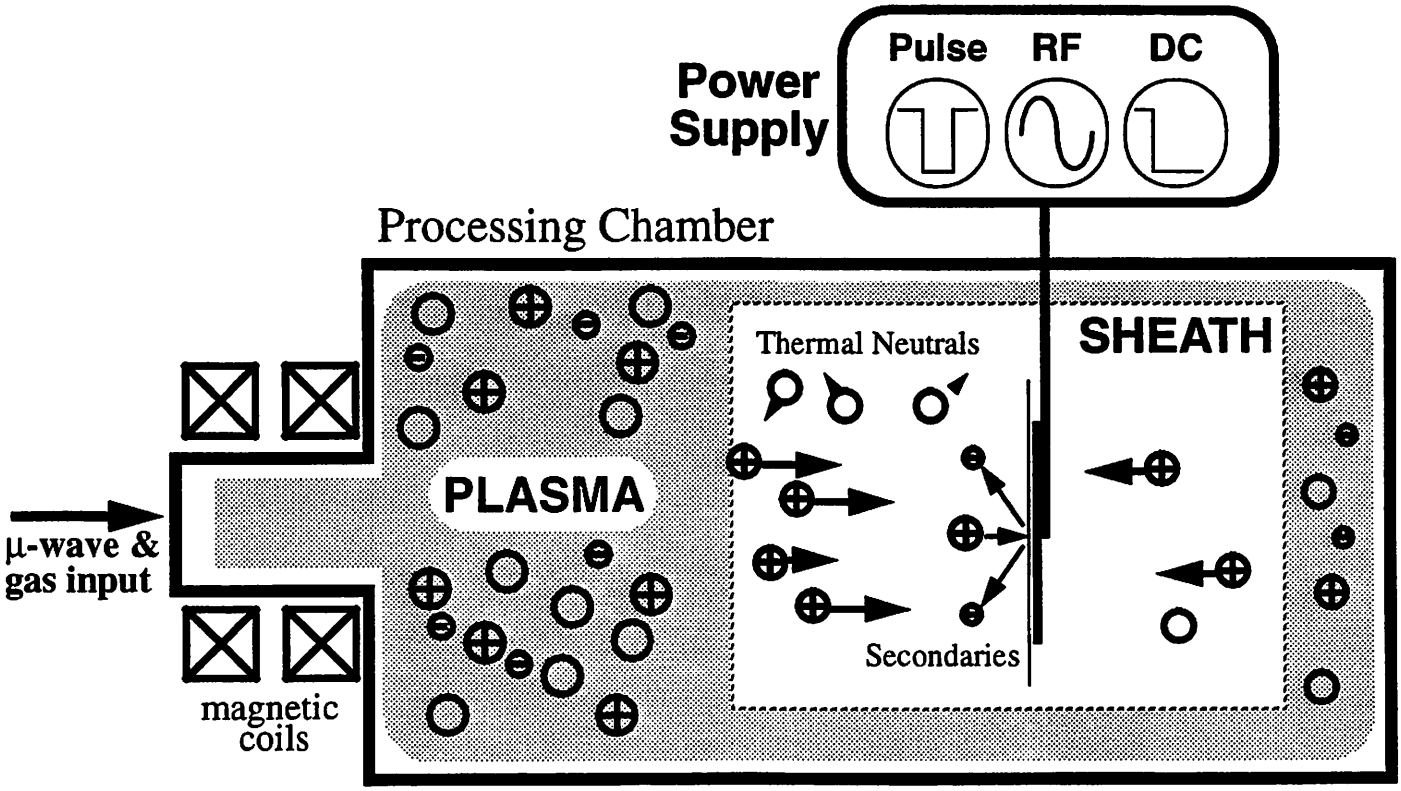


Figure 1-1 Concept of Plasma Immersion Ion Implantation

Diagram of Plasma Immersion Ion Implantation. An ECR 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.

implantation where the implant time scales with the square of the wafer radius (Figure 1-2). For comparison, typical PIII dose rates can exceed 1mA/cm^2 over the entire wafer, while an extreme high beam implanter current achieves 100mA, which when divided by the wafer area yields the current per cm^2 [1-1, 1-2]. To the first order, PIII would be faster for wafers larger than 4.5 inches.

Due to beam optics, conventional implantation currents deteriorate sharply at lower implant energies, below 10keV. 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 limita-

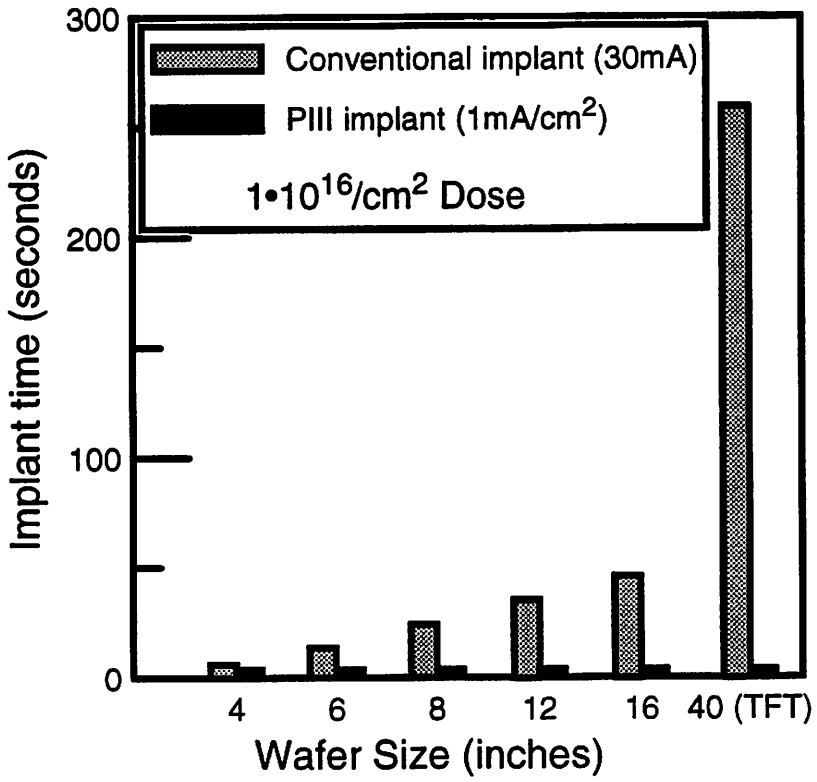


Figure 1-2 Implant Time Comparison

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

tion has led semiconductor manufacturers to look for alternatives for low energy implantation.

Beside dose rate, PIII has many 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 extremely flexible, fully scalable, and cluster tool compatible. Because the PIII machine is not specific to just implantation, it doubles as an etcher or a low temperature CVD. In principle, processes that include a pre or post implantation etch or deposition are possible all in one machine without breaking vacuum.

The applications currently under development are shallow junction formation [1-5 - 1-8], SIMOX formation [1-9 - 1-18], trench doping [1-19, 1-20], palladium doping for copper plating [1-21, 1-22], and metallurgical hardening by nitrogen implants [1-23, 1-24, 1-25].

A multitude of issues need to be investigated before full PIII implementation. Understanding the relationship between substrate bias voltage, implant energy and dose rate, the mechanisms of gate oxide charging, and the extent of the implant energy spread are main concerns. This requires 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.

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

References

[1-1]G. Ryding, T.H. Smick, M. Farley, et. al., "The IBIS 1000 SIMOX Production Implanter," to be published in Conference Proceeding of Ion Implant Technology Conference, 1996.

[1-2]K. Tokiguchi, T. Seki, K. Amemiya, and Y. Yamashita, "Advanced Microwave Ion Source for 100mA-class SIMOX Ion Implantation", to be published in Conference Proceeding of Ion Implant Technology Conference, 1996.

[1-3]S. Mofatt, "Life Below 1keV," to be published in Conference Proceeding of Ion Implant Technology Conference, 1996.

[1-4]L.M. Rubin, "Effect of Beam Energy Purity on Junction Depths in submicron Devices," to be published in Conference Proceeding of Ion Implant Technology Conference, 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 Elec. Dev. Lett.*, 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," in *Beam-Solid Interactions: Fundamentals and Applications, MRS Symp. Proc.* Vol. 279, ed. by M. Nastasi. Pittsburgh, PA: MRS, pp. 255-259, 1993.

[1-7] E. C. Jones, S. Im and N. W. Cheung, "Characterization of Ultra-Shallow P^+/N Diodes Fabricated using Plasma Immersion Ion Implantation" Implantation Technology - 92, ed. by D.F. Downey. Elsevier Science Publishers, pp. 373-376, 1993.

[1-8]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," in *Low Energy Ion Beam and Plasma Modification of Materials, MRS Symp. Proc.* Vol. 223, ed. by J.M.E. Harper. Pittsburgh, PA: MRS, pp. 115-119, 1991.

[1-9]J.Min, P.K.Chu, Y.C. Cheng, S.Im, S. Iyer, N.W. Cheung, "Fabrication and Characterization of SIMOX wafers by Plasma Immersion Ion Implantation." Conference Digest International Conference on Plasma Science and Technology 94 Chengdu China. pp 153, 1994.

[1-10]J.Min, P.K.Chu, Y.C.Cheng, J.B.Liu, S.Im, S.Iyer, 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, N.W. Cheung, "Synthesis and Characterization of SIMOX by PIII." *1995 International Conference on Materials and Process Characterization for VLSI (ICMPC 94)* pp. 238, 1994.

[1-13]J.Liu, S.S.Iyer, C.Hu and Nathan 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, pp.1010, 1995

[1-14]J. Min P.K. Chu, Y.C. Cheng, J.B. Liu, S.S.K. Iyer, 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, N.W. Cheung, "A New Development of PIII for SIMOX Structure." *Chinese Journal of Semiconductor*, 16 (B), pp. 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*, 67 (16), pp. 2361-2363, 1995.

[1-18]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.

[1-19]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.

[1-20]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.

[1-21]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.

[1-22] 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.* Symposium. Mater. Res. Soc, pp. 745-755, 1992.

[1-23]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.

[1-24]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.

[1-25]5. Malik, S.M.; Muller, D.E.; Sridharan, K.; Fetherston, R.P.; and others, "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):1015-19, Feb. 1, 1995.

2 Plasma Immersion Ion Implantation Setup and System

2.1 Berkeley PIII Reactor

A schematic of the PIII reactor developed in the Berkeley Plasma Assisted Lab 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, minimizing wall losses and improving plasma uniformity. A standard dual system of a diffusion pump and turbo pump with automatic crossover attains a base pressure near 1 microtorr. A 1500 watt ASTEX 2.45GHz microwave source supplies the power for plasma generation. The microwaves travel through waveguides, a 3-stub tuner, and finally couple to the machine through a quartz glass. Two electro-magnets, in a mirror configuration, generate the required magnetic field of 875 gauss for electron cyclotron resonance at 2.45GHz. The wafer holder handles wafers up to 12 inches in diameter, and slides from anywhere between 20 cm and 45 cm away from the source chamber. Wafers are either loaded from the top or from the back door. Theoretically, any gas source can be used as an implant source, but currently BF_3, He, Ne, SF_6, N_2, H_2, SiF_4, CF_4, 0_2, H_20, and Ar are available on the system.

2.2 Diagnostic Tools

A variety of diagnostic tools are available, and have access to the machine through 10 side and 9 back portholes. A baratron measures pressure 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

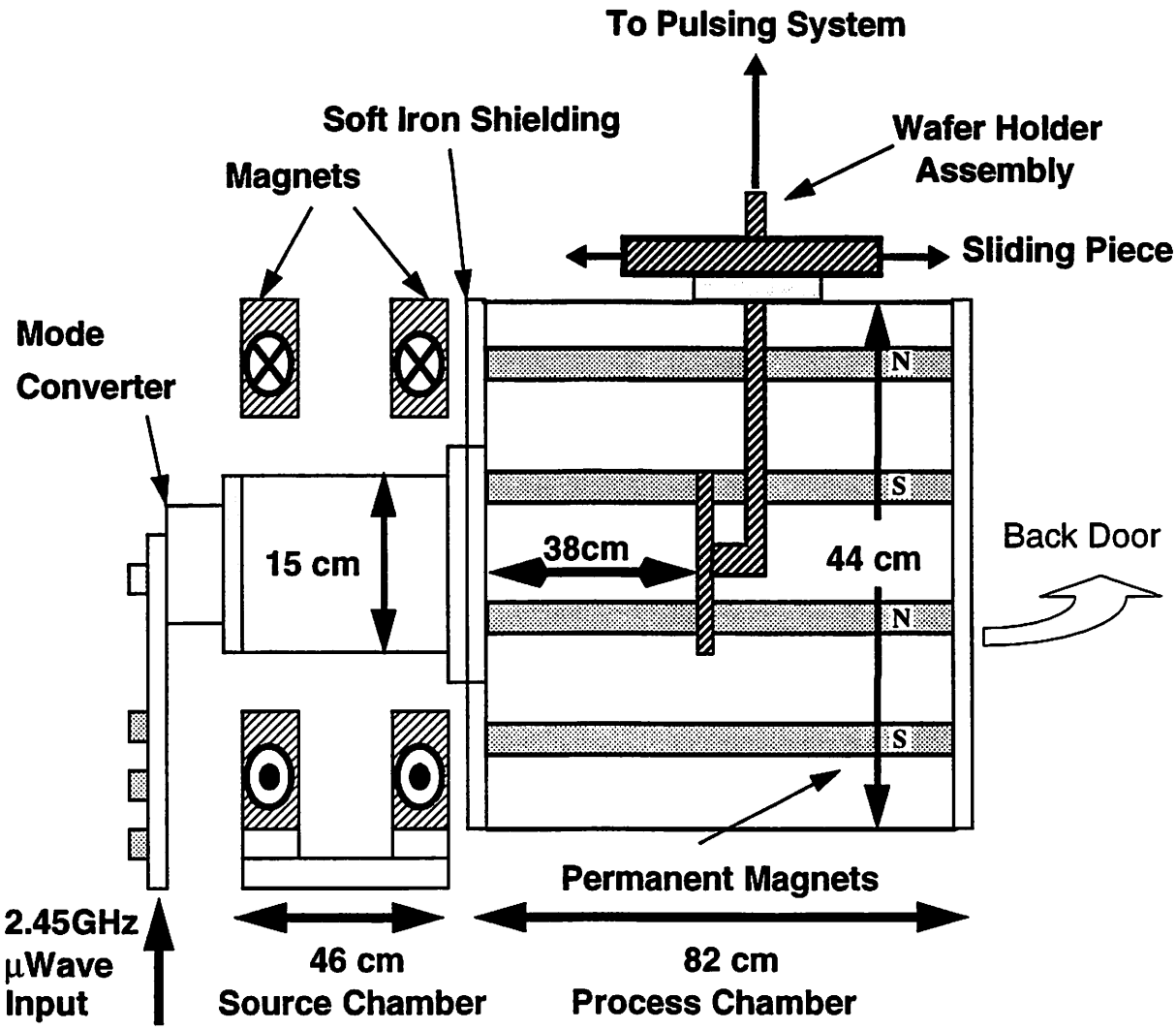


Figure 2-1 Schematic of Berkeley PIII machine

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.

gauge reading from the turbopump. 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. A mass spectrometer maps the mass and energy of the ions, while a Langmuir probe is used to extract the electron temperature, electron density, plasma floating potential, and the plasma potential. Finally, an optical emission spectrometer

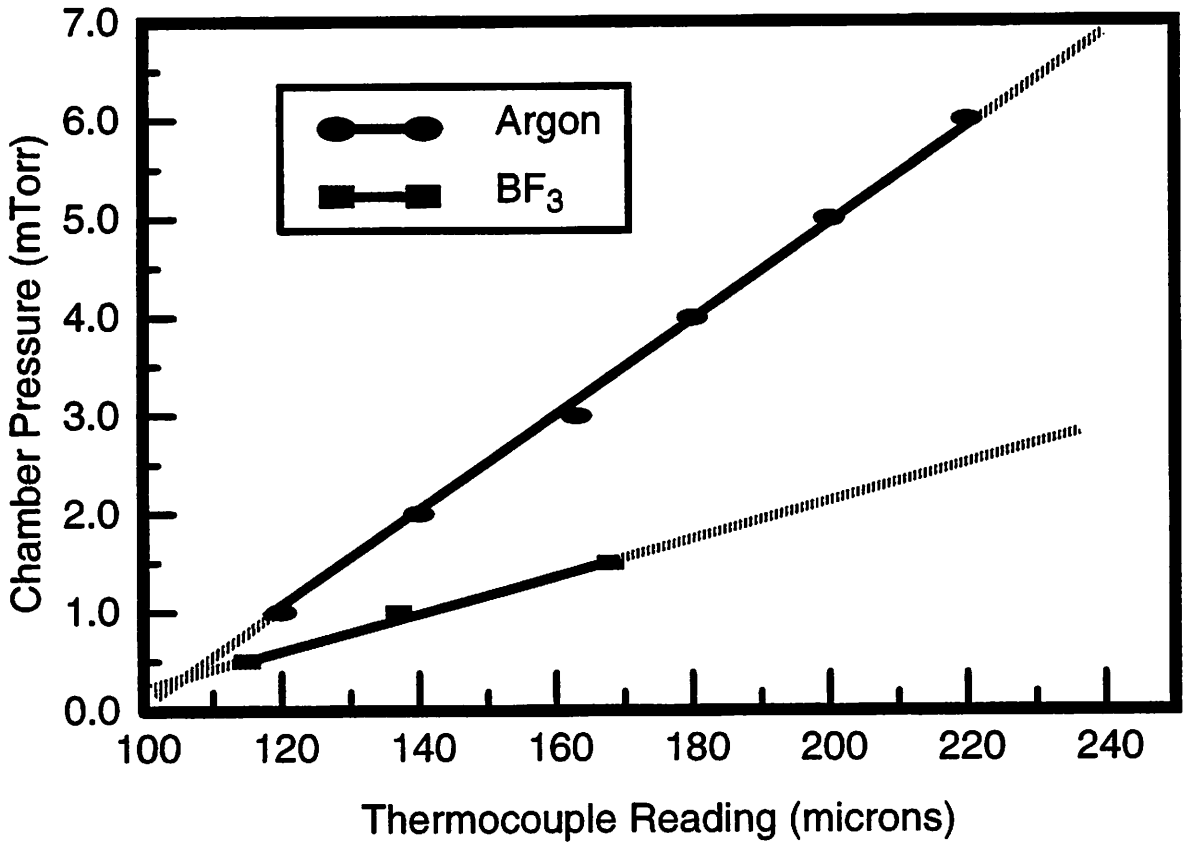


Figure 2-2 Thermocouple Pressure Calibration Curve

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.

plots the photon intensity as a function of wavelength, which can be correlated to the percentage of different ions in the plasma.

2.3 Pulsing System

Figure 2-3 depicts the PIII pulsing system. A 6kV/100mA power supply and a pulse generator connect to the 25kHz/6kV modulator. The signal travels across a transmission line containing various matching elements, terminating at the wafer holder. A high voltage probe connected to the wafer holder mechanism 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

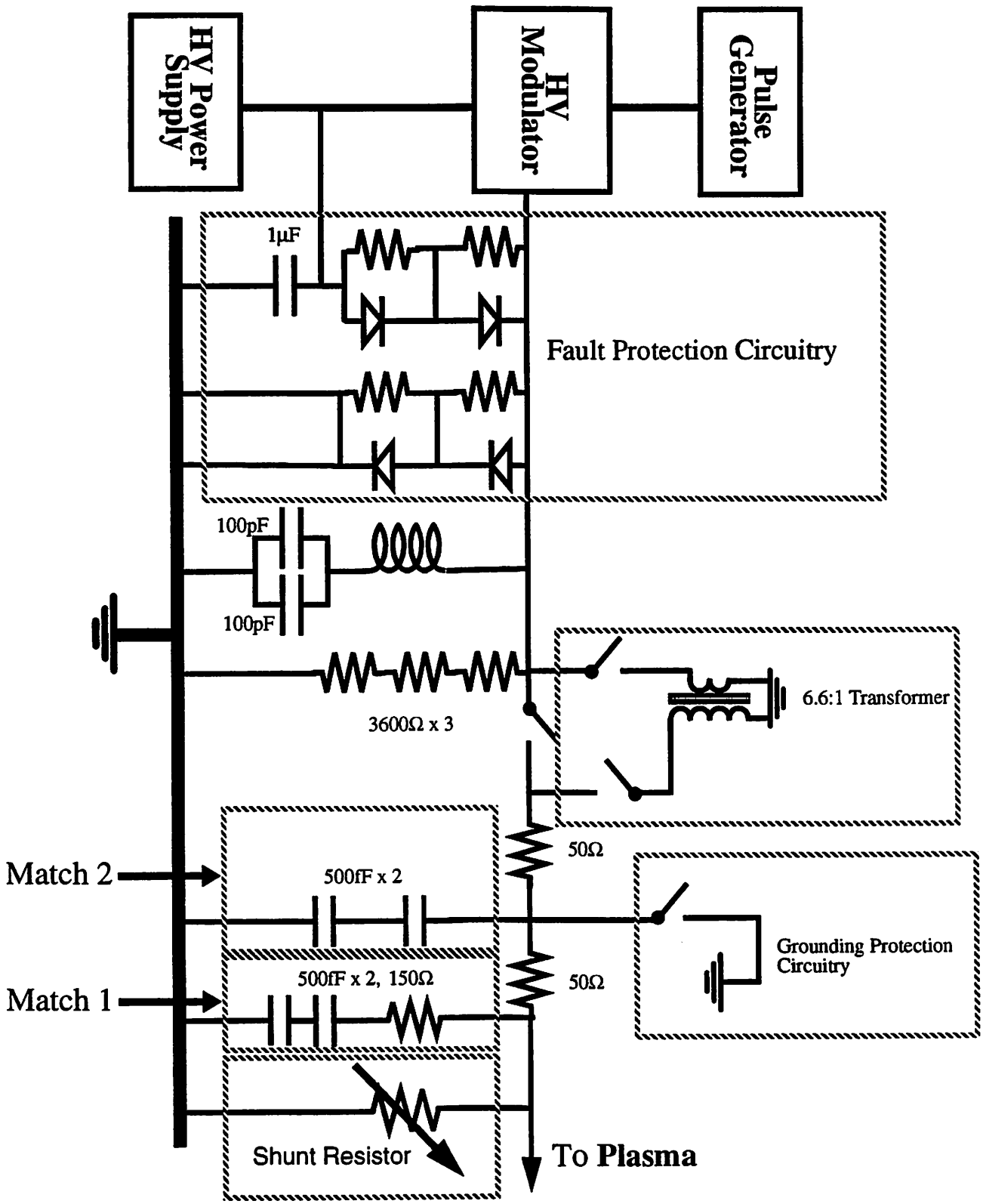


Figure 2-3 Pulsing Network

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 prevent 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.

circuits control the signal bounce. The modulator performs best with a 50 Ω terminating impedance, but since the impedance of the line depends on the plasma parameters, the matching network can not be optimized for all conditions. Typically, with the matching network shown, the fall time (defined as the time until voltage decays to 200V) is ~14.2 microseconds. When the modulator shuts off, it turns into an open, not a short to ground, forcing the capacitors to discharge through the 10kΩ resistor (3 x 3.6kΩ) or the plasma. Adding a shunt resistor diminishes the discharge time significantly. In Table 2-1 the shunt resistor reduces the fall time from over 10μs to below 1μs. The 50Ω resistor performs

Table 2-1 Shunt Resistor Effect on Pulse Fall Times

Shunt Resistor (Ohms)	Match 1	Match 2	Fall Time (μs)
none	yes	yes	14.2
50 (only 3kV)	yes	yes	0.12
270	yes	yes	0.72
270	no	no	0.40

best, but for some reason the maximum pulse voltage is capped at 3kV. The 270Ω resistor, without either match network, performs next best but suffers from signal bounce.

The shunt resistor draws extra current that becomes restrictive at high pulsing frequencies.

$$I_{sh} = \frac{V_{pulse}}{R_{sh}} t_w f_p \qquad (2\text{-}1)$$

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. In light of the 100mA limit of the power supply, a 270Ω resistor wastes 50% of the total current capacity of the power sup-

ply for a 6kV implant at a duty factor of just 2.25% or 2.25kHz with 1μs pulses. Therefore, a short fall time must be balanced with maximum pulse frequency.

3 PIII Coupled Plasma Model

3.1 Introduction

To properly model PIII, the plasma, the IC structures, and the substrate bias must all be accounted (Figure 3-1). The plasma model determines the plasma ion and electron currents to the wafer surface, and the ion impinging energies. The IC structure models calculate all the voltages and currents present on 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.

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.

To calculate the ion current, the sheath thickness as a function of time must be known. 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. The nominal implant pressure is 1 mTorr, which results in a low energy ion mean free path for Argon of ~3 cm,

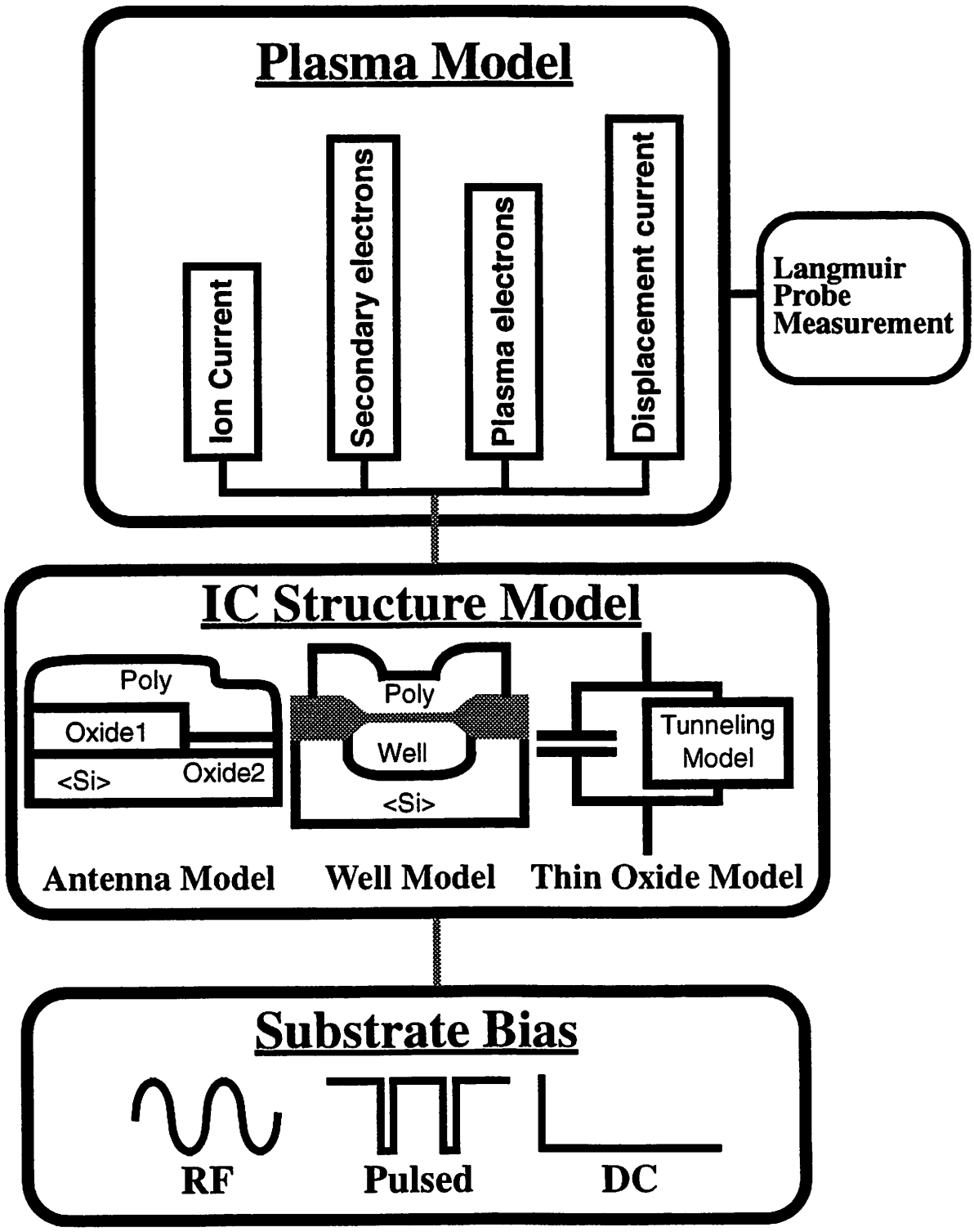


Figure 3-1 The PIII Model

The three sections of the PIII model: Plasma model, IC structure model and the Substrate bias. The Langmuir Probe provides all the parameters for the plasma model. The PIII model is inherently modular, and accommodates more complicated structures simply.

and a mean free path of ~14cm at 10keV [3-8]. With a typical ion density ~10^{10} cm^{-3} the maximum sheath width for a 10kV / 1μs pulse is ~2.4 cm. Since the sheath width is typically less than the mean free path, the sheath is assumed collisionless. During the implant pulse, the ions travel across the sheath in ~100ns, less than the nominal 1μs pulse width, allowing for the accelerating field to be assumed frozen during transit. Combined, these two assumptions imply that the ions bombard the surface with the instantaneous bias potential. A Quasi-static Child Law sheath is assumed to exist for the entire pulse duration, since it forms in the order of tens of nanoseconds, much less than the pulse width.

By applying these assumptions, the governing equations 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 distributed sheath velocity for ECR plasmas.

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

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

Solving Equation (3-4) for the sheath width, and plugging this result into Equation (3-2) determines the ion current as a function of time passing through the substrate.

Implanting ions with high voltage ejects secondary electrons, which the large sheath potential accelerate away from the wafer surface, thereby amplifying the total positive current,

$$J_+ = J_i(1+\gamma(V_i)) \tag{3-5}$$

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]:

$$\gamma \approx k\sqrt{V_i} \tag{3-6}$$

where k is an empirical fit parameter. For Al, $k \approx 0.0696$. En [3-10] contains yields for other substrate materials. For Al, secondary electrons exceed the positive ions for voltages greater than 200V, while for a 10keV ion, γ is near 7. Therefore, secondary electrons dominate positive charge deposition, and must be included in considering gate oxide charging.

To simplify the plasma electron current, a single temperature Boltzmann distribution is utilized. With Boltzmann electrons, the plasma electron flux to a surface is:

$$J_e = \frac{1}{4}qn_i v_e e^{\frac{-(V_p - V_s)}{T_e}} \tag{3-7}$$

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 (3-7) is no longer valid, and J_e is capped at $\frac{1}{4}qn_i v_e$. One situation where this may occur is with dielectric implantation just after the fall time (see section 4.5 on page 42).

Lastly, both the displacement current due to the changing sheath potential and the displacement current due to the changing sheath capacitance are included.

$$J_{disp}(t) = C_s(t) \cdot \frac{d}{dt}(V_s(t)) + V_s(t) \cdot \frac{d}{dt}C_s(t) \tag{3-8}$$

where J_{disp}, C_s, and V_s are the displacement current density, sheath capacitance, and sheath voltage. Displacement currents emerge during large applied bias voltage swings, which occur 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) - (8) form the fundamental equations for the PIII Plasma Model.

To solve the above equations simultaneously requires measurement of 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 experimental work demonstrates that this plasma model accurately determines the plasma currents [3-1, 3-6].

3.2.1 Wafer Structure Models

Modeling gate oxide charging requires the combining of the plasma equations with the mathematical descriptions of the device structures [3-1, 3-6, 3-7]. Most structure models are built from simple models of resistors, capacitors, inductors, diodes, and transistors. The thin gate oxide model consists of a capacitor in parallel with the Fowler-Nordheim and Direct tunneling models (Figure 3-2), which are known to be the main charging damage mechanisms 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 with a parallel capacitor, which includes both junction and transit time capacitances.

3.2.2 Substrate Bias

The substrate bias is included in the model by specifying a voltage or current sources. Non-ideal source effects are easily included by adding in transmission lines, internal source resistors, transformers, and the like. All the simulations in this paper utilize an ideal source.

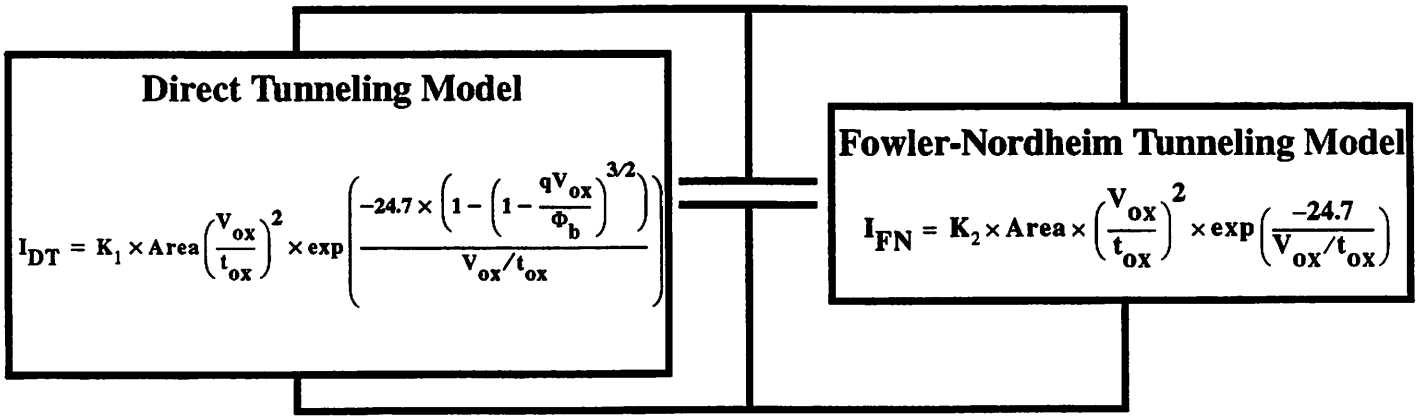


Figure 3-2 Gate Oxide Model

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.

3.3 Coupled and De-Coupled Models

Under the full implementation, all the elements of the model interact and all the currents and voltages are solved simultaneously (Figure 3-3). The only interaction between the wafer structures and the substrate bias with the plasma is through the surface voltage of the wafer. If it can be assumed that the surface voltage of the wafer is equal to the applied bias, then this full interaction is superfluous. In this case, the plasma currents, except for J_e, and substrate bias can be solved independently of the wafer structures (Figure 3-4). The plasma electron current must always be solved in conjunction with the wafer structures, since J_e is sensitive to fractions of a volt differences in surface voltage. The only situations that necessitate a fully coupled model are when the substrate contains thick dielectrics, typically greater then 5μm, which occur with buried oxide layers or dielectric substrates.

With the fully de-coupled model, the plasma currents, except for J_e, are independent of the wafer structures, allowing the creation of a library of plasma conditions and substrate biases. Then, for each wafer structure setup, the library is accessed for the

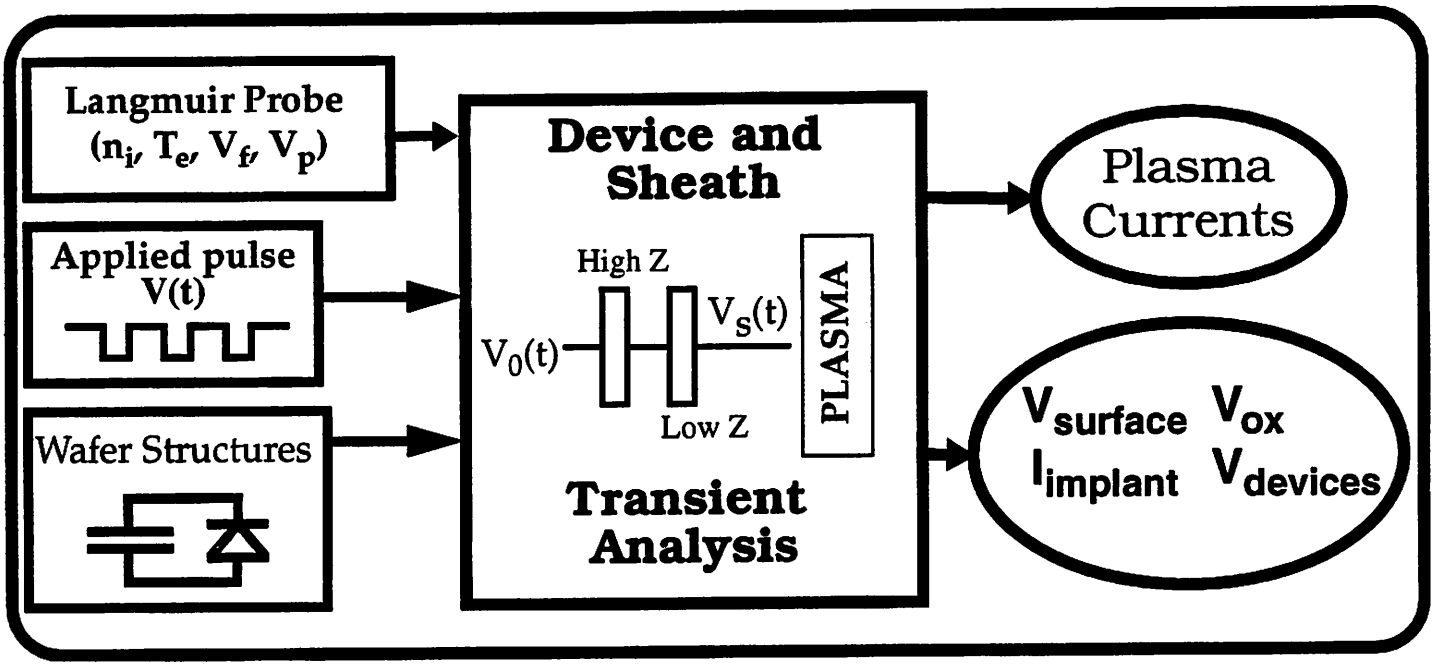


Figure 3-3 Fully Coupled PIII Model

The Fully Coupled PIII Model solves the sheath, plasma currents, surface currents, and surface voltages simultaneously. This is imperative when high impedance devices exist in the substrate.

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

We have used two different computer programs for implementing the PIII plasma model, the circuit simulator SPICE and the general purpose matrix solver MATLAB. Each program is better suited for different simulation conditions.

Since SPICE is a full circuit simulator, it already contains all the models necessary for the wafer structure models and the substrate bias. It is quite easy to add more devices or complexity to these sections of the model. On the other hand, the implementation of the differential equation in SPICE is cumbersome, and sometimes experiences convergence difficulties.

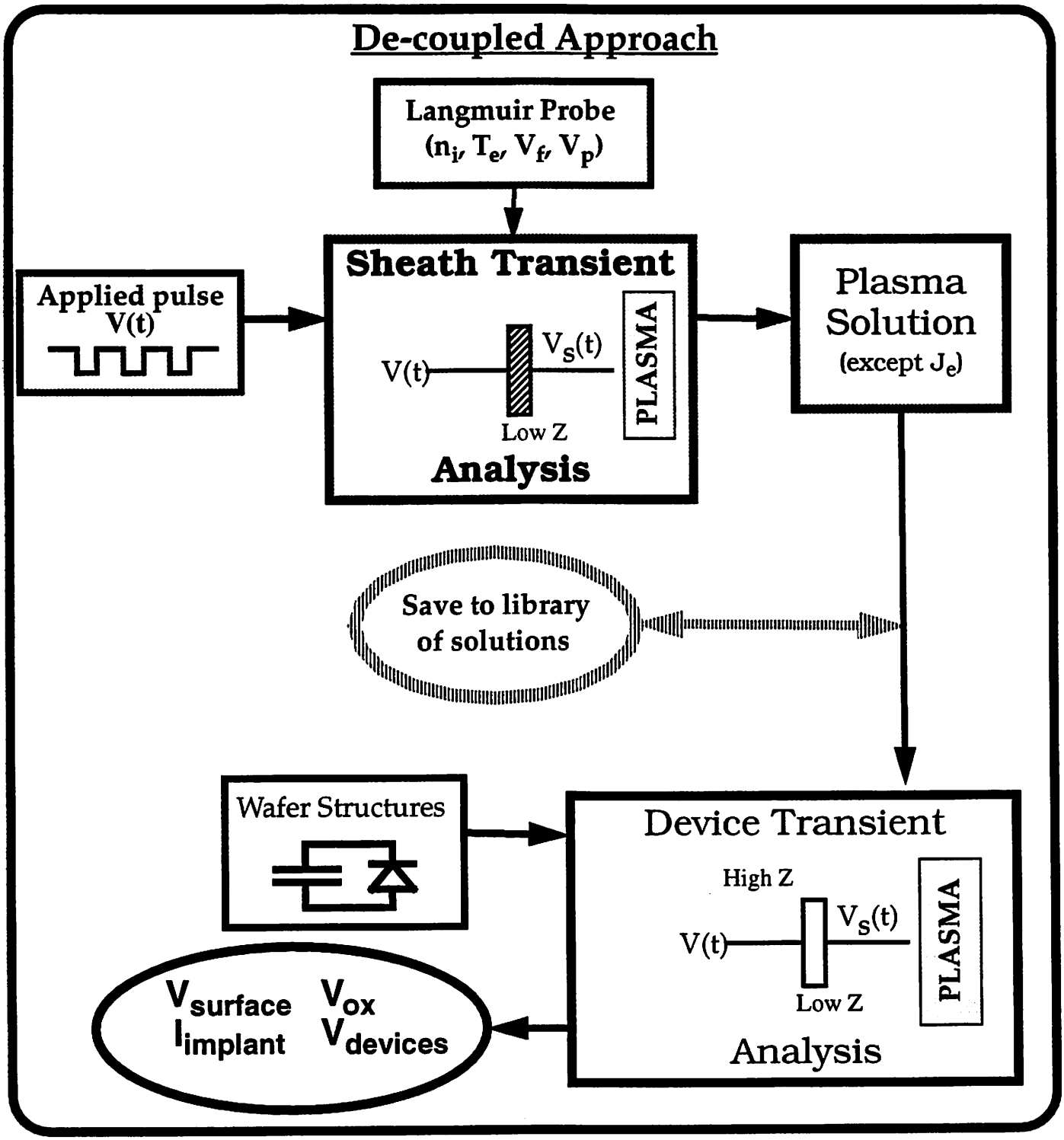


Figure 3-4 De-Coupled Modular PIII Model

In the de-coupled approach, the sheath thickness and the plasma currents, except J_e, are solved independently of the wafer structures, allowing a magnitude increase in computational speed. 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, allowing them to be computed only once. Then, for each different wafer structure set-up, the library is accessed for the plasma currents.

MATLAB contains an extensive library of differential equation solvers. If one method does not converge, it is trivial to switch differential solving methods. Since MATLAB does not include electrical models, these must be programmed in, which becomes laborious for complicated circuits. MATLAB also works best in the de-coupled mode, and has an extensive collection of file storage functions, that makes the construction of the plasma solution library seamless.

In all, 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

Combining a plasma model with models for the wafer structures and substrate bias forms a complete picture of PIII. The plasma model consists of physical equations, and contains only physical parameters supplied from a Langmuir Probe measurement. 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 applies to all 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 plasma model fully characterizes the 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 Implan-

tation", *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", *J. Applied Physics*, vol. 69, pp. 2904-2908, 1991.

[3-4] M. A. Lieberman, "Model of Plasma Immersion Ion Implantation," *J. 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," *J. 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," *J. Vac. Sci and Tech. 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," *J. 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" to be published in IEEE Trans. on Plasma Science 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] Emmert, G.A., "Model for Expanding Sheaths and Surface Charging at Dielectric Surfaces during Plasma Source Ion Implantation", *J. Vac. Sci. and Tech. B*, vol. 12, pp. 880-883, 1994.

4 Ion Implant Energy Spread in PIII

4.1 Introduction

Implant depth and profile prediction and simulation are necessary tools for users of implantation systems. With PIII, the unavoidable energy spread complicates profiling simulation. In contrast, conventional implantation has a minuscule energy spread. Assuming a collisionless plasma sheath, there are three main sources of energy spread in a PIII implant: the rise time and matrix sheath formation, the fall time, and voltage buildup on the substrate surface. Depending on the implant conditions either one of the three sources will dominate the energy spread. This chapter describes the sources of energy spread and some methods for estimating the implant energies.

The general definitions for a pulse implant are shown in Figure 4-1. Rise times are generally fast, and are less than 50ns for many pulsers. The fall time is usually considerably larger, and may range from less than a microsecond to tens of seconds, depending on the pulsing network. The on time for typical implants ranges from a microsecond to tens of microseconds. All the simulations in this chapter use an Argon plasma.

4.2 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

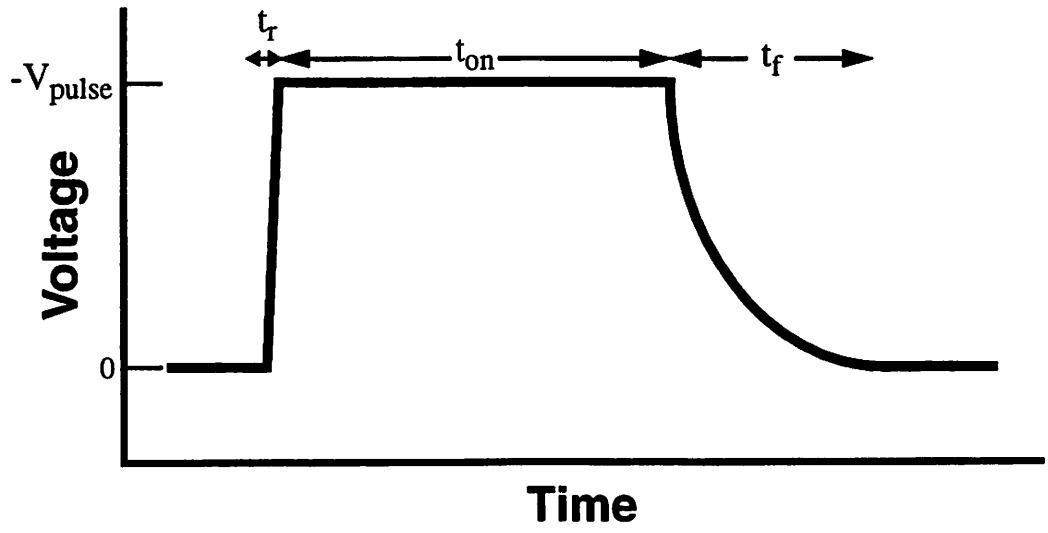


Figure 4-1 Definitions of Pulse Parameters

A typical voltage pulse showing the definitions of 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.

than 1GHz). Integrating Poison's equation twice, with uniform space charge, the width of the matrix sheath is:

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

where s_m is the matrix sheath thickness. Once the plasma density and the voltage pulse are known, s_m becomes constant. All the ions uncovered by the matrix sheath do not implant with the full energy (defined as V_{pulse}), but rather with the energy determined by the voltage distribution in the matrix sheath, which may be calculated from Equation (4-1) by substituting x for s_m, where x is the distance from the substrate, and V(x) for V_{pulse}. In order for an ion to implant with the peak energy, it must travel across the entire sheath. For example, an ion that happens to be half way between the edge of the matrix sheath and the substrate, will implant with $V_{pulse}/4$, rather than V_{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). Along the way, the sheath uncov-

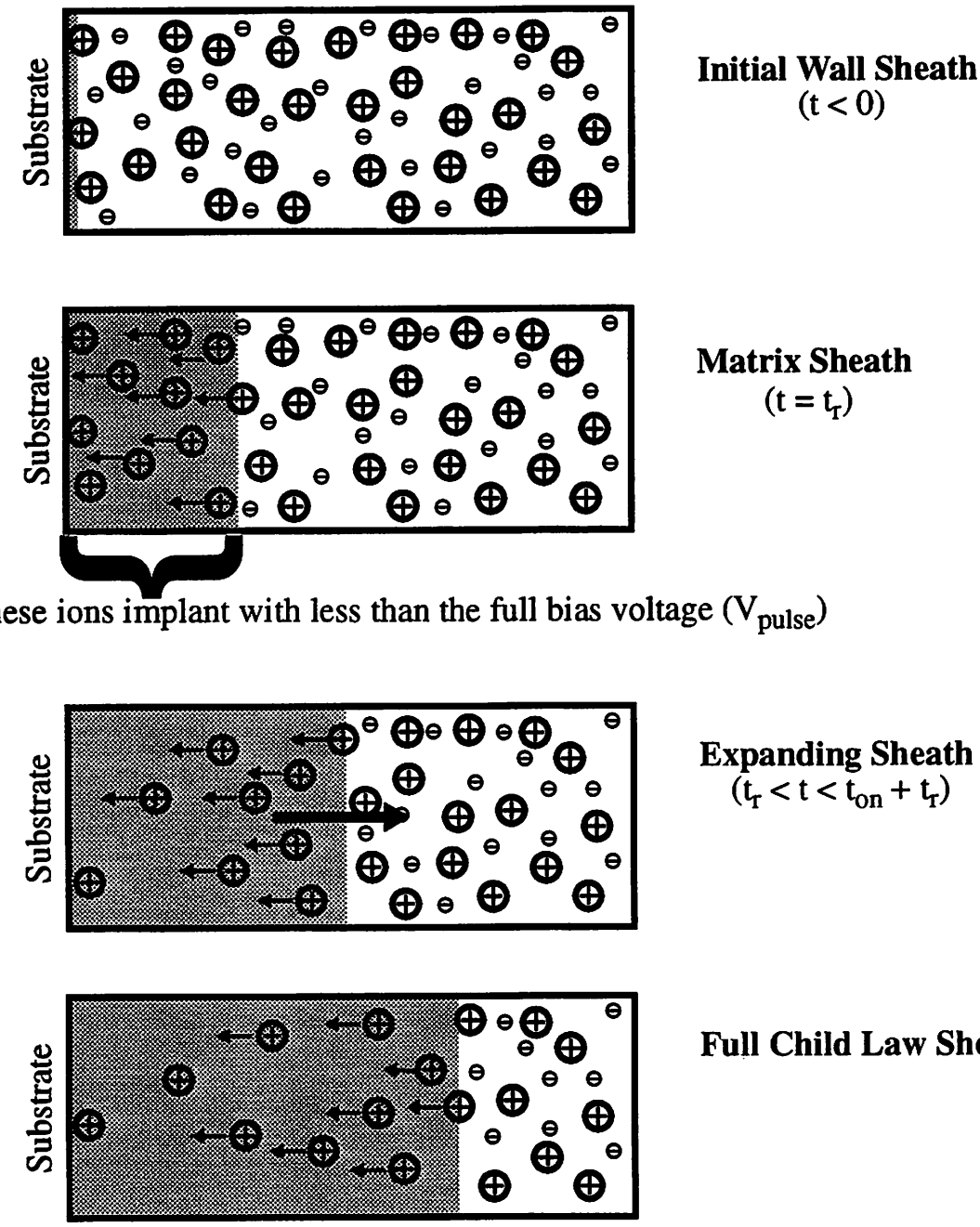


Figure 4-2 Stages of PIII: Matrix Sheath, Expanding Sheath, and Child Law Sheath

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 initial nearly instantaneous sheath formation, implant with less energy less than V_{pulse}. During the on time, the sheath expands out, uncovering more ions. Because these ions traverse the entire sheath (assuming a small transit time), they do 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 the full Child Law Sheath value, and stops.

ers more ions, which implant at the full pulse voltage, since they travel across the entire sheath (assuming the transit time is short compared to the sheath expansion rate). Lieberman [4-1] calculated the expansion rate for the sheath:

$$\arctan\left(\frac{s(t)}{s_c}\right) - \left(\frac{s(t)}{s_c}\right) = \frac{u_b t}{s_c} + \operatorname{atanh}\left(\frac{s_m}{s_c}\right) - \frac{s_m}{s_c} \tag{4-2}$$

where t, s(t), and s_c are the time, sheath thickness as function of time, and the full Child law sheath thickness. Since rise times are usually less than 50 ns, much less than normal pulse widths, they are assumed instantaneous, and the time in Equation (4-2) starts at the beginning of the pulse. A finite rise time would reduce the final sheath thickness slightly, but the effect is small. Another source of implanting ions are those that diffuse across the sheath boundary while the sheath expands. Because they traverse the entire sheath width, these ions implant with the full energy (V_{pulse}).

To calculate the percentage of ions that implant with less than the peak energy from the matrix 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:

$$P_{low} = \frac{n_{matrix}}{n_{matrix} + n_{expand} + n_{diff}} \tag{4-3}$$

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

$$n_{expand} = n_i s(t) \tag{4-5}$$

$$n_{diff} = n_i u_b t \tag{4-6}$$

where P_{low}, n_{matrix}, n_{expand}, and n_{diff} 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, and the ions that diffuse across the sheath boundary, respectively. At the end of the pulse on time, the ions that are still in the sheath will not be implanted with the peak energy, since the bias voltage will drop before they are implanted. These ions should be subtracted from n_{expand}.

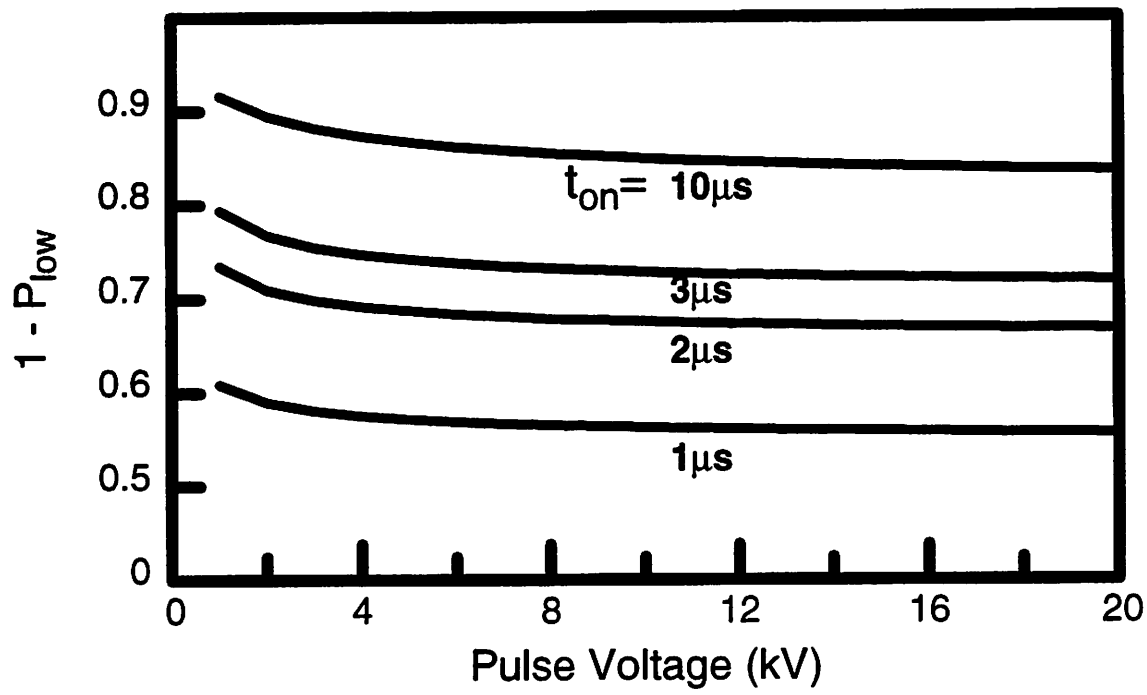


Figure 4-3 Low Energy Implantation from Matrix Sheath

This figure graphs $1\text{-}P_{low}$, which is the number of ions that implant with an energy 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 is $10^{10}cm^{-3}$.

Assuming a quasi-static Child Law sheath, the number of ions in the sheath at the end of the pulse on time is:

$$n_{smax} = \left(q \int_0^{s_{max}} \frac{4}{9} \frac{\varepsilon_o}{q} \frac{V_{pulse}}{(s_{max})^2} \left(\frac{x}{s_{max}}\right)^{-2/3} dx = \frac{4}{3} \frac{\varepsilon_0}{q} \frac{V_{pulse}}{s_{max}} \right) \qquad (4\text{-}7)$$

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. The maximum sheath width occurs near the beginning of the fall time. The sheath might continue to expand during the fall time until the sheath width exceeds the full steady-state Child law value for the dropping bias voltage. This phenomenon is more prevalent with short pulse widths, large implant voltages, and long times. This effect is negligible for most practical implant conditions, and is ignored.

Therefore, the value of s_{max} should be the actual maximum sheath width, not simply the sheath width at the commencement of the fall time

Figure 4-3 plots 1 - P_{low}, the amount of ions that implant with the peak energy (V_{pulse}) 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. Therefore, as the pulse widths lengthen, P_{low} decreases. As the implant voltage increases, n_{matrix} increases while n_{diff} is constant, resulting in a higher P_{low}.

Figure 4-4 shows how 1 - P_{low} changes with ion density. n_{matrix} is proportional to

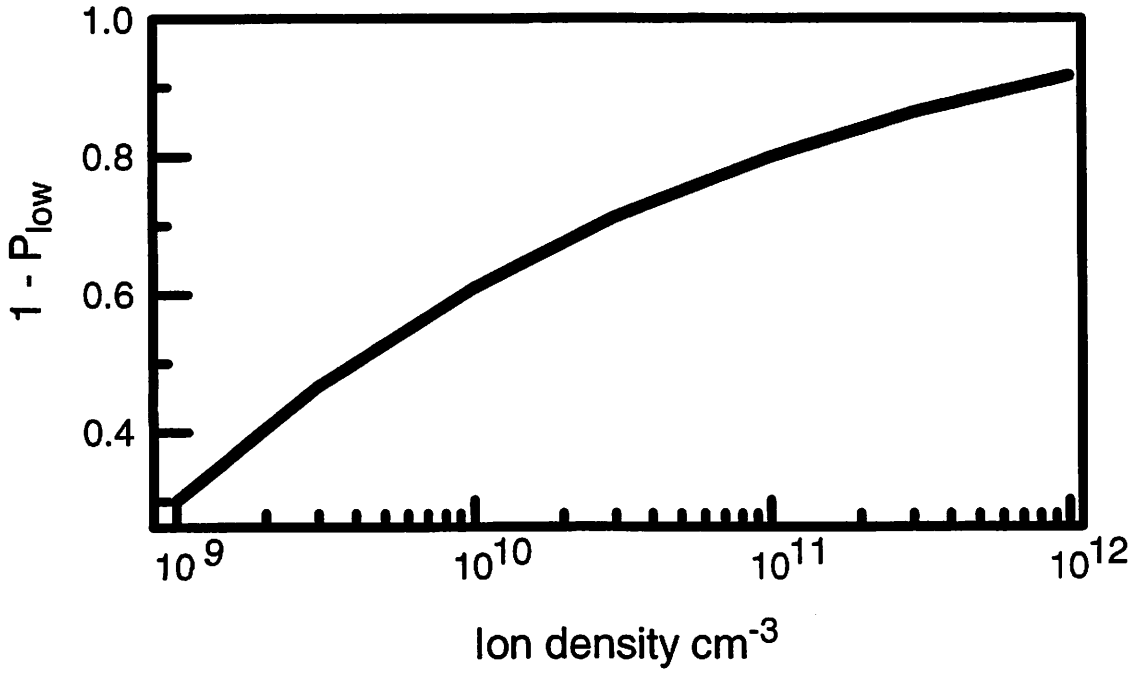


Figure 4-4 Energy Spread as a Function of Plasma Ion density

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

$(n_i)^{-1/2}$, but $n_{diff} \propto n_i$; so as the ion density increases P_{low} will decrease (Equation (4-3)). 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 quiescent matrix sheath contribution to the energy spread of a PIII implant. Overall, the quiescent matrix sheath theoretically limits the implant energy integrity of a PIII implant pulse.

4.3 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 the full energy (V_{pulse}). Since fall times can be comparable or actually longer than the on time, its contribution to low energy implantation can be significant. The charge implanted during the fall time is equal to 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 maximum sheath collapse rate is equal the ion diffusion velocity (the Bohm velocity), with the total collapse time equal to $\frac{s_{max}}{u_B}$ [4-2, 4-3]. If the sheath is collapsing 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 s_{max} / u_b. Shorter fall times do not decrease the energy spread, as long as the fall time is less than the inequality. (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 this inequality for a variety of implant times and voltages for a 10kV/1μs implant, the fall time simply needs to be less than 6.5μs to minimize the fall time effect on energy spread. For infinite pulse widths the maximum allowable fall time is simply s_c/u_b, since the sheath stops expanding at the full steady-state Child law thickness.

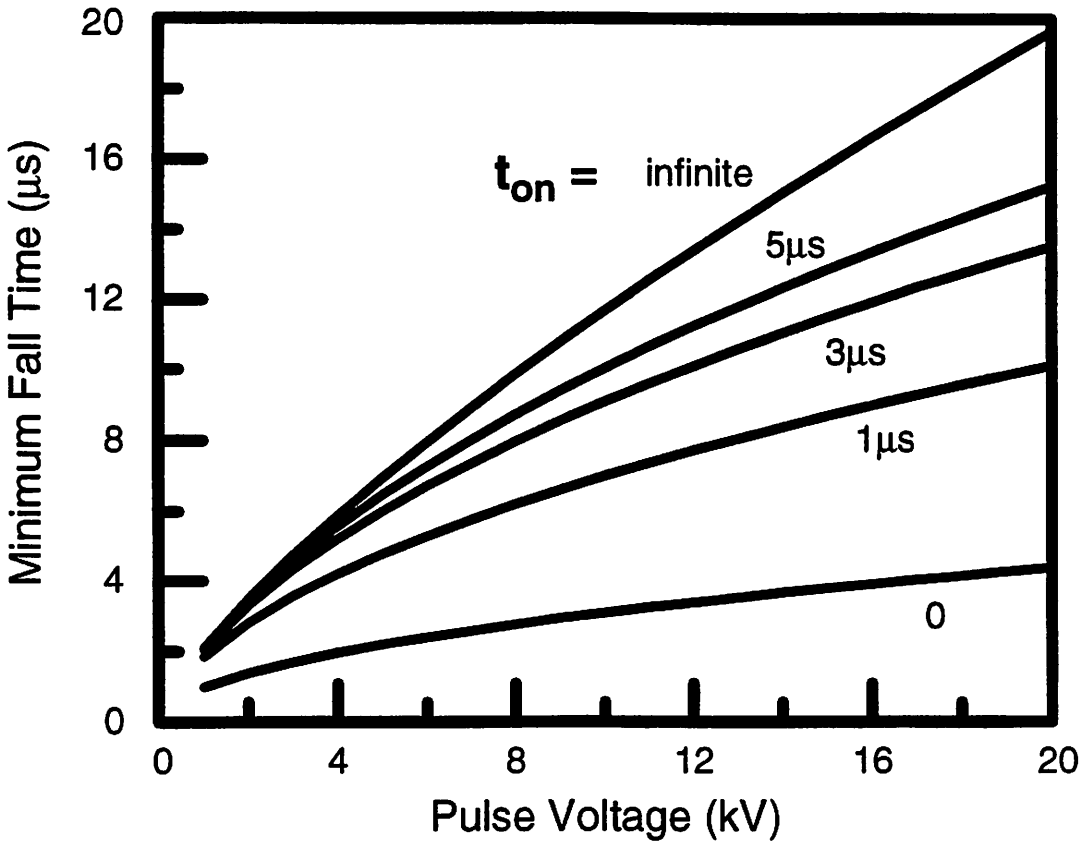


Figure 4-5 Sheath collapse time for different pulses

The minimum sheath collapse time for different pulse widths and pulse voltages. A fall time equal to the minimum fall time reduces the implant flux during the fall as much as possible. Any further reduction in the fall time has no effect. As the pulse width increases, the minimum fall time converges to the infinite on time curve. The ion density is $10^{10}cm^{-3}$.

Increasing the plasma ion density results in thinner sheaths, reducing the maximum fall time significantly. Therefore, higher ion densities require shorter fall times.

The energy profile of the ions can be estimated by three methods, depending on the time scale of the fall time. If the fall time is fast, e.g. $t_f \ll s_{max}/u_b$, then the implant energies for the fall time ions may be assumed to follow the Quasi-Static Child Law Relationship:

$$V(x) = -V_{pulse}\left(\frac{x}{s_{max}}\right)^{4/3} \qquad (4\text{-}8)$$

$$n(x) = \frac{4}{9}\frac{\varepsilon_o}{q}\frac{V_{pulse}}{(s_{max})^2}\left(\frac{x}{s_{max}}\right)^{-\frac{2}{3}} \quad (4\text{-}9)$$

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 is for long fall times or when $t_f \gg s_{max}/u_b$. 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 regime is the time period in between the two other time periods. For this range of fall time, the analytical relationship may be solved for simple cases [4-8], but is not tractable for more complex fall times. The profile may be simulated, though, by simply keeping tract of the ions as they traverse the sheath and the sheath edge in the spirit of Particle in Cell (PIC) simulations, or more simply by noting V_{sheath} when an ion enters the sheath, and assuming that is the actual implant energy.

It is significantly simpler to estimate the sum total of the ions that implant with less than the peak energy, with the same method applied to matrix sheath implantation. The ions that implant during the fall time are:

$$n_{fdiff} = \begin{cases} n_i u_b t_f - n_i s_{max} & \left(t_f > \frac{s_{max}}{u_b}\right) \\ 0 & \text{otherwise} \end{cases} \quad (4\text{-}10)$$

and

$$n_f = n_{fdiff} + n_{smax} \quad (4\text{-}11)$$

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 3μs pulse. This figure includes

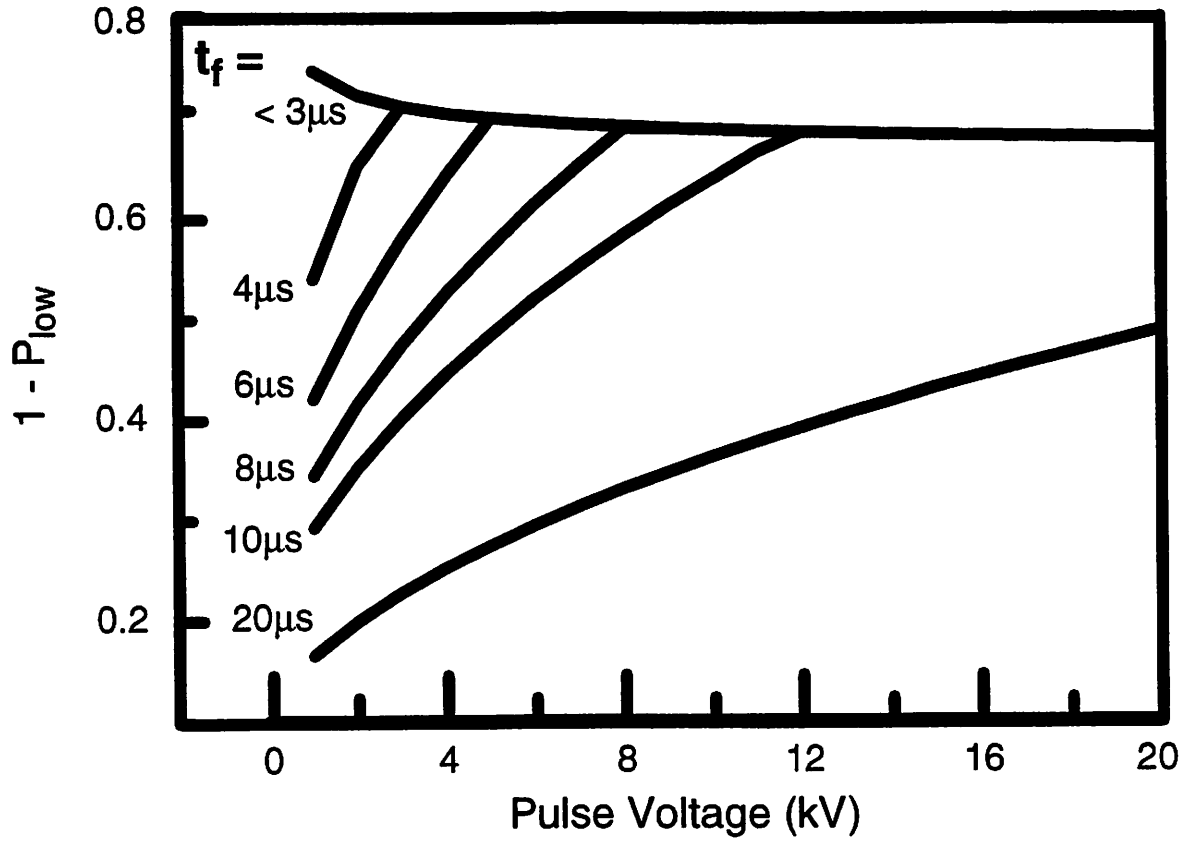


Figure 4-6 Low Energy Implantation for Entire Pulse Cycle

Cumulative 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 $10^{10}cm^{-3}$.

the ions from the matrix sheath implantation. For a 3μs on time, the maximum allowable fall time, as determined by Figure 4-5, is above 3μs for a 1kV pulse, and therefore all fall times less than this value show identical P_{low}'s, as shown in Figure 4-6. At 4μs, implant voltages less than 3kV are above their maximum allowable fall time, and therefore their P_{low}'s increase dramatically. At 10μs, all voltages below 12kV are above their maximum fall time.

It is quite important to have a fall time shorter than that determined by Figure 4-5, else the fall time corrupts the implant energy significantly. For longer fall times the

energy spread increases dramatically, and the fall time component dominates the total energy spread.

As seen in the above description, PIII pulsed implants have a significant low energy contribution, reaching 80% or higher for long fall times. Increasing the ion density does reduce the matrix sheath component, but may increase the fall time contribution because of the more stringent requirement on shorter fall times. Optimizing the implant with respect to ion density, requires knowledge of the fall time and the maximum allowable charge per pulse (as determined by charging considerations, see Chapter 5). With a short fall time, increasing the ion density to $10^{11}cm^{-3}$ from $10^{10}cm^{-3}$ reduces the low energy component by 20%.

4.4 Implant Energy Profile

The previous sections discussed the origins of low energy components to 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 profile estimation, is that the errors in the assumptions generally will be magnified, resulting in the profiles being mostly qualitative in nature.

There are three main methods for estimating the profile. First, and probably most accurate, are the Particle in Cell or similar type simulators. These simulators 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. The main problem with these simulators are that they are slow, and generally provide little insight into the mechanisms of the low energy components of the implant or the scaling of the low energy components with different plasma/implant conditions.

Another approach is to simply assume that the ion transit time across the sheath is zero, and therefore the implant energy is equal to the applied voltage when the ion reaches the sheath edge. This method provides fast profile predictions, but suffers from an underestimation of the low energy component. For fast rise times, the actual ion

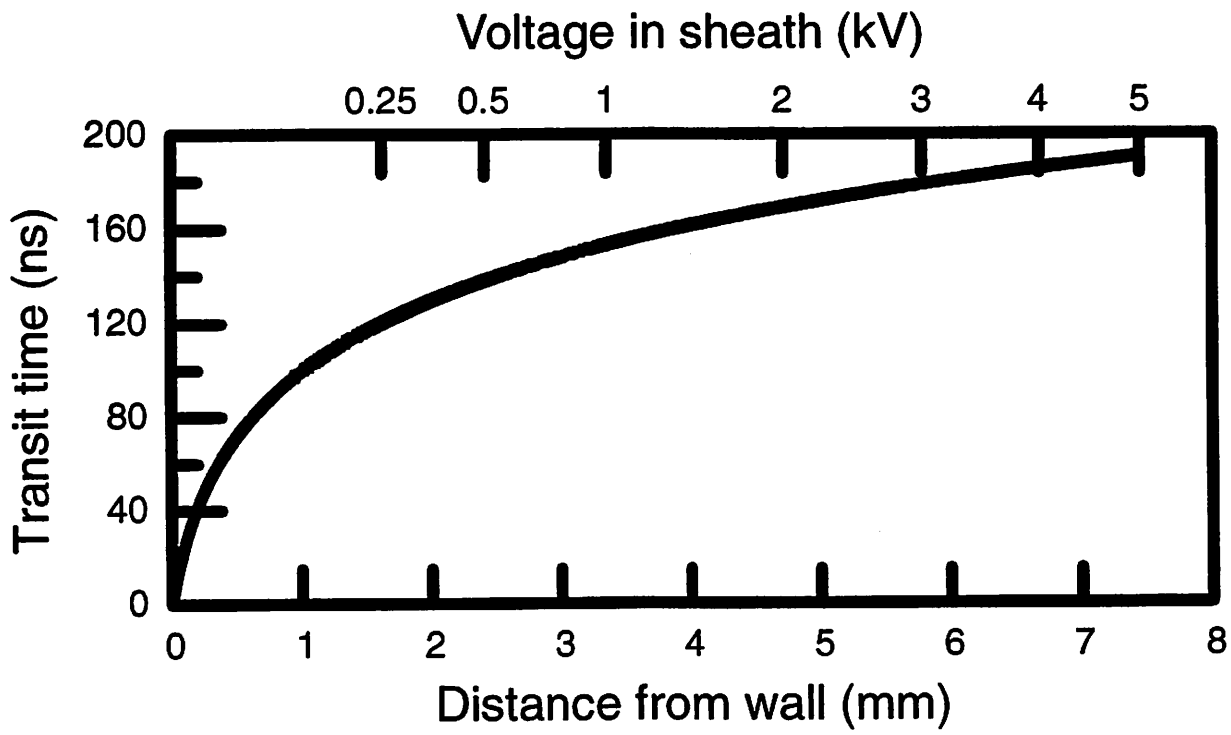


Figure 4-7 Ion Transit time for Matrix Sheath

The ion transit time for a 5kV pulse with a plasma ion density of $10^{10}/cm^3$. Note that the majority of the ions have a transit time above 100ns. These contradicts the assumption of a zero transit time. The calculation does assume a frozen electric field, and that the sheath does not expand during the transit of these matrix sheath ions.

transit time is longer than the rise time (Figure 4-7). This contradiction to the assumption will result 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 width is much wider, and the ion transit time can approach or exceed 500ns. 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 energy component for fall

times on the order of the ion transit times. This method will accurately model profiles for slow rising and slow falling voltage pulses, but this is in contrast to the goal of sharp pulses. A possible patch is to calculate the ion transit time, and then offset the current by this amount. This would increase the accuracy somewhat, especially with the fall time, but doesn't address the matrix sheath contributions.

A third method for profile prediction is to use a more analytical approach [4-8]. The obvious advantage is the insight afforded by analytical equations, and the simple extraction of scaling. Stewart et. al. have attempted an analytical solution to the problem of profile 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, that are common in current pulsing systems. By applying some of the concepts of the previous sections, a more accurate profile prediction is possible.

First for fast rise times, the matrix sheath contribution must be considered. The voltage profile of the matrix sheath is:

$$V(x) = \frac{qn_i x^2}{2\varepsilon_o} \tag{4-12}$$

where V(x) is the voltage distance x away from the wall edge. This also applies to longer rise times, except is doesn't account for ions that cross the sheath boundary during the rise time. For reasonable rise times, these ions may be ignored (but may be added if desired). For example, with a 5kV pulse, 100 ns rise time, and $10^{10}/cm^3$ ion density, the amount of ions crossing the sheath during the rise time ($3 \cdot 10^8/cm^2$) is less than 10% of the ions that are implanted from the matrix calculation above ($7.4 \cdot 10^9/cm^2$). If the rise times become excessively long, than the ions that cross the sheath should be accounted, and may be done so in a similar manner as [4-8], by assuming the transit time is near zero (which is a good assumption for long rise times).

Previous profile predictions ignored the effect of the ions in the sheath at the onset of the fall time. For fast fall times, this may lead to a significant undercounting of the low energy component of the implant. For a quasi-static Child Law sheath, the transit time, assuming a frozen E-field is:

$$t_c = 3s_{max}\sqrt{\frac{M}{2V_{pulse}}} \tag{4-13}$$

where t_c is the transit time across the quasi-static Child-Law Sheath. This equation underestimates the actual time, since it assumes a frozen field. If we assume zero field, i.e. a zero fall time for the pulse, the transit time is calculated:

$$v = \sqrt{\frac{2V(x)}{M}} \tag{4-14}$$

$$V(x) = V_{pulse}\left(\frac{s_{max} - x}{s_{max}}\right)^{4/3} \tag{4-15}$$

$$t_{c2} = \frac{x}{v} \tag{4-16}$$

where v is the ion velocity, V(x) is the voltage in the sheath at the onset of the fall time, and t_{c2} is the ion transit time assuming a zero fall time.

For fast fall times, those significantly less than the ion transit time, a zero fall time might be a better assumption. In this case the profile for the ions implanted during the fall time is calculated from Equation (4-8) and Equation (4-9).

By using these new assumptions the energy implant profile is estimated for a 3μs hold time, 5kV implant pulse in Figure 4-8. The y-axis shows the probability distribution function cut up into 100V energy bins. The fall time of the pulse is 1μs, which is faster than the ion sheath collapse time. The dashed line is the estimation for the current-voltage comparison method using the zero transit time assumption with a 100ns rise time. The ion current was determined using the plasma model, which is known to accurately predict the ion currents. The solid line is the prediction using the new method, which

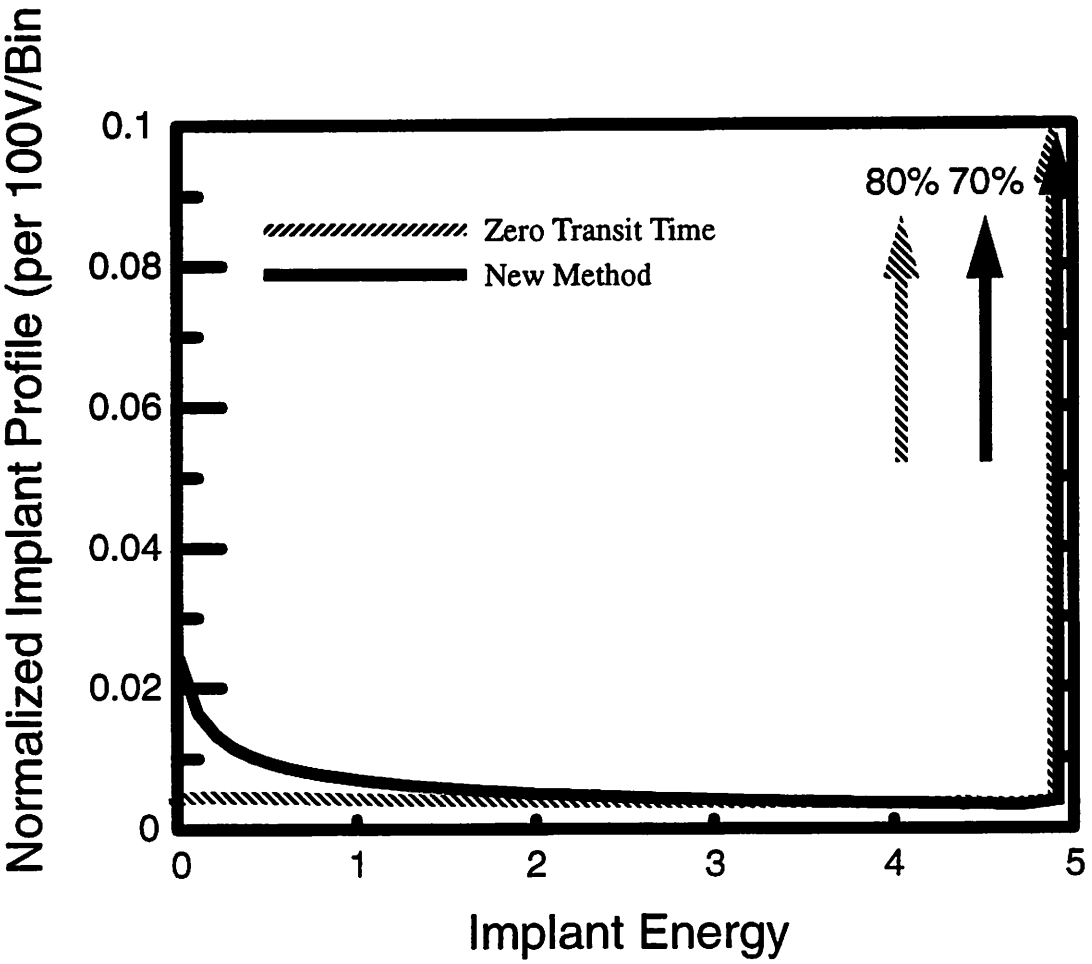


Figure 4-8 Estimated Implant Energy Distribution.

Estimated implant energy distributions for a 3μs/5kV pulse. 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 dashed line assumes a zero ion transit time with a 100ns rise time. The zero-transit time predicts 80% of the ions implanting 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 1kV, which mostly results from the matrix implantation.

explicitly accounts for the matrix implantation. The zero transit time method with a 100 ns rise time predicts that 80% of the ions implant with the peak energy (V_{pulse}), while the new model estimates only 70%. If the rise time is reduced to 50ns, the difference increases to 15%. By examining the profile, 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 clear that the zero-transit time method is inaccurate for rise times less than 100ns; It probably becomes sufficient for rise times nearer to 500 ns.

Overall, for fast rise times (our pulser rise time is about 50ns), it is imperative to consider the matrix contribution to the low energy implant. Neglecting to do so will cause an overestimation of the high energy component of the implant by around 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). The new model proposed for estimating the profiles still is less accurate than the full PIC simulators, since it still assumes a frozen E-field, even though the sheath is expanding. This effect would result in ions implanting at energies near the peak, but not at the peak. The energy loss would directly depend on the rate of sheath expansion.

4.5 Dielectric Implantation

We have investigated PIII for the two main dielectric substrate applications: thin film transistor and silicon-on-insulator technologies. Thin film transistors (tft) are a key technology in liquid crystal displays. There at least four different implantation steps in a tft process flow: source/drain, poly gate, hydrogenation, and poly grain size control. All of the implants dictate high doses with implant energies ranging from 20-100 keV. Silicon-On-Insulator technology, with buried oxide layers approximately 50-500 nm thick, promises faster devices and a simpler CMOS process flow than conventional bulk wafers.

Implantation with dielectric substrates introduces new energy spread mechanisms. The first complication arises with the coupling of the voltage pulse to the plasma; by capacitive division some of the voltage couples to the substrate, reducing the effective

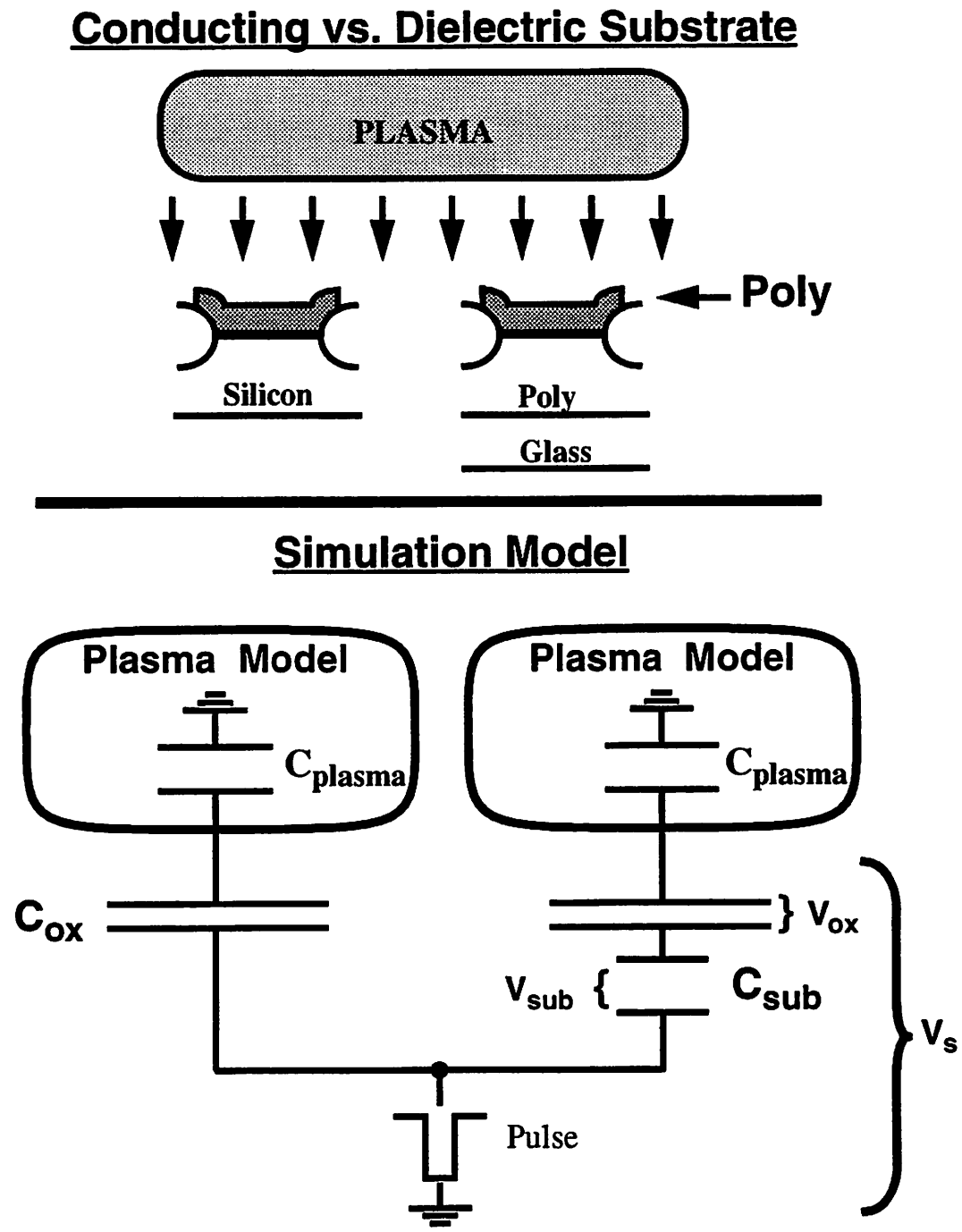


Figure 4-9 PIII Model for Dielectric Substrates

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 capacitance 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 the simulator, dielectric substrates are modeled as leakless capacitors.

implant voltage. Then, during the implant, deposited positive ions establish a voltage drop across the dielectric substrate. A typical 10keV implant deposits ~$3 \cdot 10^{-4}$ Coulombs/cm^2, and builds-up ~4keV over a 0.5 mm glass substrate, reducing the implant voltage by the same amount. Using the PIII model for dielectric substrates, the three stages of PIII, rise, hold, and fall time are re-evaluated in terms of energy spread for dielectric implantation. The PIII model for dielectric substrate implantation is illustrated in Figure 4-9.

4.5.1 Sheath Voltage During Implantation

There are three stages of PIII, the rise time, the hold time, and the fall/off time, each with distinct implantation characteristics. Before pulsing but after plasma exposure, the surface charges to the plasma floating potential. The sheath width, a function of ion density, is small, usually less than 1mm. During the rise time, the applied voltage pulse (V_{pulse}) capacitively couples to the sheath and glass substrate, with thicker dielectric substrates reducing the coupling efficiency to the plasma. Initially, the sheath expands rapidly producing a large plasma ion current, which implants with less than the full pulse potential. The implanting ions eject secondary electrons, amplifying the deposition of positive charge. For high implant voltages, the secondary electrons dominate the surface charge deposition with a yield (γ) in the 1-20 range [4-4]. Previous papers [4-6] ignored the secondary electrons, vastly underestimating the surface charge. The extremely large plasma ion current builds-up a significant surface charge and substrate voltage drop, which reduces the sheath voltage. For the small substrate thicknesses of interest (0-2mm), the reduction in sheath voltage is mostly due to charge deposition rather than capacitive coupling losses. However, capacitive coupling becomes significant for thicker substrates. For the 0.5 mm substrate case shown in Figure 4-10, the maximum sheath voltage is 15kV. This translates to a coupling efficiency ($\eta = \frac{V_{max}}{V_{pulse}}$) of only 75%, while only 8% of the total loss is attributable to capacitive coupling losses.

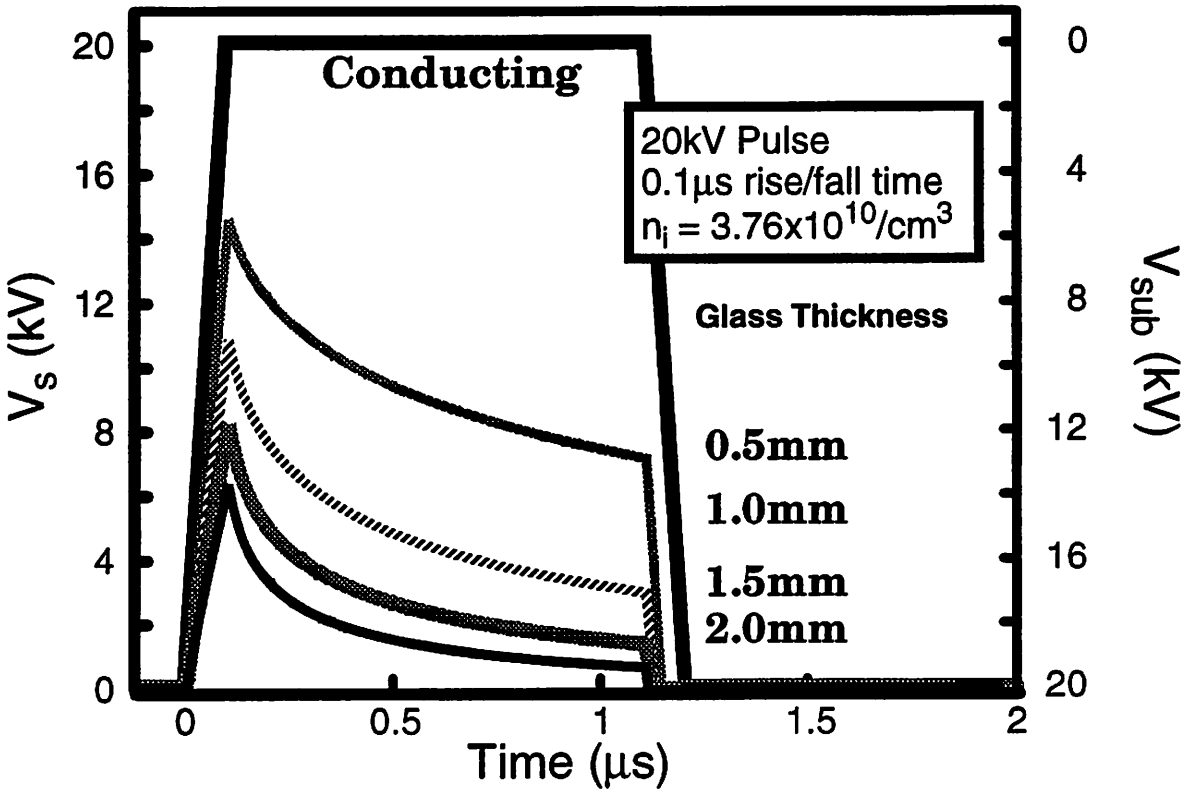


Figure 4-10 Sheath Evolution with Glass Substrates

V_{sheath} during implantation with varying substrate thicknesses for a 20kV, 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.

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 is so severe that the sheath voltage can be extinguished after only a couple of microseconds. For this discussion, the self-extinguishing time is defined as the point at which the voltage build-up across the substrate attenuates the surface voltage by 90%. For example, if V_{pulse} equals 10kV, the self-extinguishing point occurs when the voltage across the substrate is 9kV, reducing the instantaneous implant energy to 1keV.

The falling edge of the pulse capacitively couples to the sheath, actually causing the simulated sheath potential to become negative, an extremely non-equilibrium situa-

tion. To resolve this, the simulator limits the electron current to the electron saturation value (Equation (3-7)). 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).

As shown in Figure 4-10, the sheath voltage evolution varies with the thickness of the glass. Thicker substrates worsen the capacitively coupling of V_{pulse} 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}). The value for δ is directly dependent on the total deposited positive charge. Therefore,

$$\delta \propto \text{Dose/Pulse} \cdot (1 + \gamma(V)) \tag{4-17}$$

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

4.5.2 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 considerable spread of energies resulting from implantation during the rise and fall times [4-8]. The significant surface voltage buildup with a dielectric substrate further disperses the energy distribution. Figure 4-11 shows the implant energy distributions for each of the three stages of PIII. To reduce simulation noise, the implant dose is integrated over 400V 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. During the hold time, charge deposition reduces the implant energy. Because $\Delta V_{sheath}/\Delta t$ slows (Figure 4-11), the Dose/Bin increases with time (decreasing energy). During the fall time, J_i is small, yielding only a blip in Figure 4-11 on

the falling edge of the pulse. 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, a doubling of the substrate thickness will nearly double the energy spread during the hold time.

4.5.3 Applied Voltage Effect

Optimizing the implant energy for the tft process flow 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, since increasing the Dose/Pulse obviously increases the charge build-up on the substrate. The Dose/Pulse for PIII is approximately:

$$\text{Dose/Pulse} \approx qn_i u_b t_w + qn_i s_{max} \quad (4\text{-}18)$$

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, with $s_{max} \propto V_{pulse}^{3/4}$ (Equation (3-4)), the Dose/Pulse from Equation (4-18) increases sub-linearly with V_{pulse}. To determine the scaling of δ with applied voltage, the relationship of $s_{max} \propto V_{pulse}^{3/4}$, and the scaling of secondary electron yield with voltage ($\gamma \propto V^{1/2}$), must be combined in Equation (4-17). This results in the total energy spread, δ, being proportional to $V_{pulse}^{5/4}$. The actually scaling should be slightly less than $V_{pulse}^{5/4}$, since the implant voltage degrades during the pulse, reducing the secondary electron yield from its peak value The secondary electron yield, averaged over the entire pulse width, actually scales less than the assumed square root dependence with respect to V_{pulse}, especially for implants with a wide range of implanted energies.

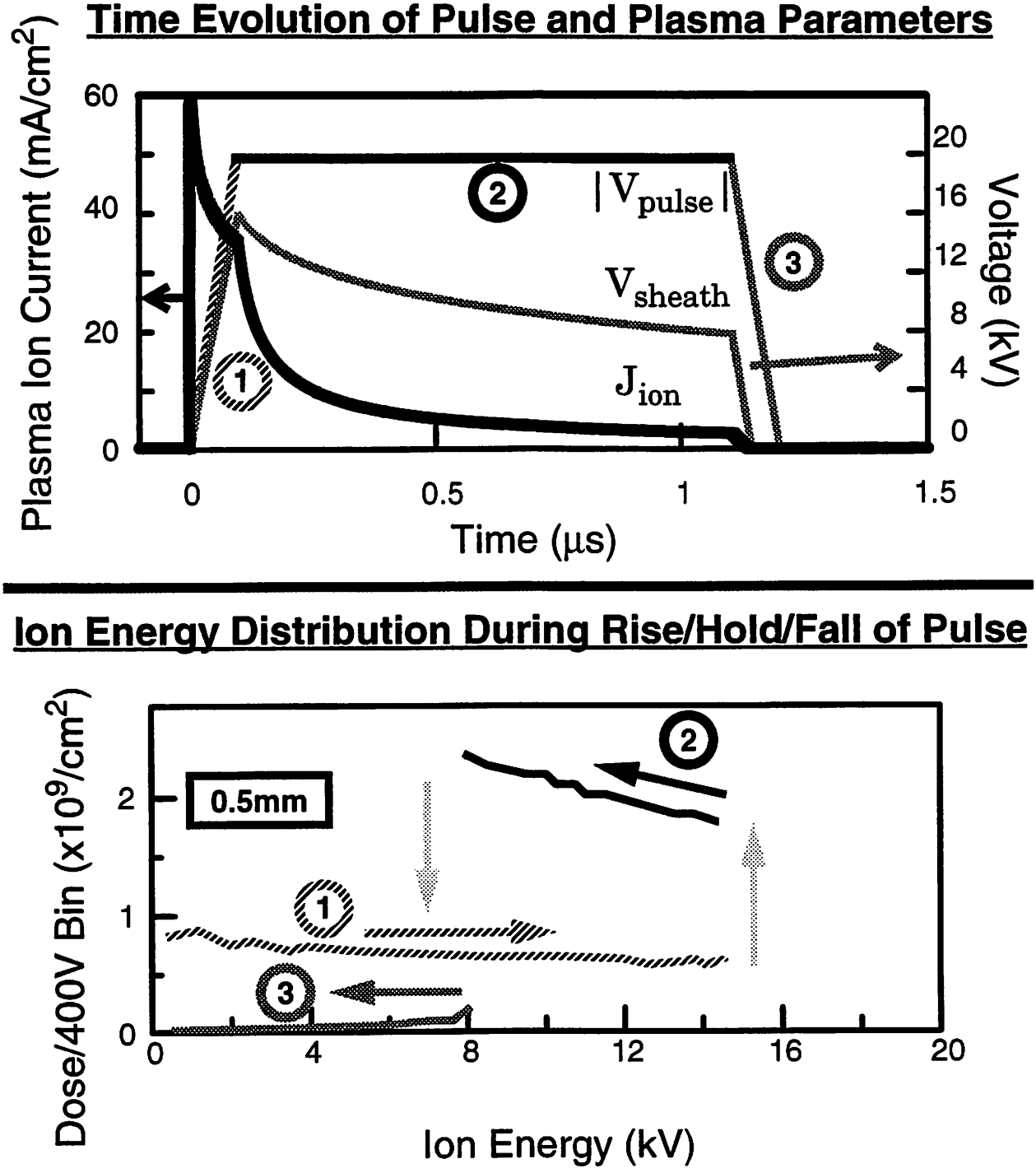


Figure 4-11 The Three stages of PIII

The implant ion energy distribution with the corresponding pulse, sheath voltage and plasma ion current with a 0.5mm glass substrate and a 20kV pulse. During the rise time, step ①, the energy distribution ramps up to the maximum implant energy. During the hold time, step ②, positive charge deposition reduces V_{sheath}, thereby lowering the implant energy. The ion current is low during the short fall time, step ③, not contributing much to the overall implant. The fall time stage could become more important for long decay rate pulses.

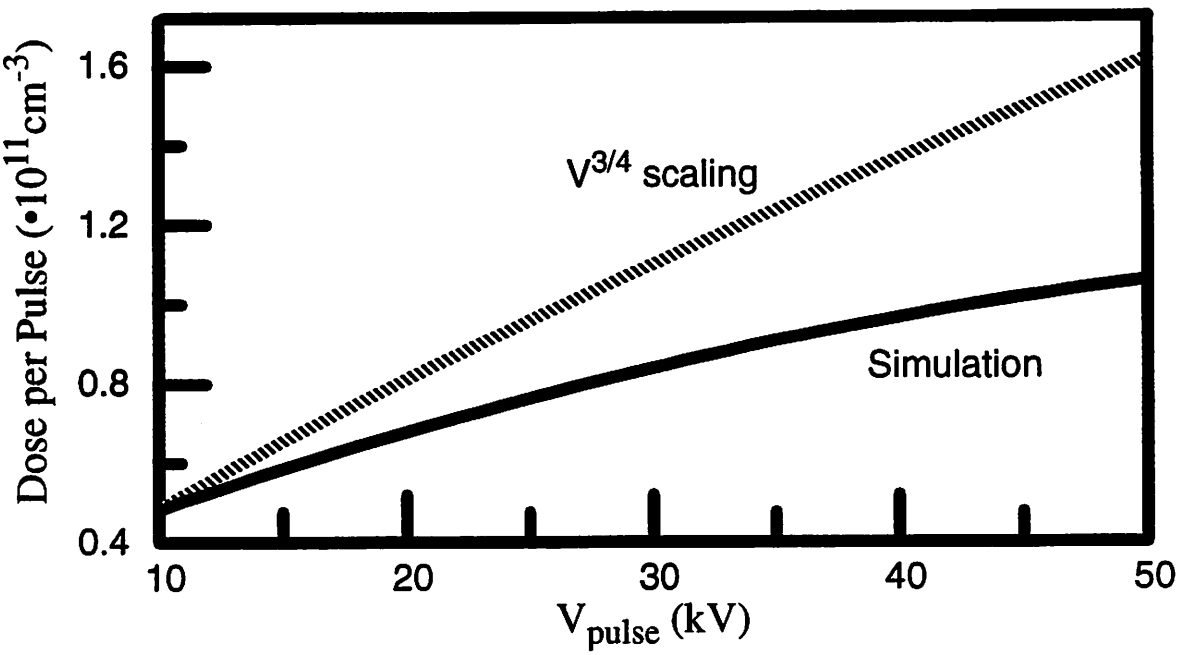


Figure 4-12 Effect of Bias Voltage on Dose/Pulse for Dielectric Substrates

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

The simulation results for energy spread and Dose/Pulse are shown in Figure 4-12 and 4-13. The Dose/Pulse indeed increases sub-linearly with applied bias; even less than the simple $V_{pulse}^{3/4}$ model. This is attributable to the sheath expansion rate not being proportional to the final Child Law sheath width. As expected, the energy spread does increase with bias voltage, but increases slightly less than the predicted $V^{5/4}$ scaling. As explained, this is attributable to the reduction of the secondary electron yield during the implant from the reduction in implant energy. Overall, increasing V_{pulse} boosts the Dose/Pulse while widening the voltage spread.

4.5.4 Ion Density Effect

The plasma ion density is a controllable parameter for most plasma sources, and therefore, it is interesting to explore the effect of changing n_i on the implant characteristics. Increasing n_i raises the Dose/Pulse, decreasing processing time, but the higher dose widens the energy spread. Equation (4-18) suggests that the dose rises linearly

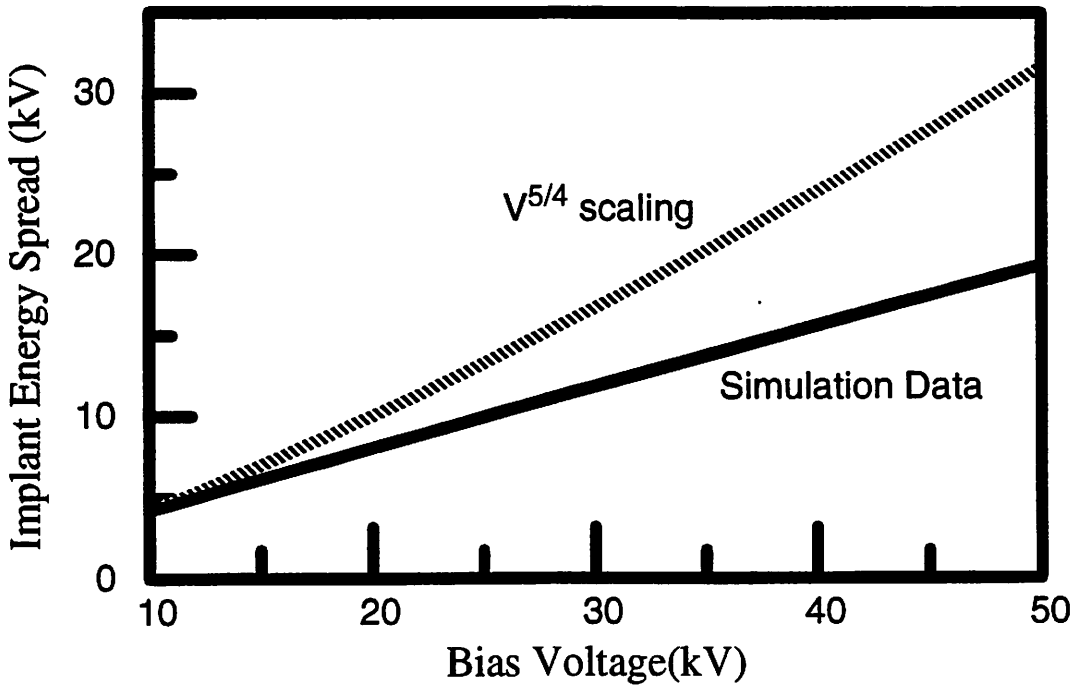


Figure 4-13 Scaling of Implant Energy Spread with Bias Voltage

The predicted scaling of implant energy spread (δ) with bias voltage is a $V_{pulse}^{5/4}$ relationship. The actual scaling from simulations is slightly less. The most obvious reason for the difference is that scaling of s_{max} with voltage does not follow the steady-state Child Law scaling, especially for shorter implant pulses.

with n_i, but the full plasma model shows a sub-linear dependence on n_i (Figure 4-14). The difference between the two results from s_{max} decreasing with n_i, (the steady-state Child Law sheath scales $s_{max} \propto n_i^{-1/2}$). In addition, the larger J_i reduces V_{sheath} during the implant, retarding the sheath expansion. Therefore, the thinner sheath and the slower sheath expansion rate combine to reduce s_{max} and the scaling of Dose/Pulse with n_i. Along these same lines, the increased current and the thinner sheath combine to reduce η (Figure 4-15), lower the mean implant energy, and widen δ.

4.5.5 Pulsing Frequency

Two fully controllable variables in PIII are the pulsing frequency (f_p) and pulse width (t_w). Both of these need to be optimized for maximum throughput and minimum gate oxide damage. The theoretical maximum dose rate for PIII occurs with a DC bias:

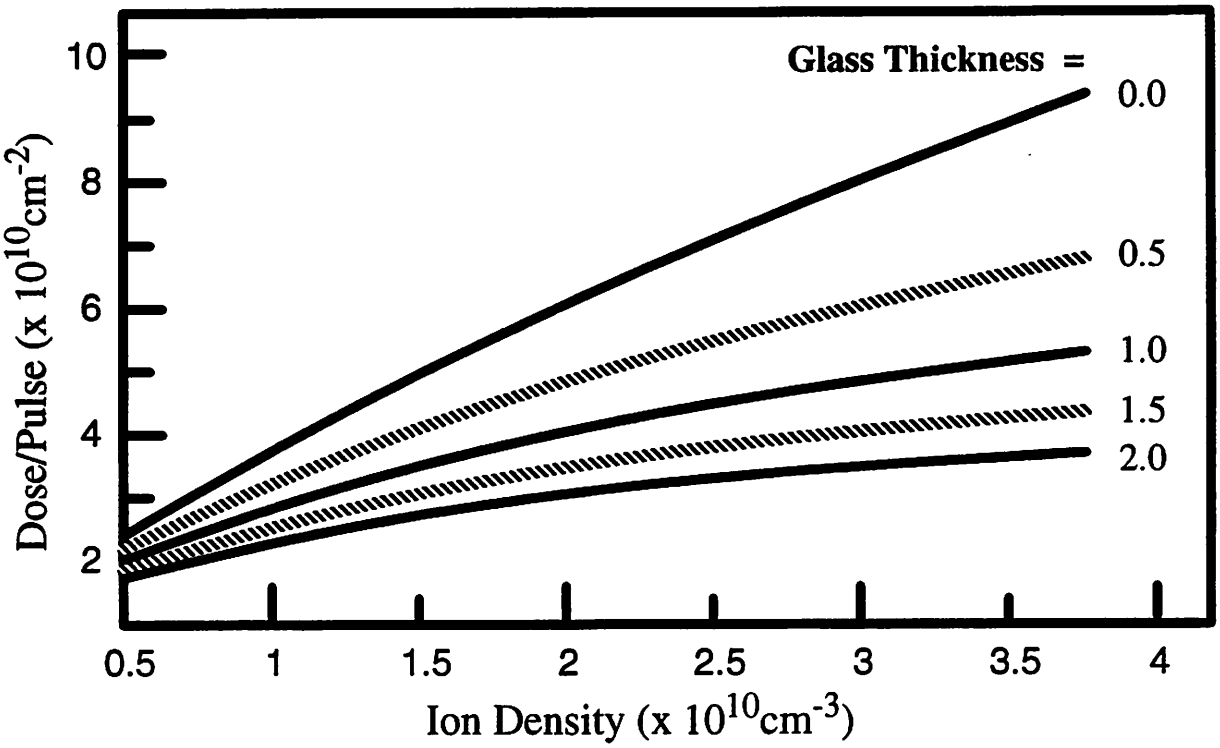


Figure 4-14 Scaling of Dose/Pulse with Ion Density

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. Data is for a 20kV/1μs pulse.

$$\text{Maximum Dose Rate} = q \cdot n_i \cdot u_b \qquad (4\text{-}19)$$

With pulse bias operation, the sheath expands during the pulse, increasing J_i above this value. When the pulse is turned off, the sheath collapses, and J_i temporarily goes below this value, so that the time-averaged current is always $q \cdot n_i \cdot u_b$. To maximize throughput, one wants as much of the total ion current implanted, rather than hitting the surface at low voltages. For dielectric substrate implantation, t_w is limited by the self-extinguishing time. If the pulse is on long enough, the charge deposited by the plasma ions and ejected secondary electrons will completely counterbalance V_{pulse}, yielding a sheath voltage close to zero. Any time that the pulse is held on after self-extinguishment is basically wasted, since the ion impinging energy is so low. For the implant conditions of interest, this occurs in the 0.5μs to 10μs range. The counterpart to t_w is f_p, or the off time

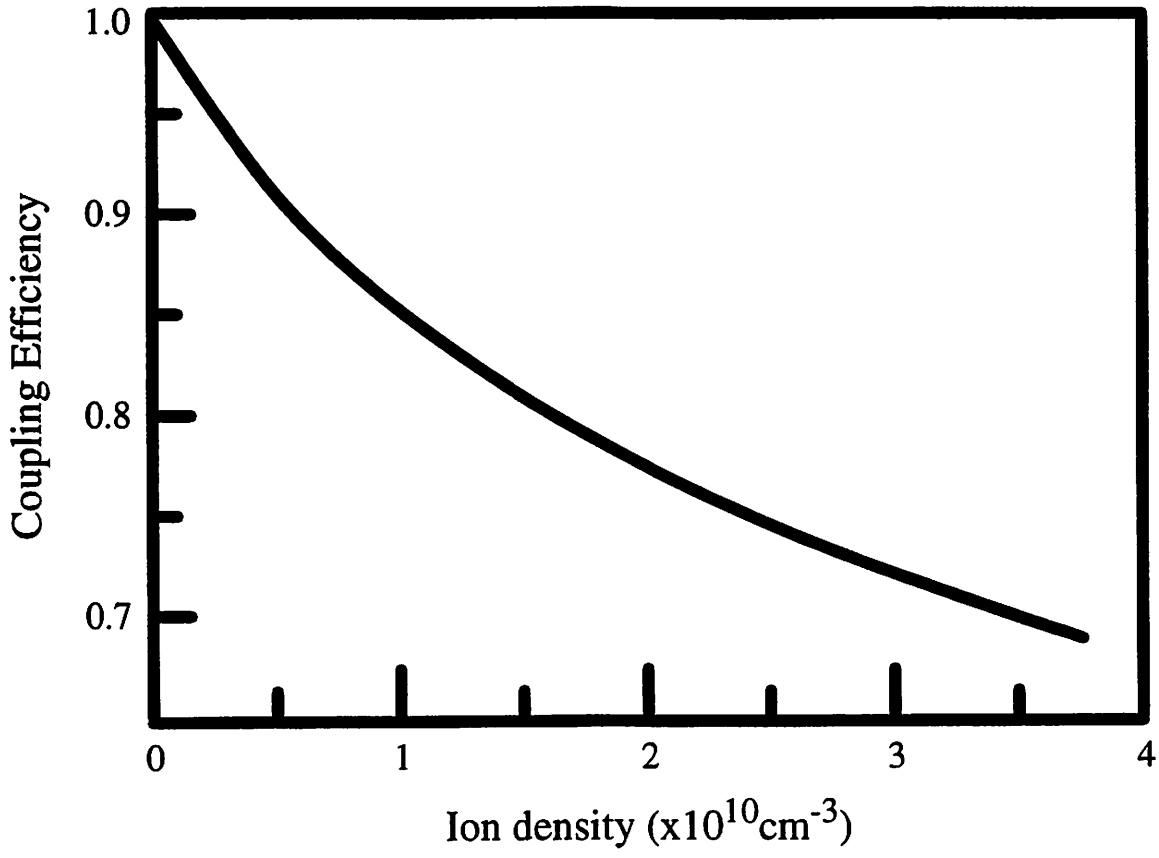


Figure 4-15 Scaling of Coupling Efficiency with Ion Density

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 20kV / 1μs pulse with 0.1μs rise times.

in between pulses. The off period should be longer than the surface charge neutralization time. As stated previously, this occurs in a fraction of a microsecond after the sheath fully collapses for dielectric substrates Therefore, the only necessary limitation on 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. For instance, with a 20kV pulse, 0.5 mm thick glass substrate, and an n_i of $3.76 \bullet 10^{10} cm^{-3}$, the pulse extinguishes itself after 6.12 μs. The sheath fully collapses 3.15μs after the onset of the fall time, or the after the fall time is over, whichever is longer. With a 2μs fall time and a 1.15ms off, time, the pulse frequency will be 109kHz. This gives a 100% efficiency (defined as $\frac{\text{Implant Dose Rate}}{\text{DC Dose Rate}}$). If the maximum pulse frequency was only

25kHz (as is the case with our pulser), the off time would be 31.88μs, yielding an implant efficiency of 23%. By properly optimizing t_w and f_p, it is possible to have an implantation current close the DC value dictated by Equation (4-19).

4.6 Conclusions

The poly-energetic nature of the PIII implant requires special attention. Implantation from the formation of the matrix sheath, and implantation from the ions in the sheath on 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 non-inherent sources of energy spread.

Implanting into dielectric substrates introduces bias coupling to the substrate and charge build-up as additional sources of energy variation. 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.

Accounting for all the sources of energy spread, understanding the limitations and the scaling trends with the implant variables scaling, allows the identification of an implant condition that yields an allowable 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, pp. 531, 1994.

[4-2] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *J. Vac. Sci and Tech. 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," *J. 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", *J. Vac. Sci. and Tech. 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," *J. Applied Physics*, vol. 70, no. 7, pp. 3481- 3487, 1991.

5 Gate Oxide Charging Damage

5.1 Introduction

Oxide charging damage is a major concern for integrated circuit plasma processing. As gate dielectric thicknesses continue to decrease, the voltage necessary to induce damage will also decrease, enhancing the charging concerns. Previous studies have shown that geometrical effects, such as antennas, affect the amount of charging damage [5-1]. It will be shown that the substrate and well type also have an effect on oxide charging.

In this chapter, gate oxide charging is discussed in terms of PIII, although the conclusions can apply to all plasma processes. In PIII, 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 this deposited positive charge. Depending on the implant conditions, charge may accumulate on the wafer, posing an oxide charging problem. En [5-2] has successfully combined the equations governing the plasma with a gate oxide model, to predict gate oxide damage. By extending this model, it is shown that wells and substrate type affect gate oxide charging.

5.2 Gate Oxide Damage Measurement

A number of methods have been developed to quantify gate oxide stress. These separate into two distinct groups, destructive and non-destructive. The more popular destructive methods are Charge-to-Breakdown (Q_{bd}), and Time-Dependent Dielectric

Breakdown (TDDB). By measuring the amount of stress that results in breakdown, and subtracting this from the virgin breakdown value, the amount of stress is calculated (i.e. Q_{bd} (before stress) - Q_{bd} (after stress) = Q (during stress)). Large statistical variations plague these methods, and their destructive nature preclude further evaluation of the gate oxide.

The major indirect measurements are interface trap extraction and threshold shifts. During oxide stress, interface traps accumulate, altering the capacitance of the MOS system. Measuring the change in capacitance after stress, as a function of voltage (C-V measurement), reveals the stressing damage. In addition to altering the capacitance, the presence of traps shifts the threshold voltage (V_t) of the MOS capacitor/transistor. The magnitude of V_t shift correlates to the amount of damage. Measuring the capacitance changes from the interface traps requires large area test structures (i.e. a large capacitance) to overcome the background capacitance noise of the measurement system. This usually requires MOS capacitors of at least 50μm x 50μm. In contrast, threshold shifts are measurable for transistors of any size. Therefore, the preferred indirect method depends on the size of the test structures.

5.2.1 MOS Capacitance

The MOS capacitance system model, including the effect of interface traps, is depicted as in Figure 5-1. When applying a gate voltage, the total capacitance of the system is determined by where the electric fields lines terminate, or where the charge forms in the substrate. 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, one only has to keep track of where the new charge is stored to determine the capacitance of the system. If charge forms in either the inversion layer, accumulation layer, or at an interface trap the 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

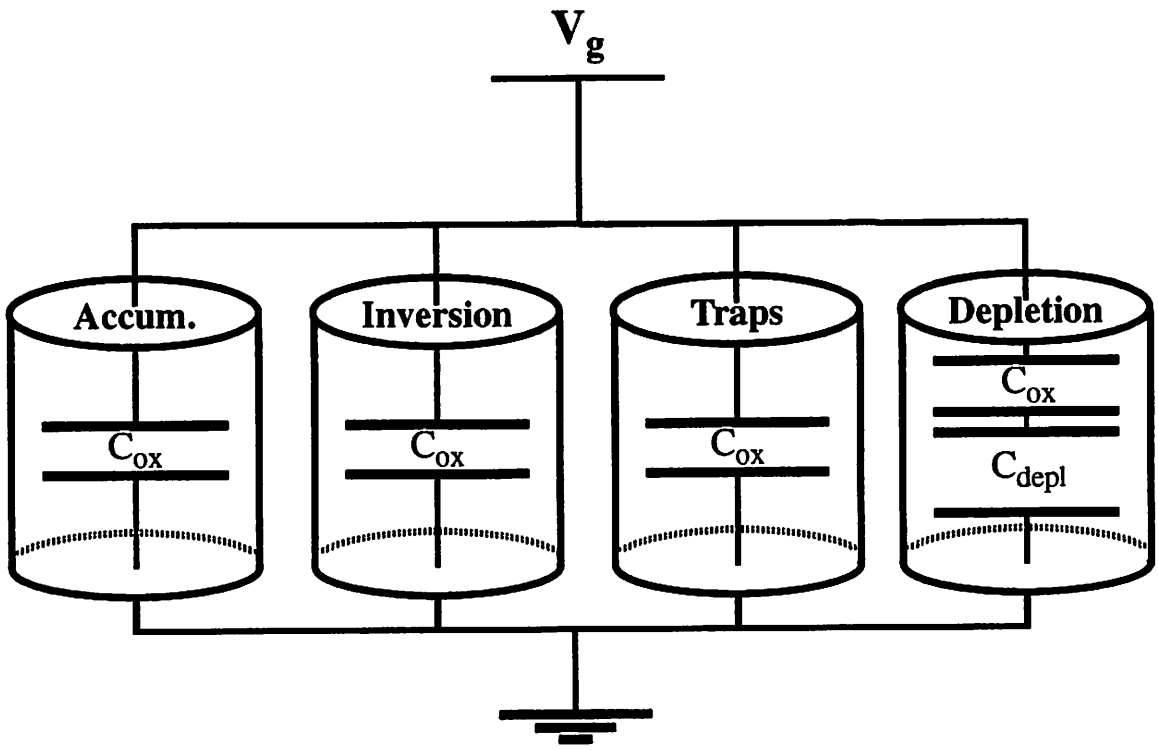


Figure 5-1 The MOS C-V System Model

The total capacitance of a MOS system is composed of four different parallel contributors. The accumulation, inversion, and interface trap capacitances are all equal to C_{ox}, while the depletion contribution is equal to the series connection of the oxide and depletion capacitances. The small signal capacitance is computed by keeping track of which barrel the new charge is stored.

depletion capacitance (since new depletion charge is always stored at the bottom of the depletion region). The total capacitance consists of the sum of the contribution of the four parallel capacitors. For instance, if half the charge goes into interface traps and the other half to the depletion region the total capacitance is $0.5\ C_{it} + 0.5\ C_{depl}$.

5.2.2 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-sensitivity 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, and only modulates depletion charge. Therefore the depletion region determines the capacitance, not the inversion region. Along the same lines, interface traps have significant time constants, and do not respond to fast (Mhz) signals. Therefore the quasi-static (low frequency) measurement measures all capacitances, while only accumulation and depletion charge respond to the high frequency measurement.

5.2.2.1 Quasi-static measurement

The Quasi-static (QS) measurement ramps the gate voltage to determine the C-V relationship of the test structure. As long as there is negligible leakage, the current drawn from the measuring device is proportional to the voltage ramp rate.

$$I = C \cdot \frac{dV}{dt} \tag{5-1}$$

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 and allowing the interface traps and inversion charges to respond. A sample QS measurement is shown in Figure 5-2.

5.2.2.1 High Frequency Measurement

During a high frequency (HF) measurement, a small amplitude (~ 0.026 V) is applied at a spot bias. The high frequency measurement attempts to measure the capacitance of the depletion and accumulation region, while not allowing the inversion and interface trap regions time to respond. This usually requires signal rates exceeding 1 MHz. Stepping the voltage, allowing ample time for carrier equilibrium, produces a full C-V curve. A sample HF measurement is shown in Figure 5-2.

5.2.3 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

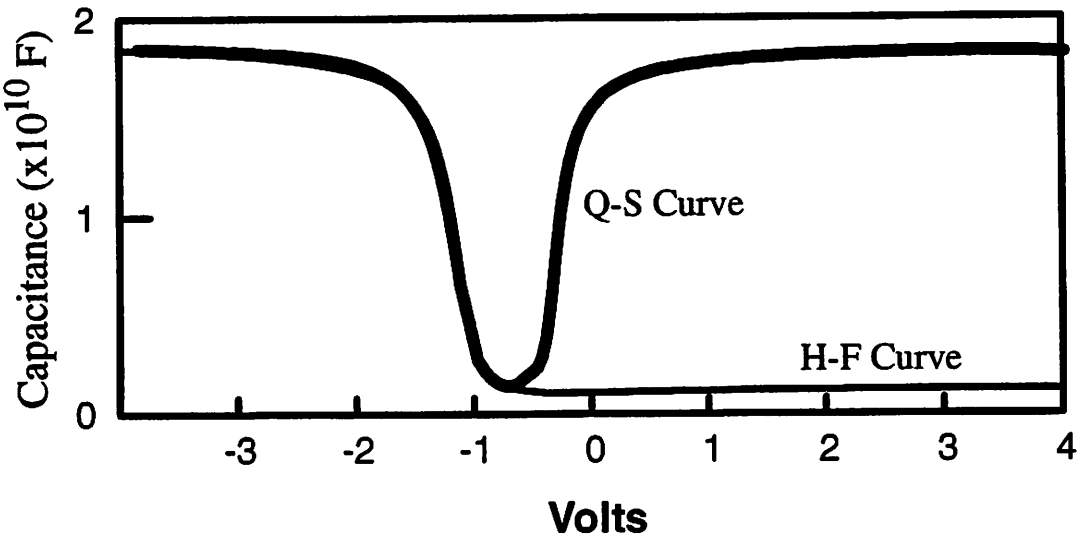


Figure 5-2 Sample "undamaged" Quasi-Static and High Frequency C-V curves.

Representative Q-S and H-F 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 to the high frequency.

corresponding measured curve. The difference between the capacitances is assumed to be due to the presence of interface traps. This method assumes a priori knowledge of the depletion capacitance, which is a function of the doping density underneath the gate oxide. Any errors in the presupposed 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.

5.2.3.1 Quasi-Static and High Frequency Comparison

The most common extraction method compares the quasi-static and high frequency measured C-V curves, eliminating many of the errors associated with the use of theoretical C-V curves. The QS measurement allows all the interface traps to respond to the signal, and therefore includes them in the capacitance. The HF measurement operates at frequencies above the interface trap rate, and therefore does not include them in the capacitance measurement. By manipulating the effective capacitance equations for the two measurements, the interface trap capacitance as a function of voltage is solved:

$$C_{it} = \{ \left(\frac{1}{C_{qs}(V_g)} - \frac{1}{C_{ox}}\right)^{-1} - \left(\frac{1}{C_{hf}(V_g)} - \frac{1}{C_{ox}}\right)^{-1} \} \quad (5\text{-}2)$$

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 [5-3]. With the interface capacitance calculated, the interface trap density as a function of V_g per eV is simply:

$$D_{it}(V_g) = \frac{C_{it}}{q \cdot A} \quad (5\text{-}3)$$

where A is the capacitor area. Typically it is more useful to integrate over a region in the bandgap for the total amount of interface traps, but first V_g must be transformed into Φ_s, the potential at the oxide-silicon interface. This is accomplished by Berglunds Method [5-4].

$$\Phi_s = \int_{V_{fb}}^{V_g} \left(1 - \frac{C_{qs}}{C_{ox}}\right) dV \quad (5\text{-}4)$$

where V_{fb} is the flatband voltage. Finally, the transformed D_{it} is integrated across the bandgap as shown in Figure 5-3. The error enlarges dramatically near the band edges, usually confining the integration to the midgap region. The silicon bandgap is ~1.12eV, so it is customary to integrate symmetrically around 0.56eV, 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.

5.2.3.1 Quasi-Static Only Comparisons

Small stray capacitances severely affect the HF C-V measurement, and locating a high enough 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 that relies solely on the easier QS method is desired. One such method which compares before and after stress QS curves is quite reliable. Any new interface traps will increase the

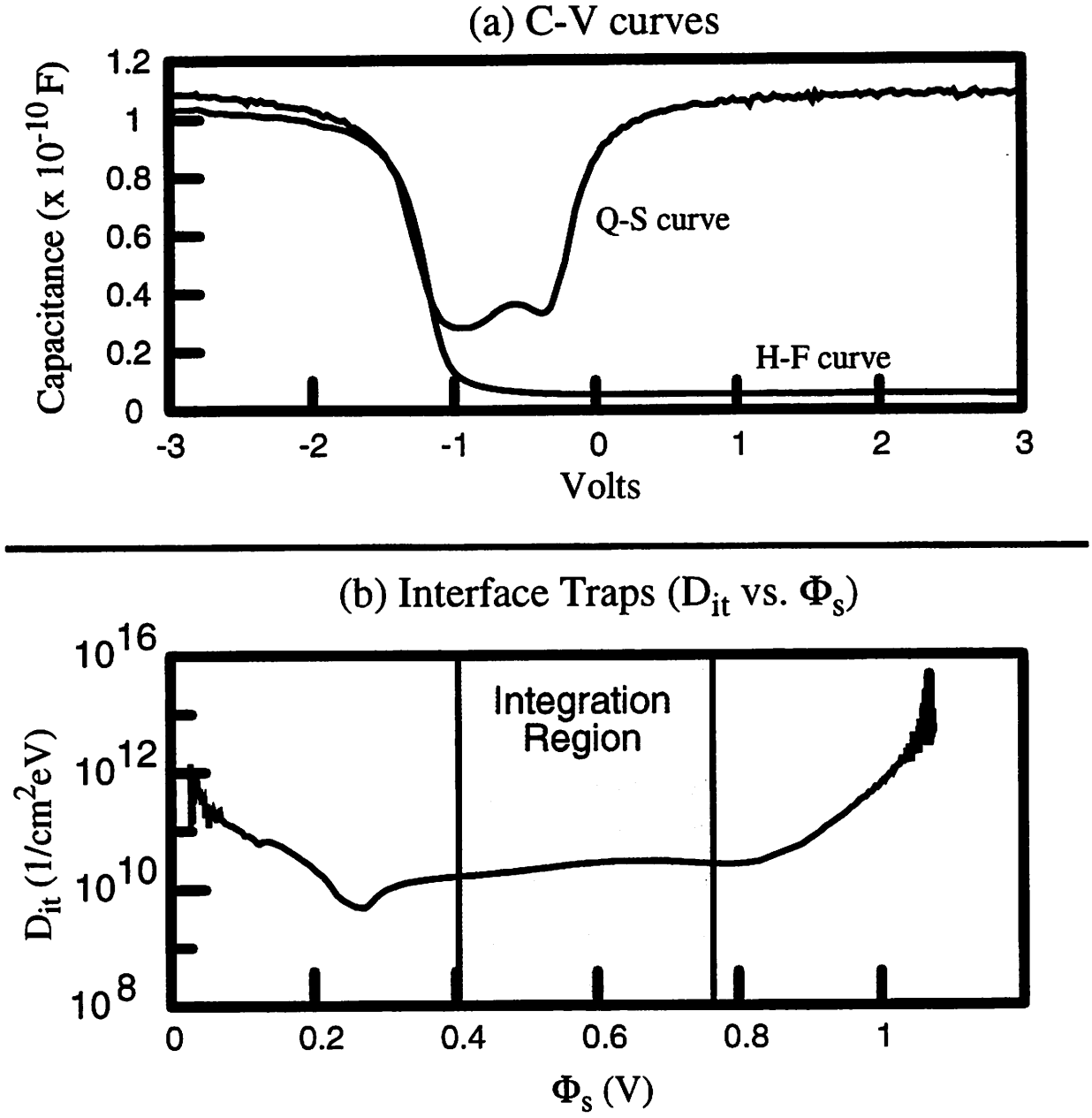


Figure 5-3 C-V Extraction of Interface Traps

(a) "Damaged" Quasi-Static and High Frequency 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 multiplying by an area factor yields a interface trap density of $1.94 \cdot 10^{11}/cm^2$ for this sample.

capacitance at a given Φ_s for the QS measurement, and the difference can be integrated, yielding the change in interface trap density. In 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}:

$$D_{it}(\Phi_s) = \frac{\left\{\left(\frac{C_{qs2} \bullet C_{ox}}{C_{ox} - C_{qs2}}\right) - \left(\frac{C_{qs1} \bullet C_{ox}}{C_{ox} - C_{qs1}}\right)\right\}}{q \bullet A} \qquad (5\text{-}5)$$

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

The requirement of before and after stress measurements, where the dual QS/HF technique requires only after stress measurements, presents the only drawback of this technique. Because of the increased ease of interface trap extraction, the Quasi-Static Only method is employed throughout this paper.

5.2.4 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 measurement set-up and those inherent in the technique.

5.2.4.1 Measurement Conditions

The Quasi-Static capacitance measurements were made with the HP 4140B picoammeter. 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 (5-1)). 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 chapter 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 the leakage currents. Leakage is monitored before each voltage sweep by examining the current at a DC 4V applied bias. Ideally the current should be zero, and for low leakage situations is less than 10fA. During the sweep, 250 data points are taken, with intermediate values calculated by simple linear interpolation. Under proper conditions, the accuracy of the Quasi-Static measurement is near 1%.

The high frequency measurements were made with the HP 4192 Impedance Meter. The main user parameter is the frequency of the measurement. 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 is set at 1Mhz, which is a compromise between the inversion layer generation rate, and the limit dictated by stray capacitances (that dominate above 10MHz). To properly exploit the internal compensation for the coaxial line reflectances, the wires must be exactly 1 meter long. Under proper measurement conditions, typical measurement errors are near 1%, or 1pF, whichever is larger.

5.2.4.2 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 [5-5] discuss these errors in detail, and they will be summarized here. First, there is error with the assumption that a 1MHz HF measurement is a true

high frequency measurement. Some interface traps will respond at 1 MHz, especially near flatband, where the trap capture time is the most rapid [5-5]. With a $10^{15}/cm^3$ doped substrate, errors in excess of 10% in the interface trap extraction occur from approximately flatband to 0.1 V away from flatband (in terms of gate voltage). 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, a false value of $10^{10}cm^{-2}/eV$ will be added 0.8eV into the bandgap. Higher substrate dopings decrease the error, since inversion onset occurs later, and a $10^{10}cm^{-2}/eV$ error arises 0.9eV into the bandgap, for a $10^{18}/cm^3$ doped substrate.

Another source of error, occurs with the calculation of the reciprocal of the difference of two nearly equal numbers in Equation (5-2) and Equation (5-5), which is called round-off error. 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 be compensated, though, 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 interface 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 $10^{10}cm^{-2}/eV$, and since the integration range is 0.32V, the minimal detectable density is $3.2 \cdot 10^{10}/cm^2$. For higher values of interface traps, the accuracy is expected to be better than 10%. Since the preferred method for interface trap extraction compares before and after stressing quasi-static measurements, the minimum sensitivity isn't so much an issue as the accuracy. This arises since the before capacitance usually has trap densities near (or even above) the minimum sensitivity already.

5.3 General PIII Oxide Charging Case

Before discussing the effect of sub-surface structures on gate oxide charging, the origin of oxide charging in PIII needs to be derived.

Because the 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. When the substrate is grounded, this is also the voltage across the gate oxide (V_{ox}). For a floating surface, the net current from the plasma must be equal to zero ($J_i = J_e$). To satisfy this condition, the surface voltage of the wafer is

$$V_s = V_p - T_e \ln\left(\frac{M}{2\pi m}\right)^{1/2} \tag{5-6}$$

where T_e, m, and V_p are the electron temperature, the electron mass, and the plasma potential respectively.

When the negative pulse is applied, the incoming plasma ions and ejected secondary electrons make the surface voltage more positive. While the pulse is on, the plasma electrons can not surmount the sheath potential ($V_{pulse} >> T_e$), making J_e zero. During the pulse-off stage, the incoming plasma electron current will tend to return V_s back to V_f by Equation (5-7).

$$J_e = \frac{1}{4} q n_o u_e e^{-\frac{(V_p - V_s(t))}{T_e}} \tag{5-7}$$

However, if the pulse frequency is too rapid, the plasma electron current will not be large enough to reduce V_s back to the initial equilibrium before the next pulse begins. For this high frequency regime, some positive charge accumulates and $V_s > V_f$ 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 (Equation (5-7). This process repeats until a new equilibrium is established, with the time-averaged plasma electron current balancing the plasma ion and secondary electron currents:

$$\langle J_i \rangle + \langle J_{se} \rangle + \langle J_e \rangle = 0 \tag{5-8}$$

$$\langle J_i (1 + \gamma(V_i)) \rangle + \left\langle \left(\frac{1}{4} q n_o u_e e^{-\frac{(V_p - V_s(t))}{T_e}} \right) \right\rangle = 0 \tag{5-9}$$

Figure 5-4 depicts the transition from the initial equilibrium to the pulsing equilibrium.

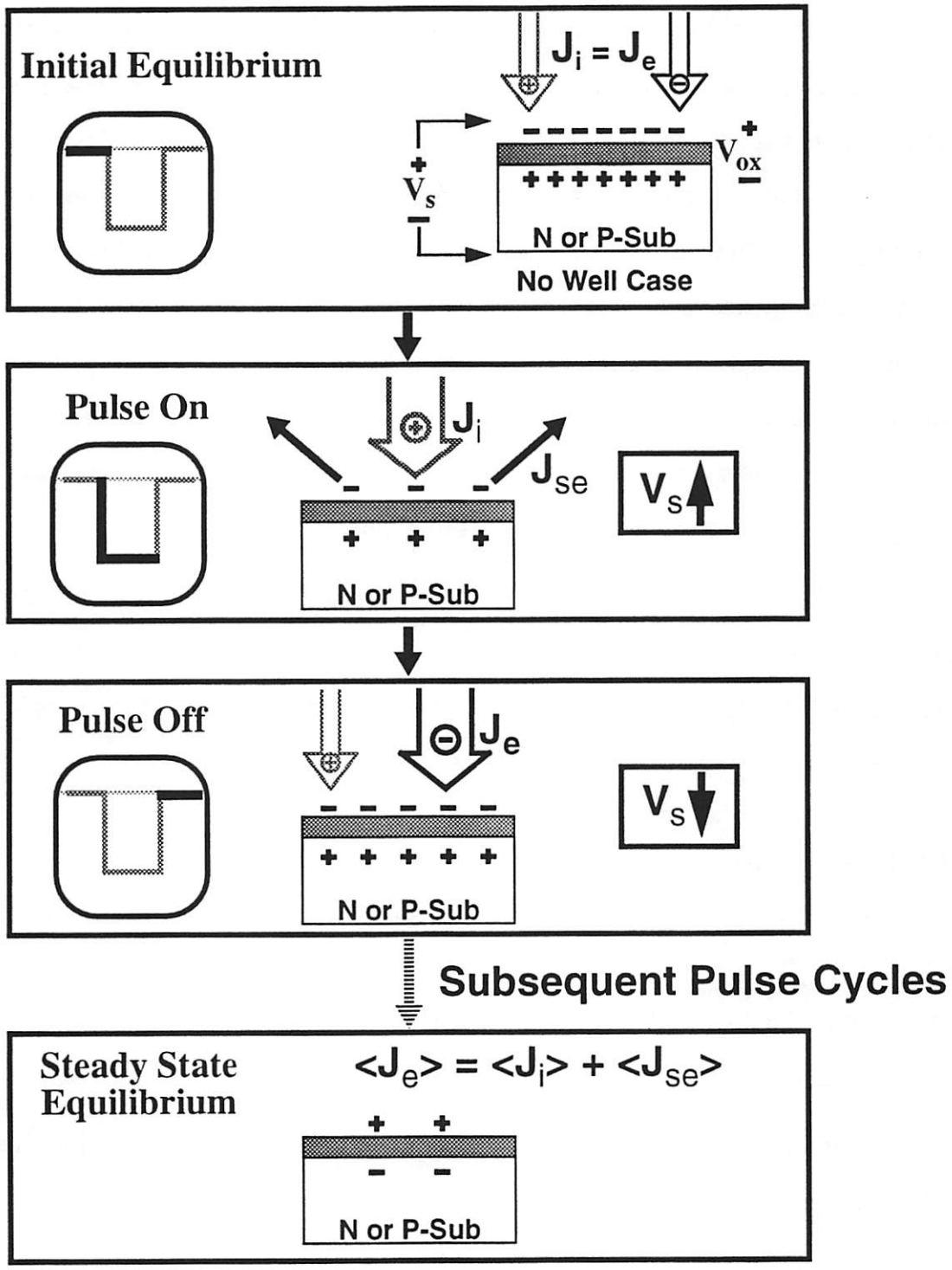


Figure 5-4 Transformation to Pulsing Equilibrium for PIII

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.

5.3.0.1 Substrate Bias Frequency Effects

The surface potential voltage that is required to balance the plasma electron with the plasma ion and secondary electron currents is highly dependent on the pulsing frequency (f_p). In the limit of $f_p \rightarrow 0$, the pulsing becomes negligible, the plasma electrons have plenty of time to neutralize the positive charge in between each pulse, and the equilibrium V_s approaches the floating potential (usually a negative value). As the frequency of pulsing increases, the time available for J_e to satisfy Equation (5-8) becomes successively smaller. To offset this reduction in time, V_s and V_{ox} increase in order to draw more electron current from the plasma. 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.

The frequency dependence of the pulsing equilibrium V_{ox} is graphed for the general wafer in Figure 5-5. Minimum damage results when $|V_{ox}| = 0$, which for the simulated 1μs/5kV ideal pulse, occurs at 150 kHz. This substrate bias effect on oxide charging has been experimentally confirmed by En [5-6].

5.3.1 Substrate Effect

The surface voltage of the gate equilibrates to approximately the same voltage, irregardless of the substrate type. But, some voltage may be dropped in the substrate if a depletion region is present, which will reduce the voltage across the gate oxide, reducing the damage.

The maximum steady-state voltage dropped in a depletion region for an inverted channel (i.e. $V_s > V_{threshold}$) is:

$$V_{deplo} = 2\frac{kT}{q}\ln\left(\frac{N_{ch}}{n_i}\right) \qquad (5\text{-}10)$$

where V_{deplo}, 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

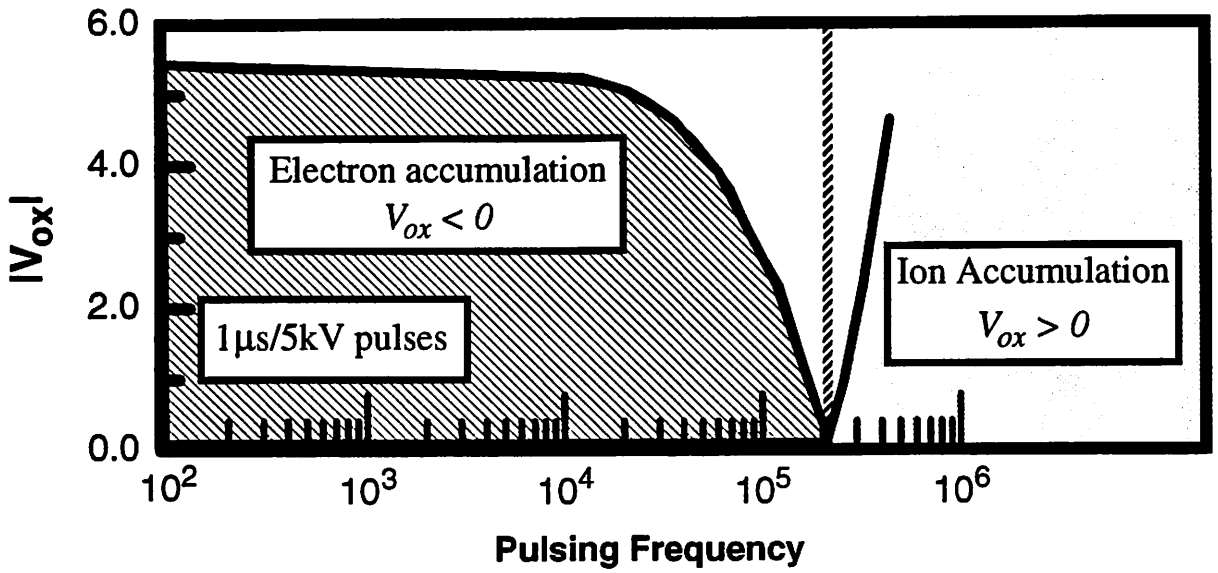


Figure 5-5 Substrate Pulsing Frequency Effect

Without any substrate bias, $|V_{ox}|$ is equal to the floating potential. As the frequency increases, the neutralizing time for plasma electrons decreases, requiring V_{ox} to rise in order to increase the plasma electron current. Eventually, the frequency becomes so high that to balance all the currents, V_{ox} changes sign from negative to positive, and increases rapidly. Simulation parameters are: 1μs/5kV pulses, 0.1μs rise and fall times, 10 nm gate oxide, $3.76 \cdot 10^{10}$ cm^{-3} Argon ion density, 4eV electron temperature, and a 13.23V plasma potential.

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} doped channel, V_{depl} 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 carriers. This results in a depletion width different than the steady-state, which decays to the steady-state value on the order of the carrier generation/recombination rate, which typically ranges from μs to ms. The larger depletion widths occurring in transient situations protect the gate oxide more than the steady-state depletion region, 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 include a depletion region and should show less exposure damage. During PIII pulsing, the voltage stress is negative for slow pulsing frequencies and positive for faster frequencies, and therefore the substrate type that shows less damage will flip with increasing pulsing frequencies.

For all simulations, the channel region under the 100Å gate oxide is doped 10^{17} cm^{-3}.

Figure 5-6 shows PIII damage for 11nm gate oxides on both N and P substrates after exposure to identical pulsing conditions, except for the different pulse frequencies. Plasma exposure was kept constant at 5 minutes, in order to isolate the dose rate effect, and remove simple plasma exposure damage as a variable. In this experiment, the pulsing frequency was never high enough to switch the surface voltage from negative to positive, and therefore the N-substrate oxides show lower damage for the entire frequency range. It is predicted that if the pulsing frequency could be raised further, the P-substrate would eventually exhibit lower damage than the N-substrate.

5.4 Well Structure Effects

Well structures are essentially p-n diodes, which can either be forward biased or reverse biased (Figure 5-7). In the forward biased mode, the well drops little voltage and is like a short. In the reverse biased mode, 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. Therefore, well structures on the wafer change the surface potential, thus altering the oxide charging damage. In the simulation two different well structures are compared, P-Well and N-Well. For each case, the substrate is doped $10^{15}cm^{-3}$.

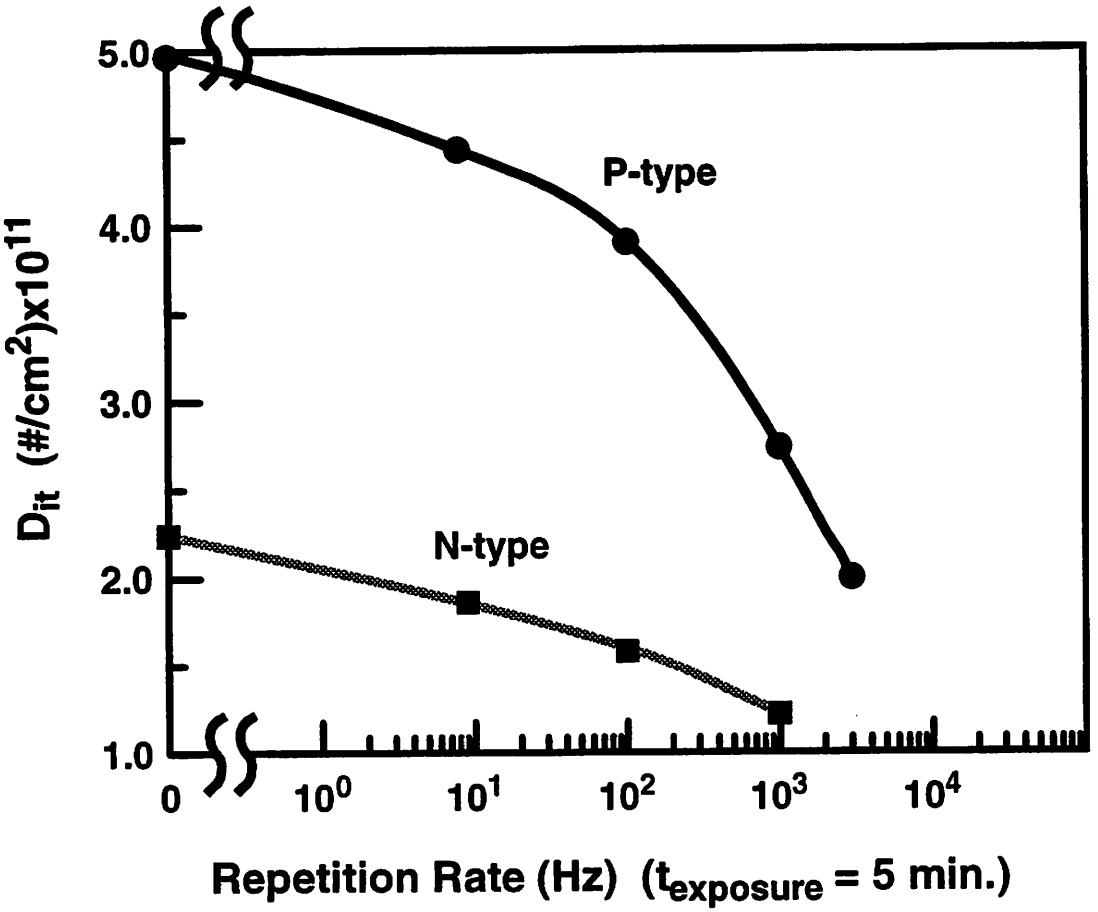


Figure 5-6 Damage Comparing N-Substrates and P-Substrates

The N-substrate shows lower damage for all during negative stressing because of the presence of a depletion region in the substrate. P-Substrates show lower damage for positive stressing. In this experiment, the relatively low pulsing frequencies resulted in negative stress for the entire range.

5.4.1 N-well

An N-well beneath the gate oxide effectively adds a diode in series (Figure 5-7). 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 $V_{ox} = V_s = V_f$. 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, and from Poisson's equation, with an abrupt, one-sided junction

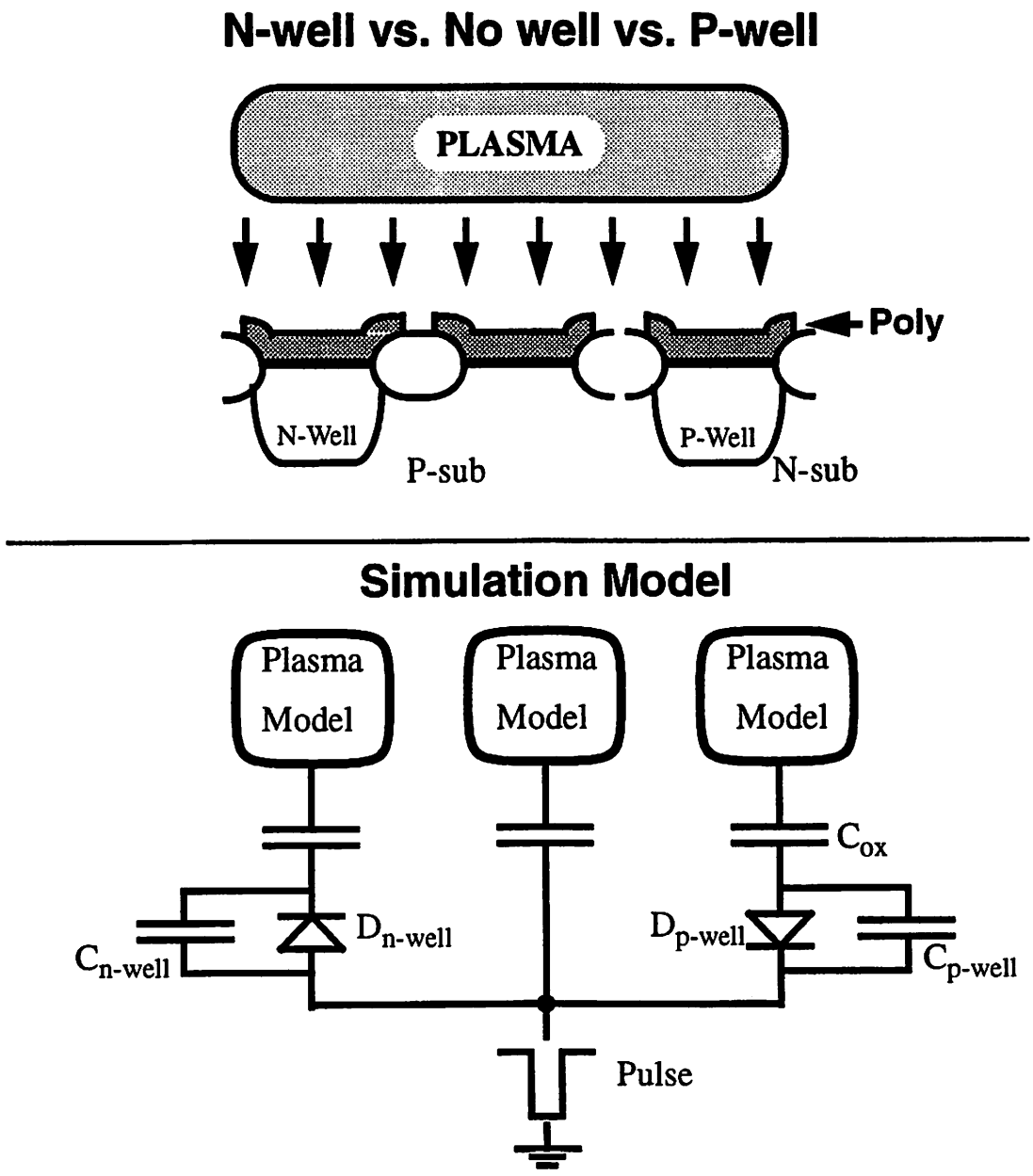


Figure 5-7 Well Simulation Model

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.

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

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 shown in Figure 5-8, when the pulse is turned off, the well voltage makes V_s more positive, increasing J_e. Additionally, because the reverse-biased capacitance of the low doped well junction is much less than the oxide capacitance, a small amount of stored charge will raise V_s significantly. 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. This results in a small ΔV_{ox} during pulsing, as compared to the no well case.

5.4.2 P-well

The initial equilibrium is the same the two previous cases, with $V_{ox} = V_s = V_f$, and $V_{well} = 0$. During the pulse, the positive charge deposition forward biases the P-well, which then drops a small forward voltage, and stores a correspondingly small amount of injected minority carriers (Figure 5-9). Then, when the pulse is turned off, the electron current deposits negative charge on the surface, which reduces the charge stored in the well and eventually reverse biases it, reducing V_s quadratically (Equation (5-11)). This super-linear reduction in V_s repels plasma electrons, decreasing the net J_e during the pulse-off stage. Therefore, in equilibrium V_s must be more positive than the no well case to compensate for the reduction in J_e, with a significant share of the extra voltage dropping across the gate oxide. This results in a larger ΔV_{ox} for the P-well case.

Table 5-1 summarizes the well effect during the initial equilibrium, pulse on, and pulse off stages. Simulated transient results for V_{ox} for the different well cases are shown in Figure 5-10. For each case V_{ox} begins at V_f, and then adjusts to a new equilibrium based on the frequency of pulsing, the duty factor, and the pulse voltage. The fig-

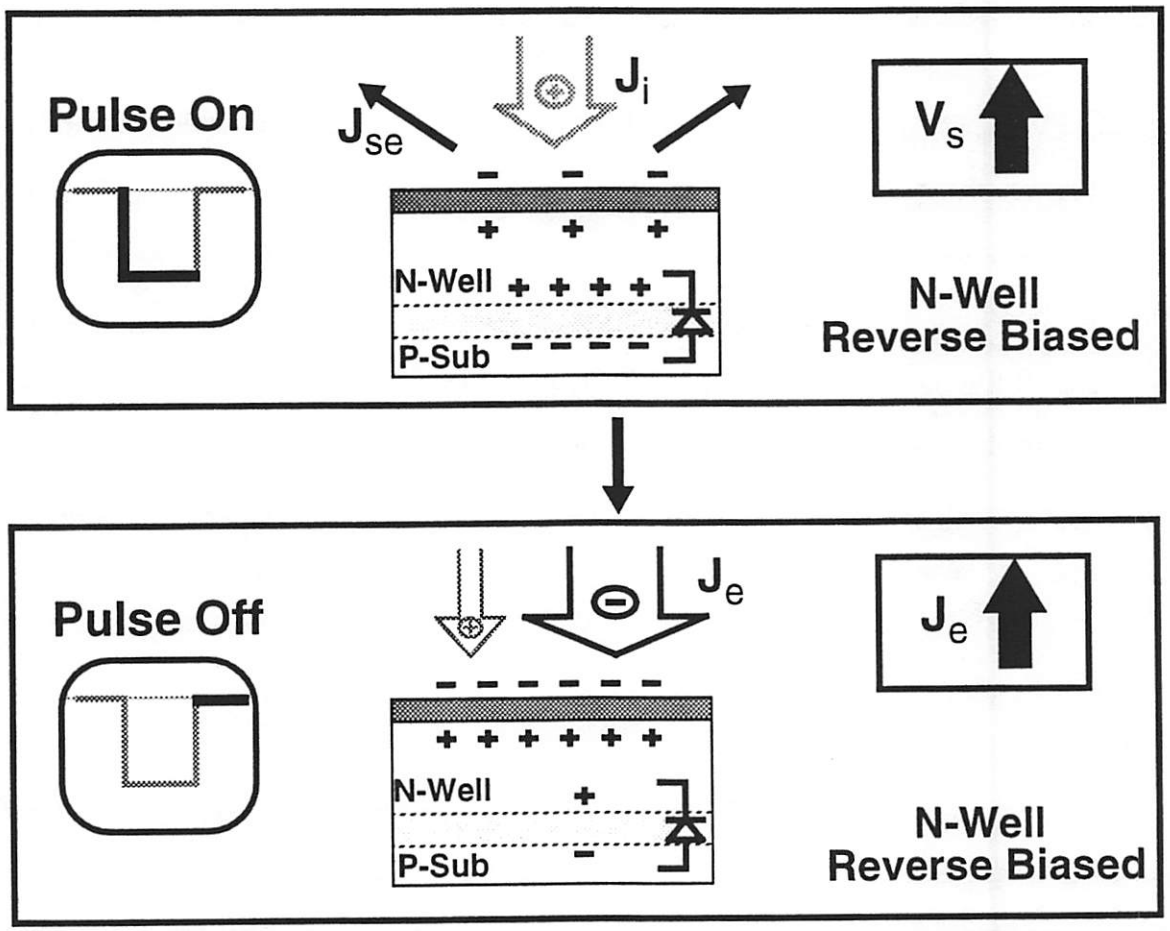


Figure 5-8 PIII Pulsing with an N-Well

The N-Well is reverse biased during PIII pulsing. The positive charge deposition reverse biases the well, which increases the surface voltage (V_s). Because of the relatively small capacitance of a well, a small amount of stored charge raises V_s considerably. This in turn increases the plasma electron current during the pulse off stage. Therefore, the well supports the extra voltage rather than the gate oxide.

ure of merit 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}.

5.4.3 Leakage Current

In Figure 5-11, the well potential is shown. The N-Well is always reverse biased with an offset from zero, which is the main reason why ΔV_{ox} is so much smaller for the N-Well case (Figure 5-10). 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

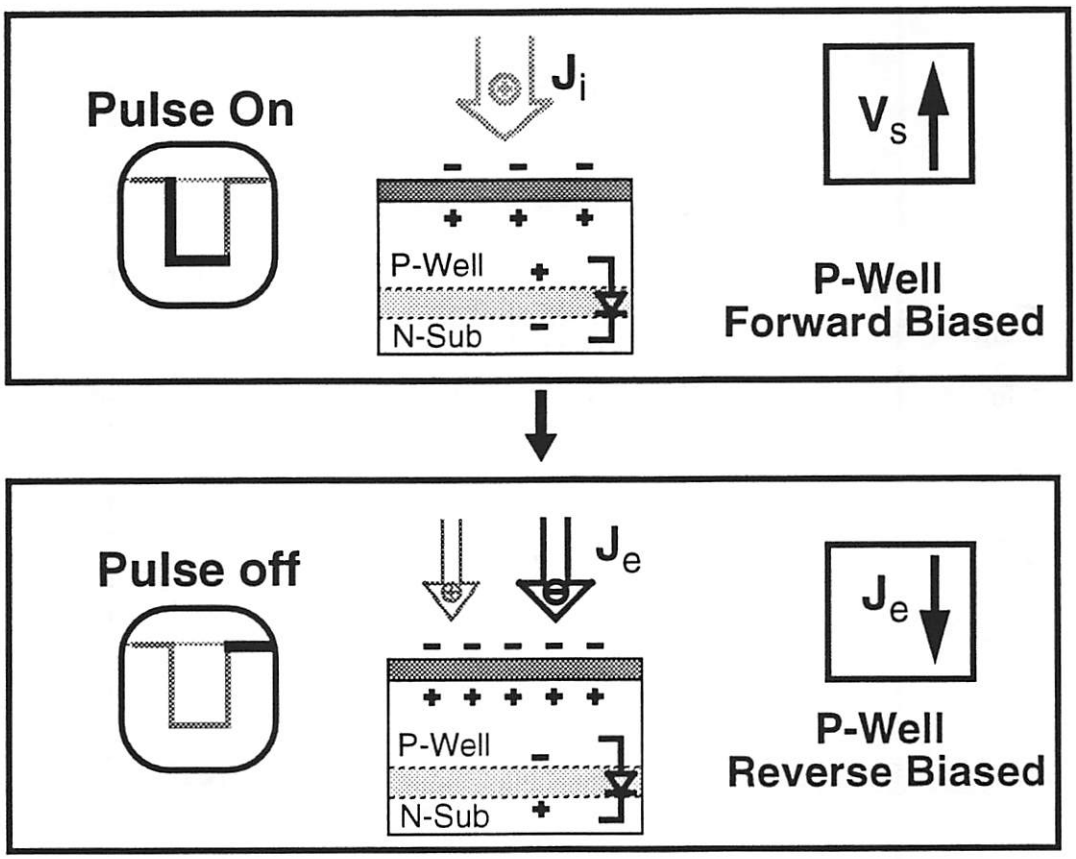


Figure 5-9 PIII Pulsing with a P-Well

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 5-1 Well Effect for the Three Stages of PIII

	N-Well	P-Well
Initial Equilibrium	no charge	no charge
Pulse On	Reverse-biased	Forward Biased
Pulse Off	Reverse-biased (Increases J_e)	Reverse-biased (Decreases J_e)

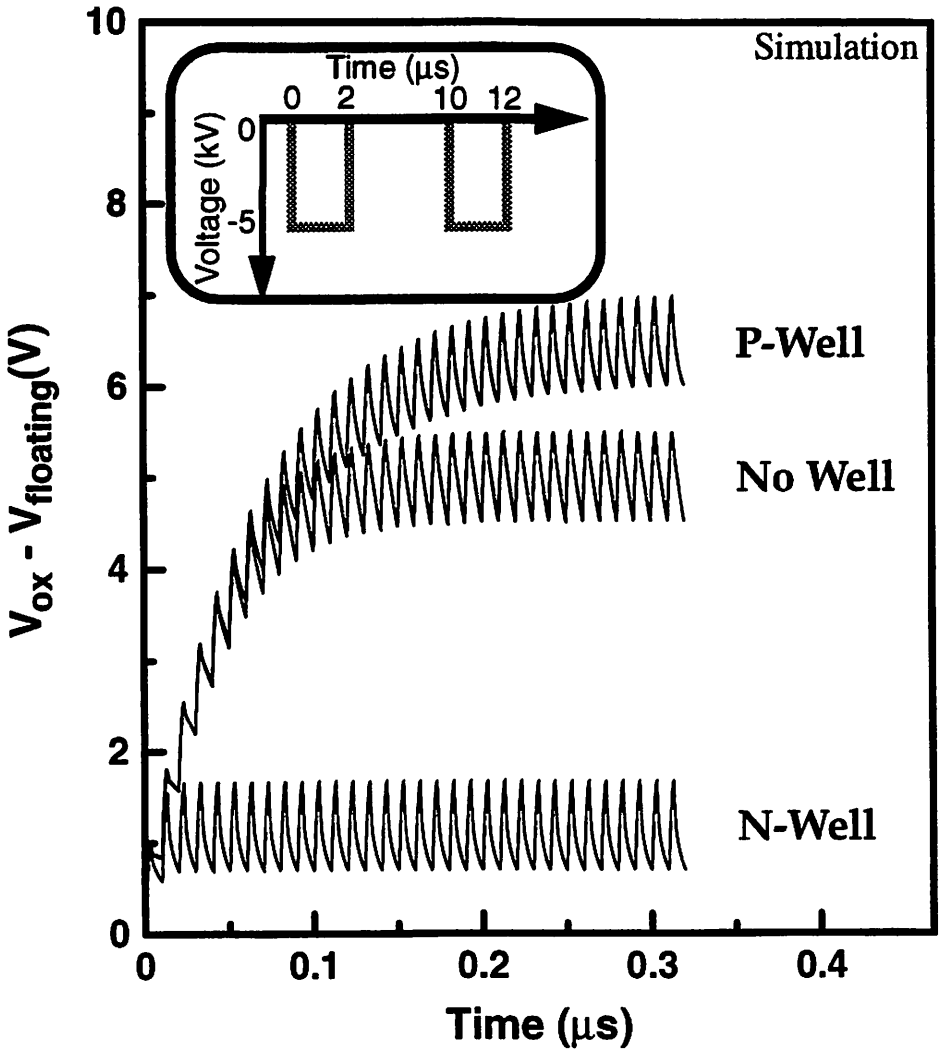


Figure 5-10 Well Effect on V_{ox}

Transient analysis of V_{ox} during pulsing with wells. V_{ox} begins at the floating potential, and then adjusts until the time average flux to the surface is equal to zero. The effect of the wells is to alter the equilibrated V_{ox}. The P-Well results in a more positive V_{ox}, while an N-well results in a more negative V_{ox}.

case (it switches from forward to reverse biased with each half cycle), the P-Well effect 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 is highly dependent on the light intensity during the plasma processing. Without light, leakage is low; for the wells fabri-

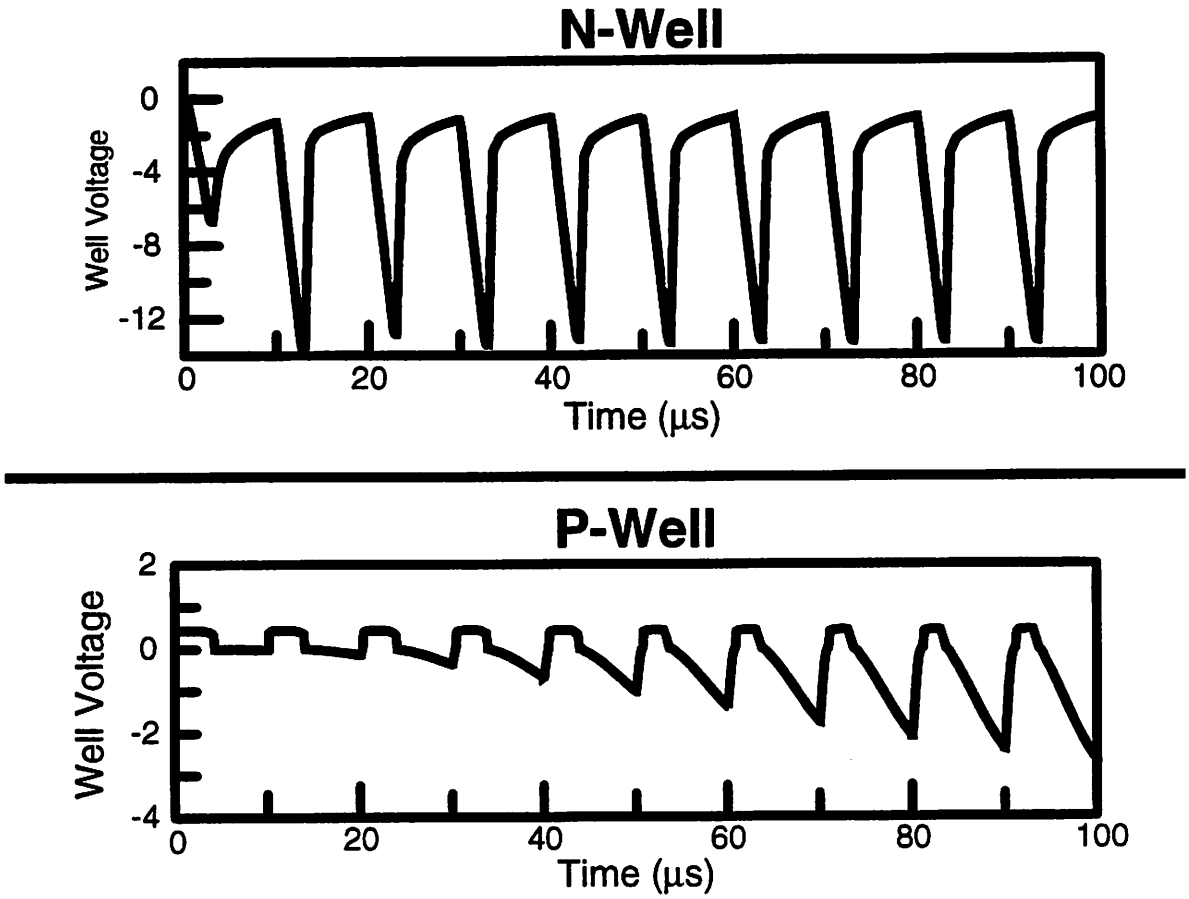


Figure 5-11 Well Voltage during Processing

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 offset, 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.

cated the leakage was less than 1μA/cm^2 at -5V. With unobstructed illumination, the leakage jumps by 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, LOCOS 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 nullified by leakage under high wafer illumination, with low absorption by overlaying layers.

5.4.4 Well and Substrate Effect

In the previous sections, the well and substrate effects were de-coupled, but to form an accurate model they must be combined. Whenever there is an N-well, the channel will be 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 contains p-type channel doping, which will have a depletion region when $V_{ox} > 0$. Therefore, both the substrate and well effects will occur during positive stressing, which occurs for PIII at very high frequencies.

5.4.5 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 11nm gate oxides were implanted at varying frequencies. The ECR power was 900W, the pulse voltage was 2.5kV, the pulse width 1μs, the pulse fall time ~35 μs, and the pulsing frequencies were from 100Hz - 22kHz. C-V measurements quantified the damage for each condition (Figure 5-12). All four curves follow the same trend, initially showing slightly higher 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 [5-7]. Then as the frequency increases further, the damage falls as predicted by simulation (Figure 5-5). 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 the surface voltage from negative to positive, and therefore the P's never showed less damage, as is predicted for very high frequencies (Figure 5-5). We expect the substrate and well effects to be more prominent when combined with antennas.

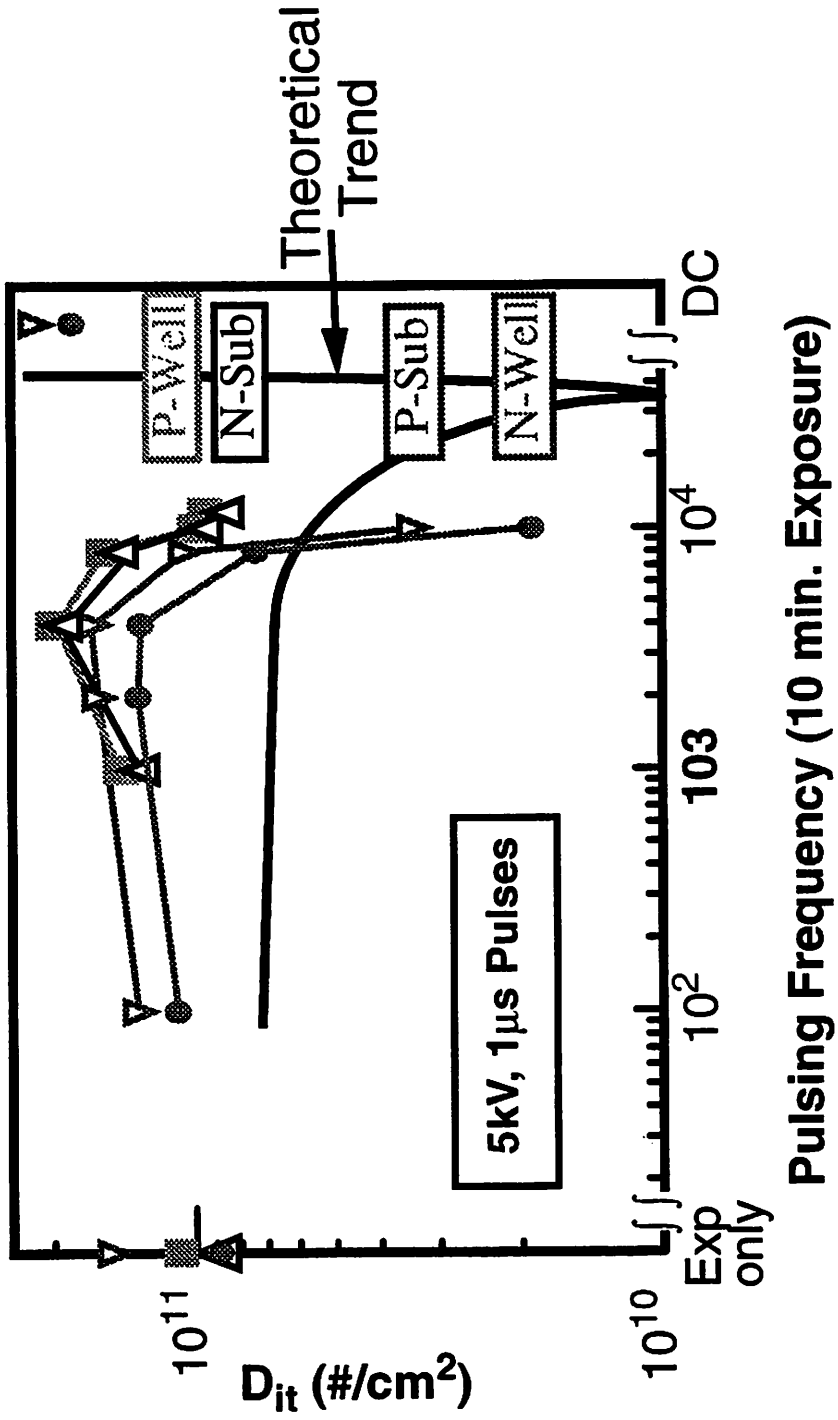


Figure 5-12 Experimental Data for Well Effect

Generated interface traps for 4 different structures: N-Well/P-Substrate, and P-Well/N-Substrate. All 4 curves follow the same general trend predicted by the model. Because of a depletion region, the n-doped channel region devices exhibit less damage.

5.4.6 The Effect of Different Well Structures

The well effects shown in the simulations are highly sensitive to the capacitance of the well. A larger capacitance well will support less voltage for the same charge, reducing the effect. 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 5-13). The lower doped side dominates the capacitance of the junction. The high doping on both sides of the triple well junction results in the highest capacitance, and the least amount of well effect.

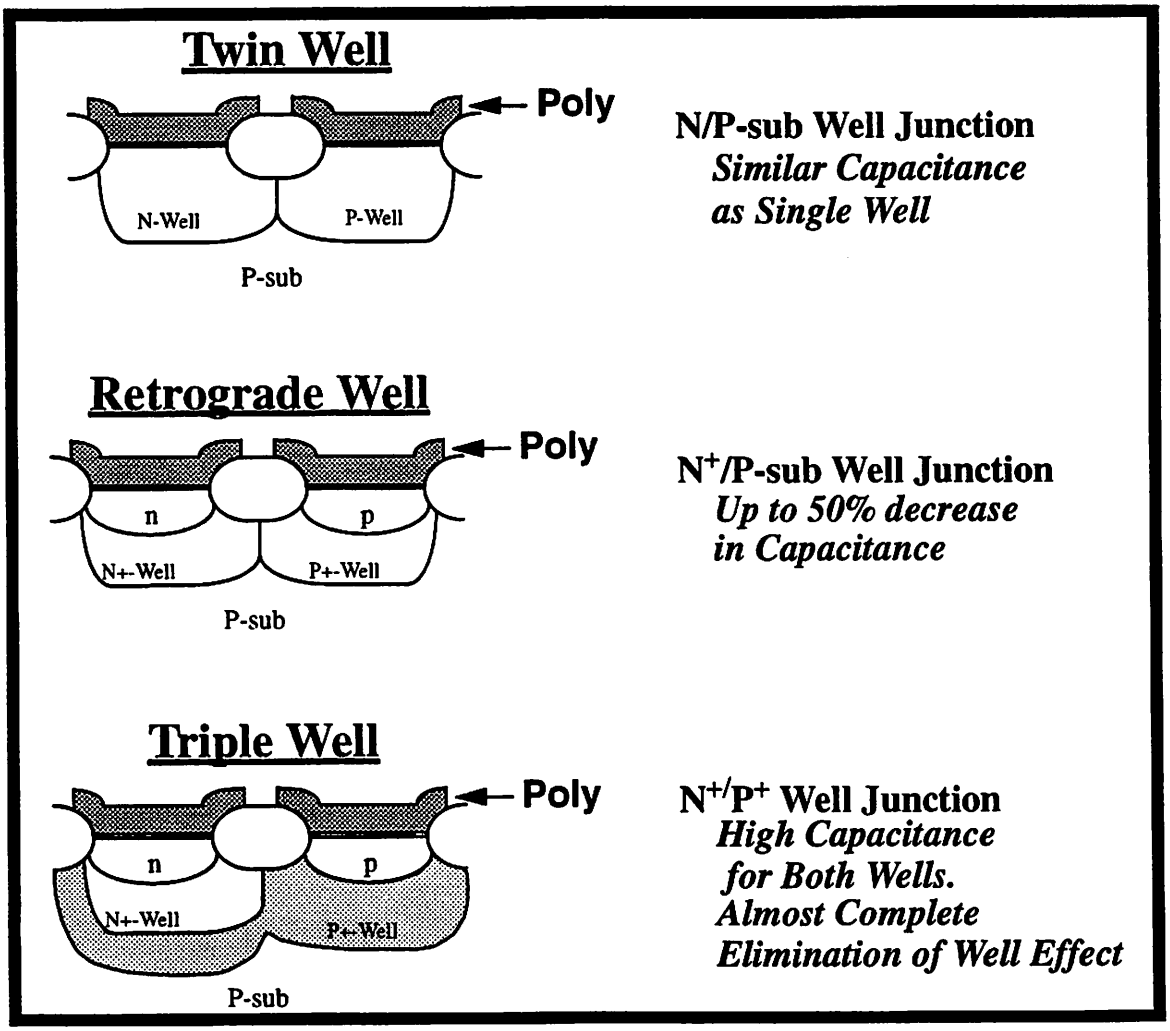


Figure 5-13 Well Effect Comparisons

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

5.5 Charging Damage and Dielectric Substrates

Using the fully coupled SPICE model, charging damage is simulated for dielectric substrates, such as occurs with thin film transistors and silicon on insulator technologies. The model assumes fully insulating substrates with negligible leakage currents. As in the well and substrate simulations, the time-average equilibrium surface potential is determined by the pulsing conditions. The effect of sub-surface structures is to 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. This does not hold when the gate oxide is attached to a charge collecting antenna, as is described in the following section.

5.6 Antenna Effect

In real wafers, 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 LOCOS oxide (Figure 5-14). The capacitance of the LOCOS oxide is much less than the gate oxide, leading to a different surface voltage across the wafer for uniform charge deposition across the wafer. Charge will then flow from the interconnect to the gate to equalize the voltages. If the electric field across the gate oxide exceeds the tunneling field, stress and damage result. This funneling of charge from a large collecting area (the antenna) to the gate oxide is called the antenna effect.

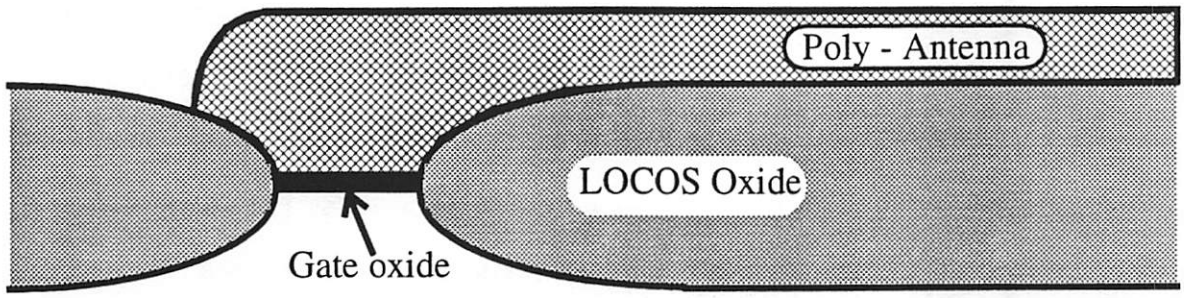


Figure 5-14 Typical Antenna Structure.

The capacitance difference between the LOCOS oxide and the gate oxide results in charge transfer to the gate oxide, increasing the gate oxide damage.

5.6.1 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 LOCOS oxide, and the ratio of the areas of the gate and the interconnect:

$$V_{ox} = \frac{(AR+1)\,Q_{dep}}{AR \cdot C_{ant} + C_{ox}} C_{ox} \tag{5-12}$$

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 ($\frac{\text{antenna area}}{\text{gate area}}$). 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 manifests from the charge deposited from single pulses. The antennas generally act as voltage like sources, since the amount of charge deposited on the antenna exceeds the charge tunneling through the gate oxide by an order of magnitude. This is shown in Figure 5-15, which shows a simulation of tunneling current from an instantaneous $10^{12}/cm^2$ deposition across the wafer, with a gate oxide of 5nm, LOCOS of 200nm, and an antenna ratio of 100. The peak electric field is 13.5 MV/cm, which decays to 10.5MV/cm in10μs. After the first couple microseconds, the tunneling current is reduced dramatically, keeping the

electric field above 10MV/cm. This response is nearly universal, with the tunneling current quickly bringing the electric field down to ~10MV/cm irregardless of antenna size.

Since the antenna effect occurs during single 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.

5.6.2 Dielectric Substrate Antenna Effect

The thick buried dielectric (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 capacitance in the gate regions is not simply the capacitance of the gate oxide, but rather the capacitance of the gate oxide 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 across the gate oxide/Box system to have the voltage equal to the field oxide/BOX value. Therefore, poly antennas should not increase gate oxide damage significantly, since the BOX layer supports the extra voltage generated by the antenna, not the gate oxide. These results have been confirmed experimentally [5-8, 5-9].

5.6.3 Well Antenna Effect

.In general, the 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 LOCOS and gate oxide. A spatially varying V_S results in charge transfer from the low capacitive region (LOCOS) to the high capacitive

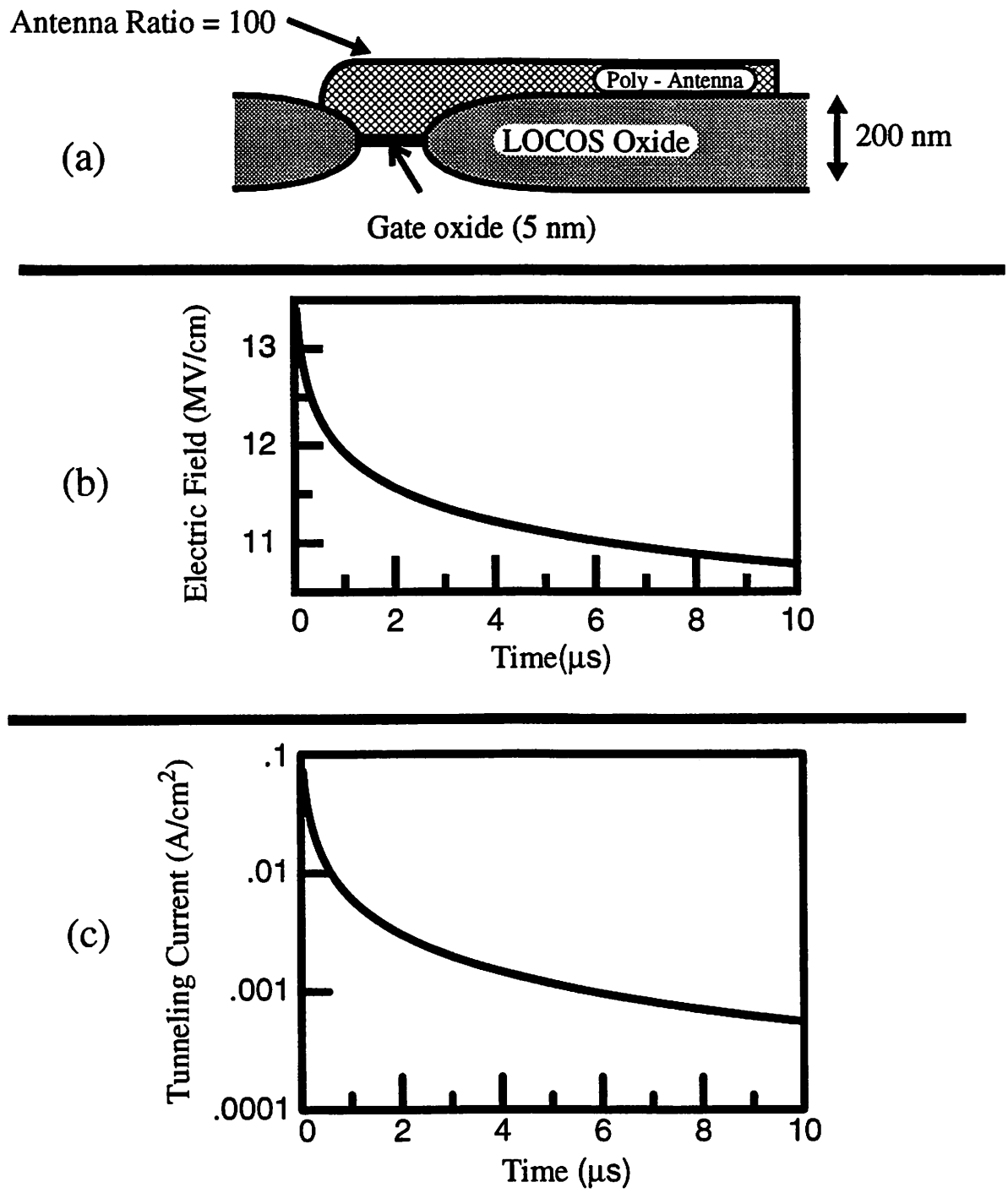
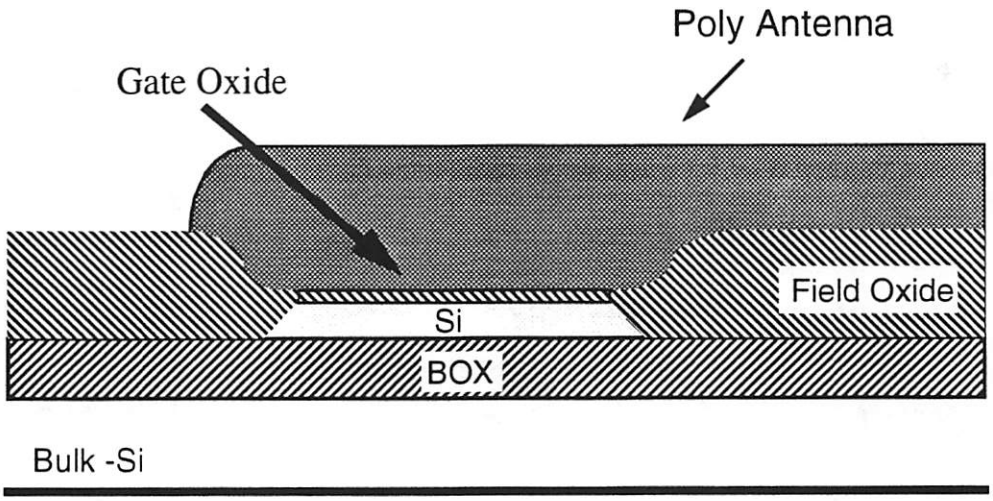


Figure 5-15 Simulated Tunneling Current and Gate Voltage with Antenna

Simulation for an instantaneous charge deposition of $10^{12}/cm^2$ across an antenna with a LOCOS of 200nm and an antenna ratio of 100. The antenna is connected to a 5 nm gate oxide. (b) shows the electric field across the gate oxide, while (c) shows the tunneling current through the gate oxide. The gate oxide quickly conducts enough charge to reduce the field to about 10MV/cm, at which time the antenna then acts like a voltage source, rather than a current source.



Simulation Model

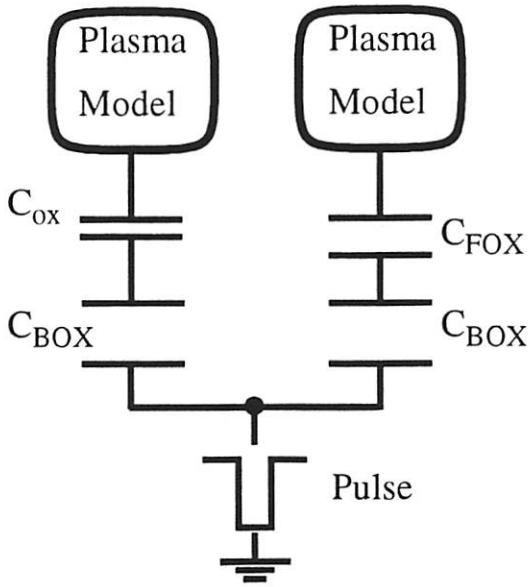


Figure 5-16 SOI Poly Antenna Device and Model

With SOI devices, a buried oxide resides beneath all device structures adding a parallel capacitance.

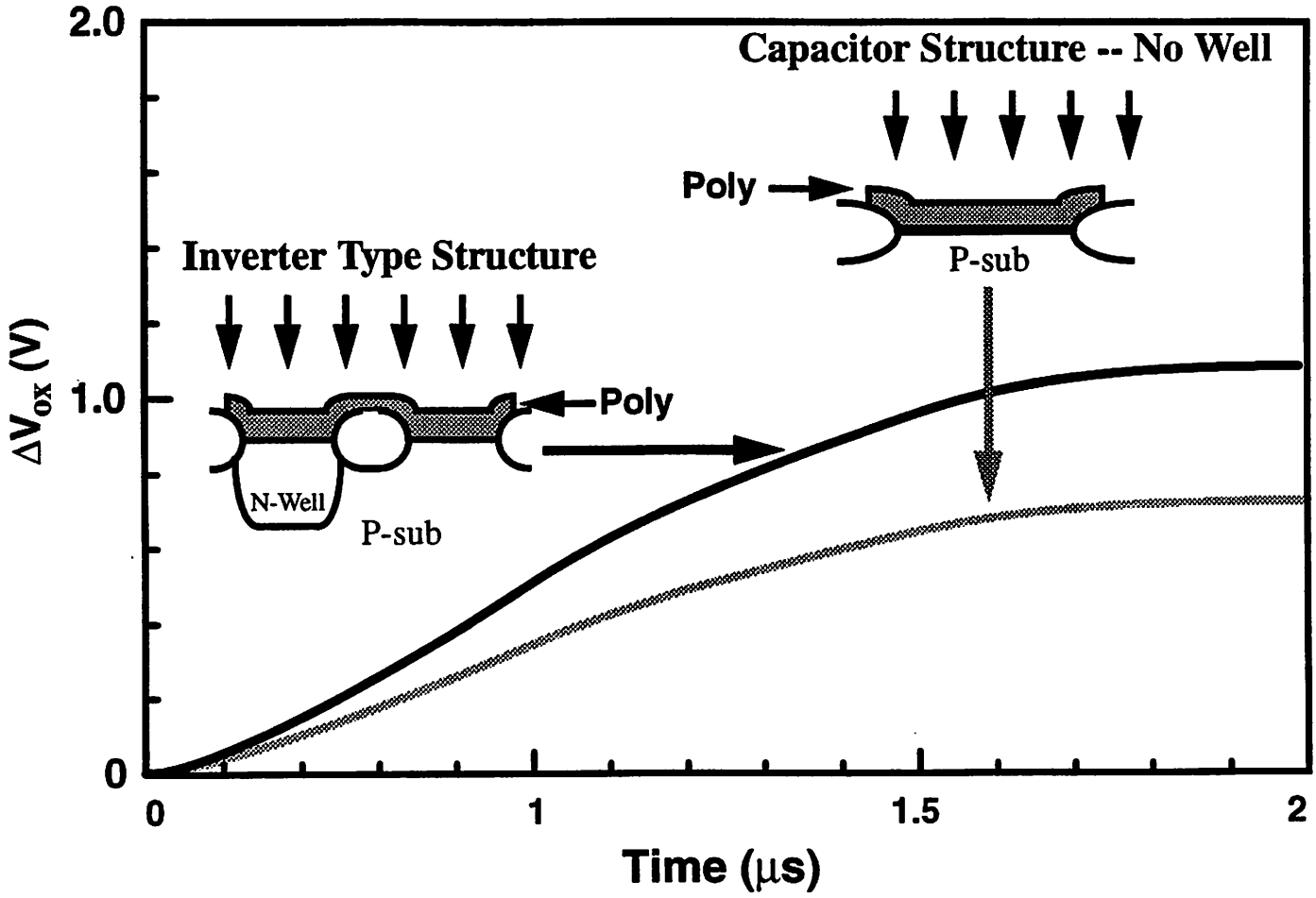


Figure 5-17 Well Antenna Effect

Change in oxide voltage as a function of time due to a pulse voltage bias for a thin oxide over P-type Si and a thin oxide over P-type Si electrically connected to an oxide over an N-well. The well capacitance generates an antenna like effect that enhances the P-substrate gate oxide voltage, as compared to the no well case. The simulation conditions are a 5kV/ 2μs pulse with 100nm gate oxides.

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

$$C_{eff} = \frac{1}{\frac{1}{C_{ox}} + \frac{1}{C_{well}}} \qquad (5\text{-}13)$$

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 the well 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 N-well gate (lower capacitive region) to the P-substrate gate (higher capacitive region), increasing the stress for the P-substrate oxide. In Figure 5-17, the well antenna effect nearly doubles V_{ox} for the P-substrate gate, 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.

5.7 Conclusion

Gate oxide charging continually presents a question mark for plasma processing, especially PIII. As gate oxides become thinner, concern becomes greater. In order to measure gate oxide charging damage, interface trap densities are extracted by C-V methods. With plasma exposure, the wafer surface potential equals the plasma floating potential, usually a negative value. If V_f is great enough, oxide damage may occur with 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 current during the pulse. The faster the pulsing frequency, the more positive the equilibrium surface potential must be to attract additional plasma electrons. Since the initial equilibrium voltage is negative (it is V_f), as the surface potential becomes more positive it must go through zero at some pulsing frequency. This frequency is usually quite high, 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 nega-

tively, while a P-well oxide charges more positively. The amount of well effect present depends on the leakage rate, but as long as the rate is longer than the pulse width, some effect will be present. 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. Overall, wells and substrate type can have an impact on oxide charging, and must be considered in the formulation of a global charging model.

Antenna effects arise from spatially varying surface potentials resulting from spatially varying capacitances across the wafer. These capacitance variations arise in a number of ways, either from capacitive differences from gate oxide to LOCOS or from N-Well to P-Well. Charge transfers from the low capacitive regions to the high capacitive regions equalizing the voltages. These currents increase the voltage stress for the gate oxide.

Through simulation and experiment gate oxide charging can be understood and controlled in plasma processing and PIII.

References

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

[5-2] W. En and N. W. Cheung, "Analytical Modeling of Plasma Immersion Ion Implantation Target Current Using the SPICE Circuit Simulator," *J. Vac. Sci. and Tech. B*, vol. 12, no. 2, pp. 833-837, 1994.

[5-3]E.H. Nicollian, and J.R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc, New York, pp. 332, 1982.

[5-4]C.N. Berglund, "Surface states at steam grown silicon-silicon dioxide interfaces", *IEEE Transactions on Electron Devices*, vol. 13, pp.701-710, 1966.

[5-5] E.H. Nicollian, and J.R. Brews, "MOS Physics and Technology." John Wiley and Sons Inc,

New York, pp. 319-372, 1982.

[5-6]W. En. B.P. Linder, and N. W. Cheung, "Modeling of Oxide Charging effects in plasma processing," *J. Vac. Sci. and Tech. B*, vol. 14, no. 1, pp. 552-559, 1996.

[5-7]K.Scheugraf and C. Hu, "Reliability of Thin SiO_2", *Semicond. Sci. Tech.* 9, pp. 989-1004, 1994.

[5-8]M. J. Sherony, A.J. Chen, K.R. Mistry, D.A. Antoniadis, and others, "Comparison of Plasma-Induced Charging Damage in Bulk and SOI MOSFETs", *1995 IEEE SOI Conference Proceedings*, pp. 20-21, 1995.

[5-9]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.

6 Conclusion

Construction of the plasma model has a wide array of applications in plasma processing. The complete model merges three different sub-sections: 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. The wafer structure model is a translation of the structures into circuit equivalent devices. Finally, the substrate bias is modeled as a voltage or current source with parallel and series non-ideal elements. The model may be solved in a fully-coupled mode, where all three sections are combined and solved simultaneously. If the substrate is conducting (i.e. a simple bulk silicon wafer), the model may be solved in a de-coupled mode, where the plasma ion currents and plasma sheath thicknesses are solved independently of the wafer structure model. The de-coupled mode affords a magnitude increase in computational speed. Two platforms have been used for the modeling work, MATLAB and SPICE. SPICE allows easy incorporation of extra circuit elements and the modeling of the wafer structures, and solves the fully coupled mode best. On the other hand, MATLAB contains more flexible differential equation solvers, but lacks the built in circuit models of SPICE.

Gate oxide charging during plasma processing, and more specifically to PIII, is one application of the model. During plasma processing, the plasma initially biases all

floating surfaces (including gate oxides) to the plasma floating potential, usually a negative value. During pulsing, the plasma ions and ejected secondary electrons deposit positive charge on the wafer surface. In between pulses, the plasma electrons attempt to neutralize this positive charge. The steady-state gate voltage during processing is that which balances all of the plasma currents such that no net current reaches the wafer. For PIII, low pulsing frequencies induces a negative gate bias. As the pulsing frequency increases, the gate bias becomes more positive, passing through zero, and eventually becoming highly positive.

By exposing large area gate capacitors to a varying pulsing frequencies, the charging damage trends predicting by the simulation have been confirmed. Comparing before and after quasi-static C-V measurements provides a tool for monitoring gate oxide damage. As pulsing frequency increases, charging damage initially increases due to the wafer heating, than starts to decrease as the gate bias becomes more positive and approaches zero.

The presence of a depletion region beneath the gate oxide diminishes charging damage, and therefore the level of damage correlates to substrate type. N-substrates contain a depletion region during negative stressing, and P-substrates during positive stressing. Experimental data confirm the substrate effect. The presence of wells also affect charging damage. N-Wells result in more negative stressing of the gate oxide, while P-Wells cause more positive stressing.

With PIII, the model can compute the amount of implanted ion energy spread. There are three sources of energy spread in a standard PIII implant: excessively long rise times, matrix sheath implantation, and fall time implantation. Implantation with dielectric substrates adds two more energy corrupting sources. Part of the applied bias couples directly to the dielectric substrate, and during the implant a significant voltage builds-up across the substrate opposing the substrate bias and reducing the implant volt-

age. By understanding the sources of energy spread, the pulsing conditions can be properly optimized to obtain acceptable spreads.

As oxides continue to scale, gate oxide charging becomes a larger issue. Oxides less than 5nm thick begin to operate in the direct tunneling regime, amplifying the tunneling current for a constant electric field stress. For a constant voltage stress, the electric field increases proportionally with thickness reduction, increasing the tunneling current even more. The saving grace for ultra-thin oxides is their higher charge to breakdown. To determine whether ultra-thin oxides suffer more damage, the increase in tunneling current must be weighed against the increase in damage immunity. By incorporating the correlation between tunneling current and actual damage, the model will be able to predict the damage trends with ultra-thin oxides.

Currently the model is limited to pulse biases less than 1Mhz. Additional modules need to be added to accurately model RF type substrate biases. The effect of non-ideal source effects, and the effect of the matching network on charging damage may be explored further.

In, summary, the plasma model aids in the prediction of gate oxide charging damage, and for PIII, the implantation dose and implant energy characteristics. Through the model, the plasma characteristics and the substrate bias may be optimized to achieve the desired process results.

Appendix A

Symbol Page

A	Area.
AR	Antenna Ratio which is the Antenna area divided by the gate area.
C	Generalized capacitance.
C_{BOX}	Capacitance of the buried oxide found in SOI devices.
C_{FOX}	Capacitance of the field oxide.
CV	Capacitance Voltage curve.
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}	"Damaged" Quasi-Static capacitance of a 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.
ΔV_{ox}	Change in the gate oxide voltage from intial equilibrium.
D_{it}	Interface face trap density. The units are cm-2ev-1 or cm-2 depending on the context.
D_{well}	Name for diode in well model.
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_i	Plasma ion current
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_{se}	Secondary electron current density.
K_1	Materials constant for direct tunneling current calculations.
K_2	Materials constant for Fowler-Nordheim tunneling current calculations.
M	Ion Mass.
N_{ch}	Channel doping concentration.
N_{sub}	Doping concentration of the substrate.
P_{low}	Percentage of ions that implant with less then Vpulse during the pulse.
Φ_s	Potential of the Silicon/Oxide interface.
QS	Quasi-Static
Q_{bd}	The charge to breakdown of a gate oxide.
Q_{dep}	Total charge deposited per pulse.
Q_{well}	Charge in the well junction depletion region.
R_{sh}	Shunt Resistance Value in the pulse supply matching network
T	Temperature.
T_e	Electron temperature in electron volts.

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.
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_{sub}	Voltage dropped across dielectric substrate.
V_{well}	Voltage dropped across the well junction.
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	Pulse frequency
γ	Secondary electron yield per impinging ion.
k	Boltzmann's constant.
k	Secondary electron yield constant relating implant voltage and yield.
m	Electron mass.
n_{diff}	Number of ions that diffuse across the sheath boundary.
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 ins that diffuse across the substrate during the fall time.
n_i	Plasma Ion Density, or intrinsic carrier density in Silicon.

n_{matrix}	Number of ions in the matrix sheath.
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_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.
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

De-Coupled Library Example

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.

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 $n_i=5\bullet10^{10}/cm^3$, V_f = -5.5, V_p = 13.23, T_e = 4, and the gas is Argon. The applied bias is a -2kV / 100kHz

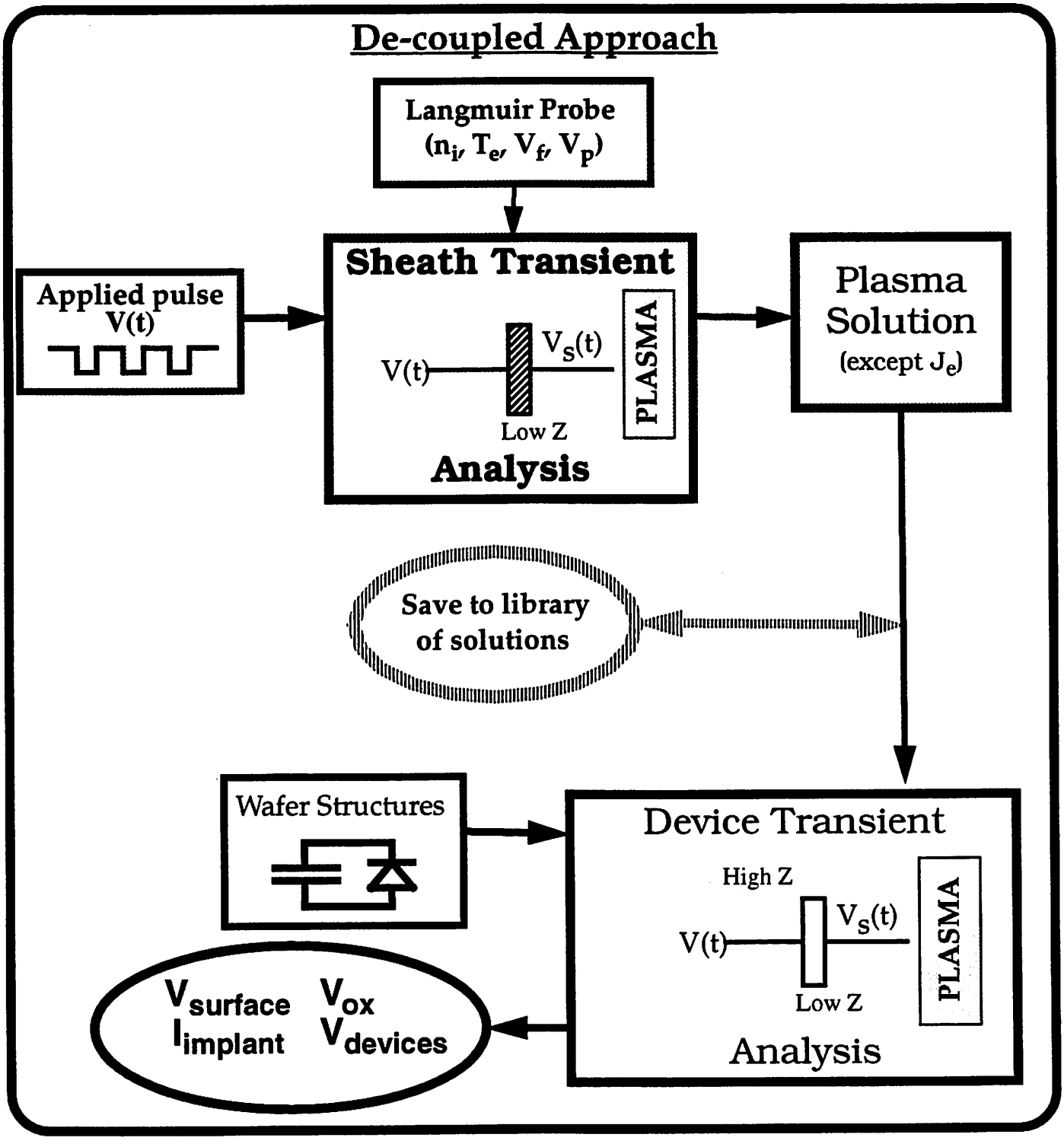


Figure B-1 De-Coupled Modular PIII Model

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. 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, allowing them to be computed only once. Then, for each different wafer structure set-up, the library is accessed for the plasma currents.

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-2). This solution is stored to disk for later retrieval.

Now that the sheath 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-3 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-3. 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 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

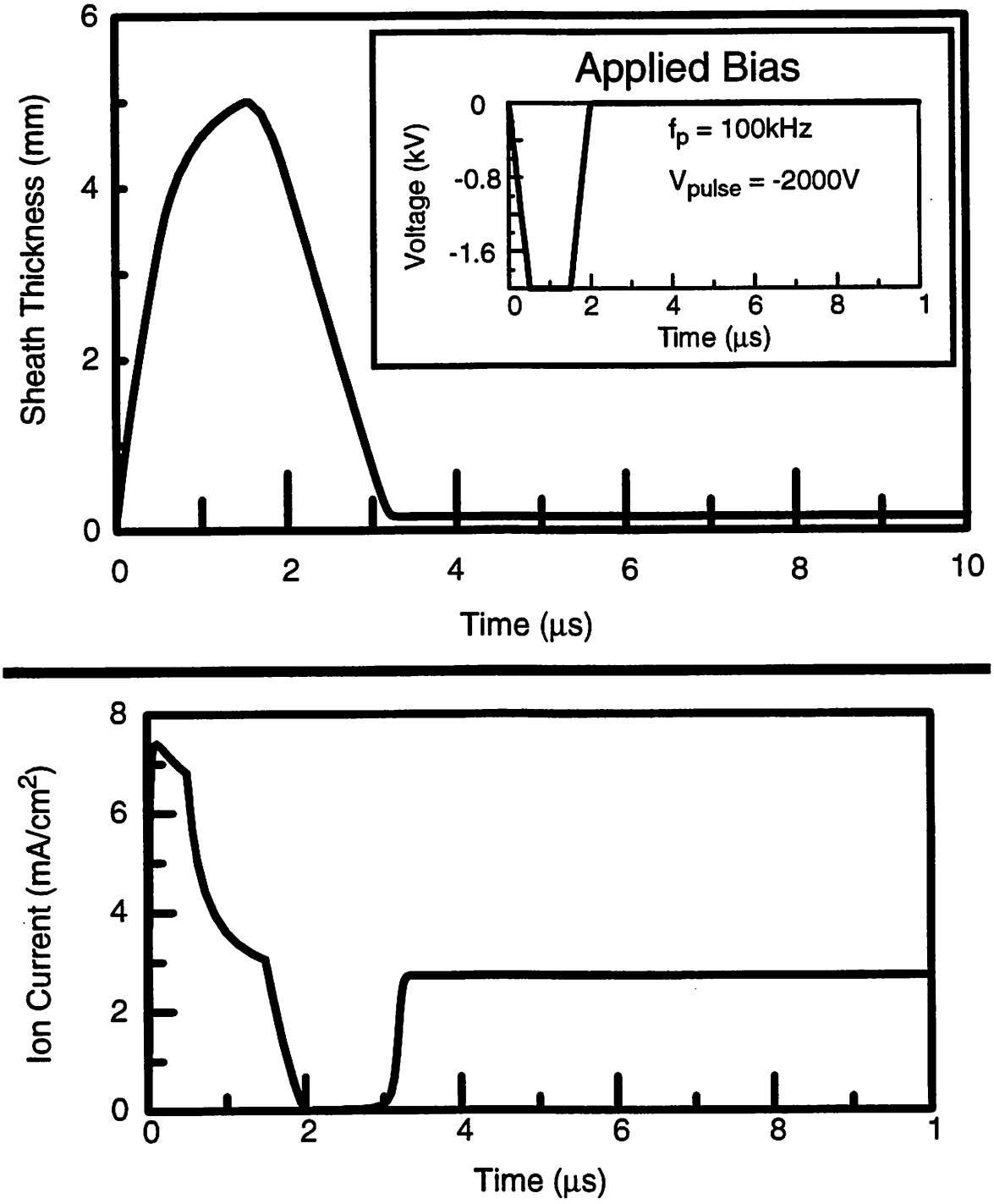


Figure B-2 Sheath Transient Analyzer Output

The Sheath Transient Analyzer output for a typical plasma condition (n_i=5•10^{10}/cm^3, V_f = -5.5, V_p = 13.23, T_e = 4) and a -2 kV/100kHz applied bias. The sheath solution shows sheath expansion to about 5 mm before the pulse ends. The ion current follows a typical curve, with 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 be the Device Transient Analyzer.

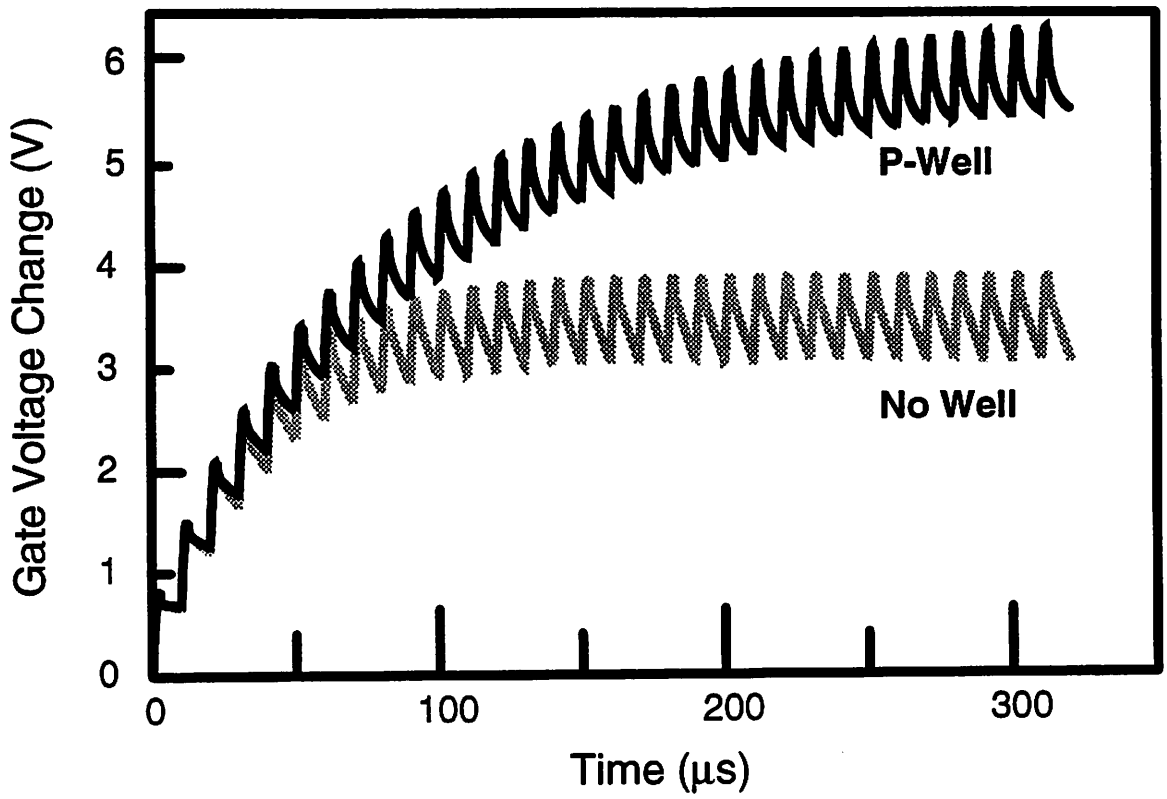


Figure B-3 Device Transient Analyzer Output

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.

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.

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 inputted innumerable times as the 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.