

**Novel Charge Integrating Pulsed I(V) Technique: A Measurement of Fowler-Nordheim
Currents through Thin SiO₂ Films**

J.C. Poler, W.S Woodward and E.A. Irene

Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

Abstract

The design, characterization and applications of a novel charge integrating pulsed current-voltage I(V) measurement are described. Tunneling transport through thin metal-oxide-semiconductor (MOS) capacitors is measured over ten orders of magnitude of current. Short pulse widths ($<1\mu\text{s}$), allow electrical characterization of these films under high current densities, without significant charge injection. A study of the quantum interference of electrons during Fowler-Nordheim (FN) conduction, is used to illustrate the measurement.

DTIC QUALITY INSPECTED 4

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification _____	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Introduction

Future ultra large scale integration (ULSI) design rules require very thin, $\sim 60\text{\AA}$, gate oxides for MOS field effect transistor (MOSFET) technology. Characterization of the thin films and their interfaces in MOSFET structures is necessary to establish reliability and performance criteria. In this paper we show the development of a technique to probe the chemical and physical properties of the interfaces in MOS structures.

By applying a large bias, ($>4\text{V}$), on the gate of an MOS structure, charge is injected into the conduction band of the insulator. The resulting Fowler-Nordheim¹ transport mechanism can be described as:

$$I_{FN} = A F_{ox}^2 \exp[-C/F_{ox}] . \quad (1)$$

Where the Fowler-Nordheim (FN) current, I_{FN} , is related to the electric field across the oxide, F_{ox} , and the constants C and A are determined from the band structure of the device. Figure 1. shows that electrons can tunnel through the triangular potential barrier, and propagate ballistically through the conduction band of the insulator toward the semiconductor substrate held in accumulation. The abrupt change in potential at Si/SiO₂ interface, enable the coherent reflection of electron waves from the silicon surface. For thin films, ($\sim 50\text{\AA}$), where there are very few inelastic phase destroying collisions, the wave packets will interfere upon emergence from the barrier in the inverted triangular cavity. The tunneling transmission coefficient for this system was first derived by Gundlach².

The resulting interference pattern is displayed as oscillations superimposed on the

FN tunneling currents as was shown by Maserjian³ *et. al.* who first experimentally realized this quantum oscillation phenomena. These oscillations are sensitive to small changes in the electron potential near the collecting interface. We attempt to use the positions and amplitudes of the oscillations to elucidate the structure and composition of the Si/SiO₂ interface.

High electric fields (>10MV/cm) across the oxide film drive large amounts of injected charge into the device. This charge does not thermalize, but rather accelerates ballistically to a substantial, (>2eV), kinetic energy⁴. The hot electrons can cause permanent damage to the film which appears as excess tunneling currents in subsequent current-voltage, I(V), characterization. Hot electron degradation also wears out the insulating properties of the film and the structure tends toward pre-mature dielectric breakdown.

In order to avoid these problems during measurement, one needs to limit the number of energetic electrons injected into the sample. This need implies a requirement for very fast current measurements at high fields. To this end we have employed a double pulse technique in conjunction with a novel charge integrating pulsed acquisition board. This board can measure currents over ten orders of magnitude and acquire data in the high field regime using sub-microsecond pulses.

Experimental

Sample Description. MOS samples were prepared on p-Si(100) substrates. The wafers were RCA cleaned⁵ followed by a dip in hydrofluoric acid (HF). Films of SiO₂ were grown at 800°C in dry O₂, containing 2% HCl in a single walled tube furnace. The films were grown to a thickness of 50Å. Polycrystalline silicon was blanket deposited and doped. Aluminum was evaporated onto the wafer and gates were defined lithographically to areas of 2.03x10⁻³cm² yielding a capacitor structure of ~1300pF. Back side contact was also made by aluminum deposition. All the wafers received a post metalization anneal at 400°C in a mixture of 10% H₂ in N₂ for 20 minutes.

Hardware Description. The data acquisition hardware consists of seven major components as illustrated in figure 2.

I. V_{app} DAC. This digital to analog converter controls the amplitude of the bias pulse, V_{app}. V_{app} amplitudes of 3.000V to 8.120V may be selected with a resolution of 5mV. Although resolution requirements placed upon this DAC were modest (10bits) a much more stringent specification applies to DAC step size uniformity. Any non-uniformity of step size (also known as *differential linearity*) will appear as noise in the tunneling current data. Binary ladder circuit topology, the most common implementation for commercially available DACs, inherently suffers from differential linearity of a magnitude comparable to the DAC's resolution.

This effect is of particular concern in the current application due to the great sensitivity of tunneling current to V_{app}. Calculations show, that a variation of V_{app} equal to one DAC step size (5mV or 1/1024) causes a change in tunneling current of the same

order as the tunneling current oscillation itself. To reduce the differential linearity related noise in acquired tunneling data to an acceptable level (e.g. 1% of anticipated oscillation amplitude) using conventional DAC technology would require a V_{app} DAC of a resolution of 1 part in 100,000 or more than 16bits.

To avoid this effect entirely, the DAC used in this apparatus is implemented using a pulse width modulation technique which displays an inherent absolute uniformity of step size. The 10MHz frequency reference (see TIMEBASE below) divided by 1024 to derive a 9766KHz waveform whose duty cycle can be varied digitally from 0 to 100% in increments of 1/1024. This waveform is applied to a high speed CMOS logic element (74AC74) supplied with precision ground and 5.000V references instead of the usual logic ground and 5V power rails. The element thus produces an output waveform with a time-averaged area (DC component) precisely variable from 0.000 to 5.120V. This signal is low pass filtered by a 3 pole (18db/octave) critically damped filter and level shifted to produce the 3.000 to 8.120 V_{app} reference.

II. CMOS Switch. High speed and high precision application and removal of the bias pulse to the device under test is the function of the CMOS switch. Implemented with a large number (24) of parallel connected high speed CMOS logic devices, this switch displays a low 'on' resistance ($<1\Omega$) and high speed (4ns).

III. High Speed Integrator. Accumulation of sample current is performed by a precision, high speed analog integrator. Under software control, the integrator performs a number of other functions including analog to digital conversion of acquired charge to 16bit resolution.

IV. Timebase. Generation and measurement of acquisition timing events (e.g. bias pulse duration and integrator analog to digital conversion ramps) are referenced to a single 10MHz quartz resonator.

V. Sample Resistance Measurement Circuit. At high tunneling currents, an important correction to V_{app} is needed to account for voltage drop across probe and sample resistances. The resistance measurement circuit provides a means of characterizing these resistances in the context of a shorted MOS capacitor. The resistance value so obtained may then be used to estimate the IR compensation needed for V_{app} .

VI. V_{ref} . All voltage levels developed by the data acquisition circuits are reference to precision voltages of +5.000 and -1.000 generated by the voltage reference circuit.

VII. PC Bus Interface. Control of the acquisition circuits is achieved through the execution of I/O commands issued by the IBM compatible microcomputer directing the experiment. Decode of these commands and generation of corresponding control signals is performed by the PC bus interface.

Hardware Characterization. It was necessary to be able to accurately and precisely apply pulses of know height and width. The DC pulse heights were measured with a high impedance electrometer. By adjusting potentiometers on the DAC we could set the voltage offset and gain such that there was less than 1mV deviation between the applied and measured signals. This accuracy was maintained to within 1-2mV of drift over

several weeks. Pulse width accuracy and reproducibility was examined using the high impedance input of a 400MHz digital oscilloscope. The pulses had rise and fall times of less than 4ns. The deviation of the measured pulse width to the applied pulse, t_{app} , was typically less than 0.2%, and the deviations were independent of the magnitude of the pulse. Figure 3 shows a typical oscilloscope trace for a pulse of height $V_{app} = 7V$ and width $t_{app} = 1\mu s$ (a), and the signal resulting from this pulse at the integrator summing point (b). The pulse is being applied to the substrate of a MOS capacitor with an accumulation capacitance of $\sim 1300pF$. There is $\sim 400\mu A$ of current being passed through the sample during this acquisition. Even under these circumstances, the pulse shapes seem well behaved.

The drive pulse typically exhibits a ringing on the rising and falling edges of $< 50ns$. When we probe the integrator's summing point, which is driven toward ground, we also observe a transient behavior. The deviation from ground and the extent to which it evolves in time is dependent on the capacitance of the device under test (DUT). As the capacitance increases, so does this charging transient. Using a double pulsed technique, as described below, we are able to accurately account for and eliminate the effects of these transients on our data acquisition.

It was also necessary to determine the charge collection and integration capabilities of the acquisition board. The accuracy of these capabilities was accomplished using an ohmic test device. Several devices were used and the measured charge collected from a pulse was consistent with the DC properties of the DUT. For a resistor of $24.0M\Omega$ with a $400pF$ polypropylene capacitor in parallel, we determined a resistance

that differed by $\sim 0.3\%$. This calibrator was stable with time. There was $\sim 0.5 \pm 0.1 \text{ pA}$ of leakage current discharging the integrator capacitor. We were able to measure this source of error and account for it with software. Therefore we were able to reliably measure currents from $1\text{-}2 \text{ pA}$ up to $\sim 10 \text{ mA}$.

However, when measuring the tunneling currents for short pulses driving large current densities, we encountered an anomaly in our measurement technique. Under some conditions, the charge passed for a given V_{app} is not linearly dependent on the width of the pulse. Figure 4 illustrates a worst case scenario displaying the temporal dependence of the collected tunneling currents. In curve (a), we plot the amount of charge passed, Q_{app} , per pulse as a function of the width of the pulse. The differential of this data, curve (b), shows that the slope of the Q_{app} vs t_{app} curve doesn't stabilize before the integrator capacitor starts to saturate. This could be a serious problem, since our double pulse technique assumes a linear dependence of Q_{app} on t_{app} . However, this phenomena is only significant under large current densities, $> 1 \text{ A/cm}^2$. To measure the currents into the high field, short pulse regime, it is required that the offset pulse width, t_{off} be only slightly shorter than t_{app} . Although this increases the accumulation of injected charge, it minimizes the errors induced from this non-linearity.

The charge on the integrating capacitor as a function of time is displayed in the oscilloscope plot of figure 5. This pulse of $V_{\text{app}} = 8 \text{ V}$ is a good illustration of the time dependence of the FN current $I_{\text{FN}}(t)$. At the rising/falling edge of the pulse we observe the expected displacement currents from the charging/discharging of the capacitance of the DUT. Although the transients are reproducible, they are not symmetric. Therefore,

without the use of a double pulse acquisition, the measured charge would be in error. Under quiescent conditions, the slope of the trace between the transients would be constant. However, we measure a monotonically increasing slope for the high field pulses, showing directly the dynamic nature of the sample. A plausible explanation for this phenomena is local heating of the DUT being stressed by high current densities for short times. After a longer period of time under stress, the local temperature should reach an equilibrium value and the current would be independent of time, so long as the same field was being dropped across the oxide.

Software Description. The hardware was controlled via a personal computer, and data acquisition and analysis algorithms were written in the ASYST programming language. Because of the unsymetric nature of the transients resulting from the application and removal of the applied pulse, it was necessary to use a double pulse technique for accurate collection of the $I(V)$ characteristics. The nature of this technique assumes that the current passing through the sample reaches a steady state. Consequently, an additional increase in the width of a subsequent pulse should result in a linear increase in the amount of charge collected by the integrator. Two pulses of equal magnitude, V_{app} , are used. The initial pulse, referred to as the offset pulse P_{off} with width, t_{off} will collect an offset charge, Q_{off} . The charge, Q_{off} is subtracted from the charge collected from the second and longer pulse, P_{app} , and the width is subtracted from t_{app} , so as to yield the steady state current.

A flow chart of the data acquisition algorithm is shown in figure 6. To determine the width of the offset pulse we estimate the expected FN currents from a preliminary

four point $I(V)$ curve. After V_{app} is set by the DAC, the estimated t_{off} is used to generate P_{off} such that the amount of charge passed is as small as possible, to limit the accumulated fluence, but large enough to include a portion of steady state current. A correction term is derived to compensate for inaccuracies in the theoretical $I(V)$ fit. The pulse width is incremented to a value such that Q_{app} is sufficiently larger than Q_{off} keeping in mind our desire to limit the amount of injected charge but also constrained by the signal to noise ratio. These pulses are subtracted, and corrected for the previously determined leakage currents. After data storage, the applied bias is incremented and the algorithm is looped through the rest of the data acquisition.

Results

Typical results for the samples we measured are shown in figure 7a-c. The $I(V)$ curve for a 50Å MOS capacitor is shown, in 7a, to exhibit the expected FN type conduction. The $I(V)$ characteristics are reproducible on several gates across the wafer. For samples that weren't damaged by hot electrons, the data is reproducible for several scans on the same gate, as shown in figure 7c. Because this is an integrating technique the signal to noise ratio is high, which allows us to discern small changes in the oscillations. Figure 7b. shows the data plotted using FN coordinates. By rearranging equation (1), we can plot the data as a linear relationship

$$\ln\left[\frac{J_{FN}}{F_{ox}^2}\right] = A' - \frac{C}{F_{ox}} \quad (2)$$

of the field across the oxide. Where the oxide field is determined iteratively from

$$F_{ox} = \frac{(V_a + \Delta\phi_{ms} - \alpha F_{ox}^{2/3})}{d_{ox}} \quad (3)$$

The applied voltage V_a and work function difference $\Delta\phi_{ms}$ are known, and the potential drop across the accumulation layer is determined from the band bending term⁶. The slope and intercept from Eqn. 2, can yield the metal work function potential and the effective mass of the electron for this system⁷. For thicker films the data would exhibit a straight line. But, for these thin films where the interference pattern of the electrons is superimposed on the FN characteristic, we see oscillations in the FN plot. We can subtract off the FN current from our data by determining the best fit linear relationship of equation (2). For the best results we perform a least squares fit to the data in a region of high electric field where the deviations from FN transport are small. Typically we place a datum near the inflection points of the oscillations, encompassing a few of the extrema, and fit this subsection of data. These oscillations are displayed in figure 7c.

The effects of hot electron injection into these devices is apparent in the oscillations of figure 7c. We show that as the amount of injected charge under high bias increases, so does the amount of excess tunneling current in the low field. This deviation

from the FN current has been observed by others⁸. This enhancement of tunneling currents has been explained by a local potential barrier lowering due to positive charge generation in the oxide⁹. The positive charge is possibly generated from the interactions of energetic electrons with defects in the oxide. Because we wanted to study the interfacial properties of these films, we tried to avoid this type of perturbation from the measurement. At low injection fluence, we can minimize these effects.

In order to examine the functionality of the pulsed acquisition board, we compared the results to those obtained by a "DC" ramped method. A linear voltage ramp was used to drive a bipolar operational amplifier power supply and this bias was applied to the substrate similar to the pulsed technique. The currents were sensed at the gate by a logarithmic picoammeter. The data was collected via a 12 bit sample and hold, successive approximation analog-to-digital acquisition board. The pulsed method has a superior signal to noise ratio. Although acquisition time was longer, the pulsed method has a higher sensitivity, and could more accurately measure the currents at low bias. The ramped method was limited in the low bias regime because of displacement currents that were typically an order of magnitude larger than the tunneling currents measured. Subtraction of the displacement currents for the ramped technique resulted in very noisy data at low field, in contrast to the pulsed technique. Most importantly, the amount of injected charge per scan was significantly reduced for the pulsed method. Even for the fastest ramp rates ($\sim 1\text{V/s}$), the amount of charge injected is typically 2-3 orders of magnitude larger for the ramped mode versus the pulsed.

In figures 8a and b we compare the $I(V)$ curves and oscillation results for the

pulsed and ramped techniques. The data was taken on the same sample but different gates, usually in close proximity to each other. It is shown, that for the ramped measurements, the $I(V)$ and the positions of the oscillations are shifted toward a more positive substrate bias. Since the applied voltages from both measurements were confirmed to be accurate, there seems to be something about the device under test that is lowering the apparent field across the oxide. From the analysis of the ratio of the $I(V)$ data, there seems to be an offset in the potential drop across the oxide. A constant offset of about 0.2 V would explain the qualitative trends of the $I(V)$ curves. From the oscillation data in figure 8b, we show that the positions of the extrema for the ramped experiment are shifted by approximately $+0.18V_{\text{sub}}$. This offset is fairly constant except in the very high bias regime where the voltage shift is negligible.

A likely explanation for the generation of the offset potential is that the application of a bias to the sample changes the occupation state of traps in the bulk or at the interfaces. If this effect was due to negative charge at the collecting interface of the device, $\sim 6 \cdot 10^{11} \text{e/cm}^2$ would be required to produce the observed shift. It is observed that this voltage shift must be established early on in the ramped $I(V)$ scan, since even the first minima in the oscillations is shifted. This effect does not seem to be a simple charge injection phenomena since the pulsed measurement actually passes more charge through the sample than does the ramped $I(V)$ in the low current density portion of the scan. It is also observed that the shift of the oscillations is independent of the ramp rate for the "DC" method. Different ramp rates will inject different amounts of charge into the oxides. Once the bias ramp is applied, it increases monotonically until the scan is

complete. However, in the pulsed acquisition, there is a short (~50ms) delay between pulses. During this hiatus the sample is held at ground, providing an opportunity for the trapped charge to emit, thus limiting the amount of accumulated charging.

The kinetics of the trapping and emission process can be accessed with a pump-probe study of the oscillations. A charging pulse, the pump, of 3V is applied for time t_p , then the traps are allowed to change occupancy, with the sample shorted to ground, for delay time t_d . The height of the applied probe pulse V_{app} is scanned through the region of an extrema in the oscillations. The width of the probe pulse, t_{app} , is determined by decreasing an initial pulse width, t_i , by 15% after each loop such that t_{app} will be the same near the maximum of one of the oscillations. Table I. lists the voltage shifts of the first maxima as a function of t_p , t_d and t_i . The positions of the maxima was shown to shift as a function of the pump pulse characteristics. For a given t_i , the voltage shift, ΔV_j , increased with increasing pump pulse width. There is also some dependence on the delay between pulses, where as the delay increases, ΔV_j decreases. Although the magnitude of the shifts in the pump-probe analysis isn't as large as those from the ramped $I(V)$, they do demonstrate that a non-trivial trapping-emission process is a valid interpretation of the measurement.

Summary

We have constructed a charge integrating pulsed $I(V)$ technique with several novel components and design considerations for speed and accuracy. The hardware was

analyzed and shown to yield adequate pulse application and charge collection even under the rigorous requirements of large device capacitance and high charge injection fluence. A double pulse algorithm was devised to accommodate the inherent problems of displacement transients and anomalies upon application and removal of the pulse. Accurate and precise $I(V)$ curves were measured for FN tunneling through thin MOS structures. A voltage shift in the positions of the quantum oscillations referenced to that of a "DC" ramped acquisition technique demonstrated an historical dependence of the charge passed for a given pulse. The pulsing and charge collection abilities of our acquisition board enabled us to observe the dynamic nature of the MOS devices, and helped toward elucidating the nature of this apparent trapping phenomena and an understanding of the capture and emission kinetics. Following publications will comprise the details of the experiments done using these techniques.

Acknowledgements

This research was supported in part by the Office of Naval Research, ONR. We gratefully acknowledge the Semiconductor Research Corporation, SRC, for a fellowship to one of us (JCP). A special thanks to D. Buchanan and D.J. Dimaria, of the IBM Corporation, for some sample preparation, and helpful discussions.

References

1. R.H. Fowler and L.Nordheim, Roy. Soc. Proc. A119, 173 (1928).

2. K.H. Gundlach, *Solid State Electronics* **9**, 949 (1966).
3. J. Maserjian, *J. Vac. Sci. Technol.* **11**, 996 (1974).
4. M.V. Fischetti, D.J. DiMaria, L. Dori, J. Batey, E. Tierney, and J. Stasiak, *Physical Review B* **35**, 4404 (1987).
5. W. Kern and D.A. Puotinen, *RCA Rev.* **31**, 187 (1970).
6. J. Maserjian, G. Petersson and C. Svensson, *Solid State Electronics* **17**, 335 (1974).
7. J. Maserjian, *The Physics and Chemistry of SiO₂ and the Si-SiO₂ interface*, edited by C. Robert Helms and Bruce E. Deal (Plenum Press New York, 1988) pp. 497-507.
8. J. Maserjian and N. Zamani, *J. Appl. Phys.* **53**, 559 (1982).
9. Fred W. Schmidlin, *J. Appl. Phys.* **37**, 2823 (1966).

Figures and Table Captions:

Table I. Summary of the voltage shifts of a quantum oscillation maximum resulting from the charging by a pump pulse. The width of the pump pulse, t_p , and the delay time between pump and probe pulses, t_d , are varied. The experiment was performed using the initial width of the probe pulse, t_i , as a parameter.

Figure 1. Band diagram of the M/SiO₂/Si tunneling structure. Electrons injected at the metal, tunnel into the conduction band of the insulator and propagate ballistically toward the semiconductor. The electron waves can undergo quantum interference in the electronic resonant cavity.

Figure 2. Schematic description of a novel charge integrating pulsed I(V) acquisition board.

Figure 3. Pulse shape characterization of measurement while loaded with a DUT having ~1300pF of capacitance and sinking ~400 μ A. The pulse shape at the sample is well defined and accurate (a), and the resulting signal at the integrator summing point exhibits small charging transients at the rising and falling edge of the applied pulse (b).

cont.

Figure 4. Temporal dependence of tunneling currents for large current densities and short pulse widths. Integrated charge per pulse versus pulse width, t_{app} , (a), and the differential (b) displaying the non-steady state behavior of the tunneling current.

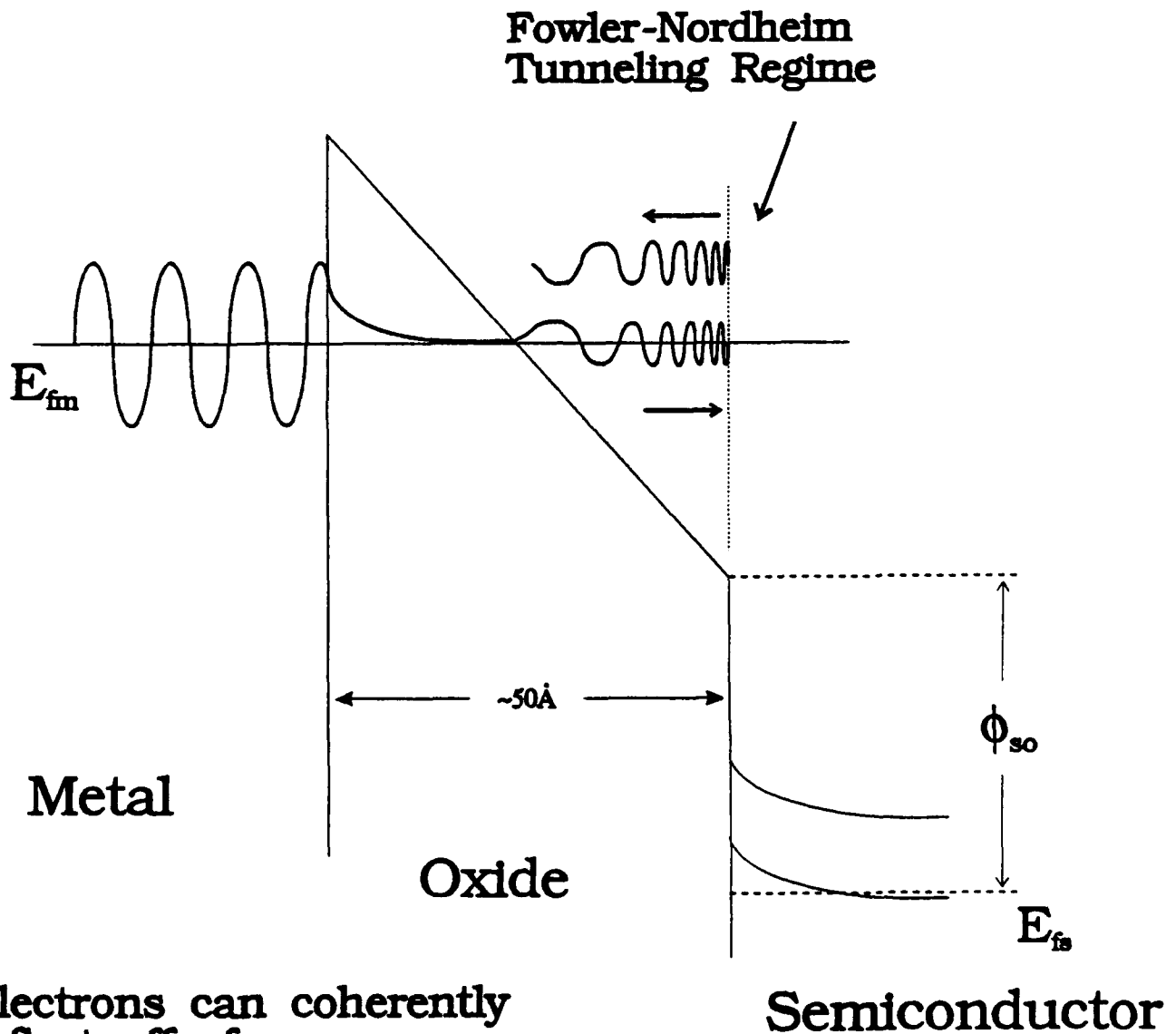
Figure 5. Signal across the integrating capacitor (Fig. 2. III), measuring the charge collected as a function of time for a $3.0\mu s$, $8.0V$ pulse. After charging of displacement currents from the rising edge of the pulse, the slope continues to change due to the time dependence of the current $I(t)$.

Figure 6. Flow chart of the data acquisition algorithm.

Figure 7. Typical results of the tunneling currents through a 50\AA oxide obtained with the pulsed acquisition method. $I(V)$ characteristics (a), tunneling currents plotted using FN coordinates (b), and the quantum oscillations in the FN current (c) are displayed. Fresh dot (solid), after injecting $7.2 \times 10^{15} \text{ e/cm}^2$ (dotted) and $17.9 \times 10^{15} \text{ e/cm}^2$ (dashed), samples are compared.

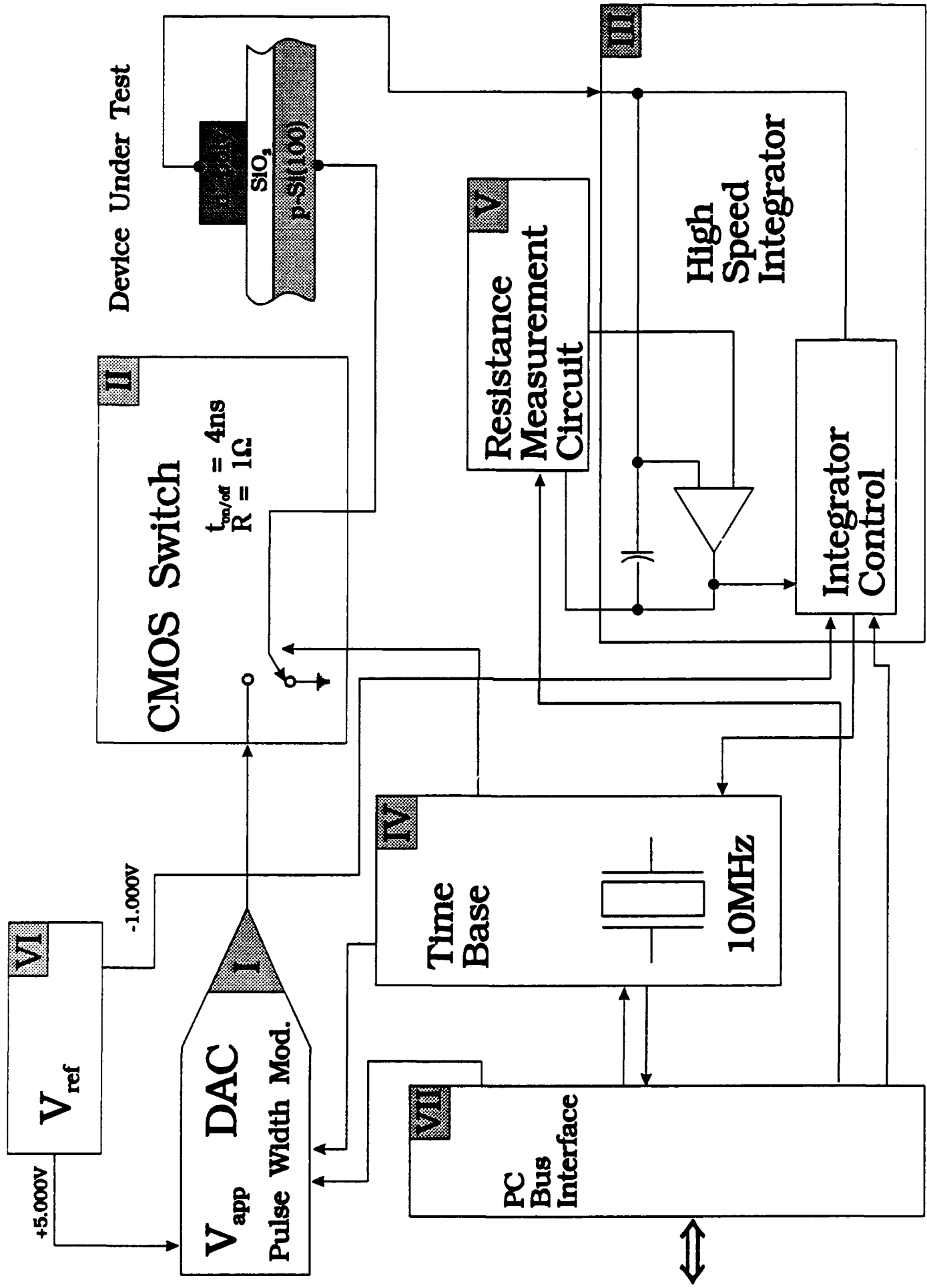
Figure 8. Typical results of tunneling currents comparing the pulsed (dashed) versus the ramped (solid) acquisition methods. The $I(V)$ characteristics (a) and quantum oscillations (b) are shifted by a constant offset of $-0.2V$.

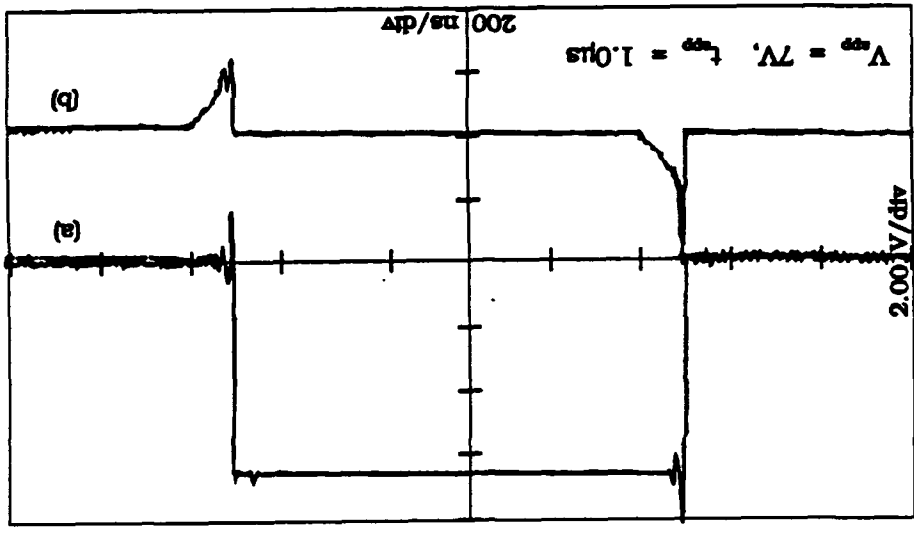
Band diagram of M/SiO₂/Si tunneling structure

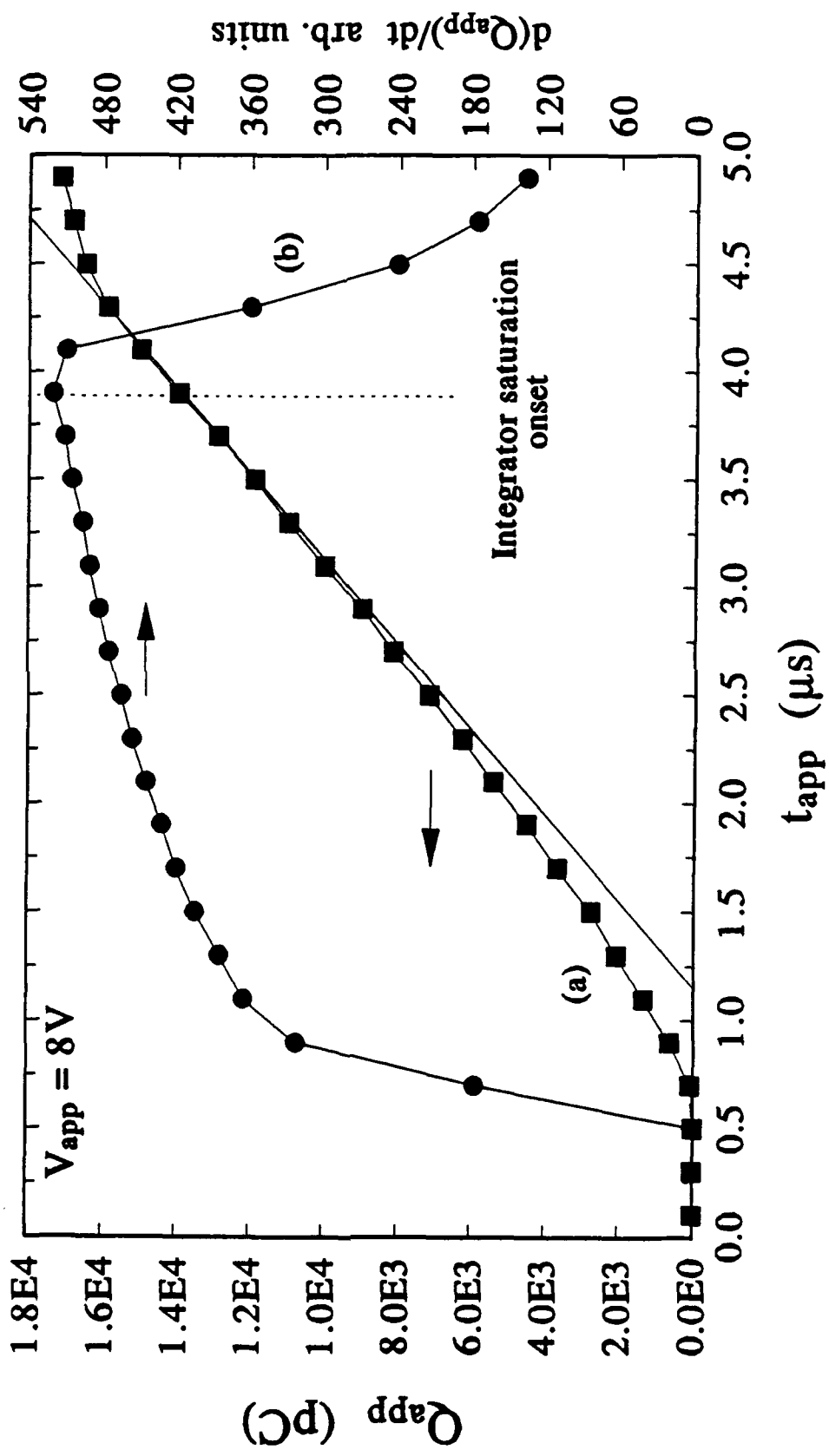


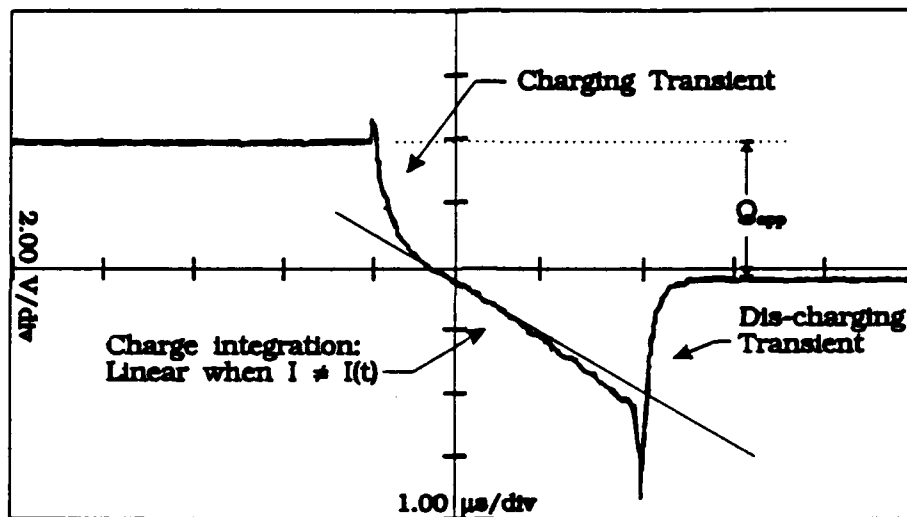
Electrons can coherently reflect off of Si/SiO₂ Interface.

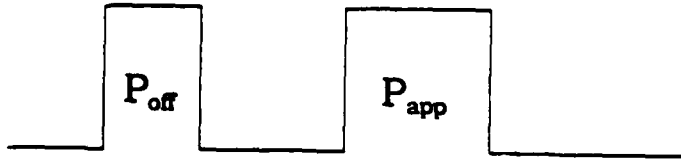
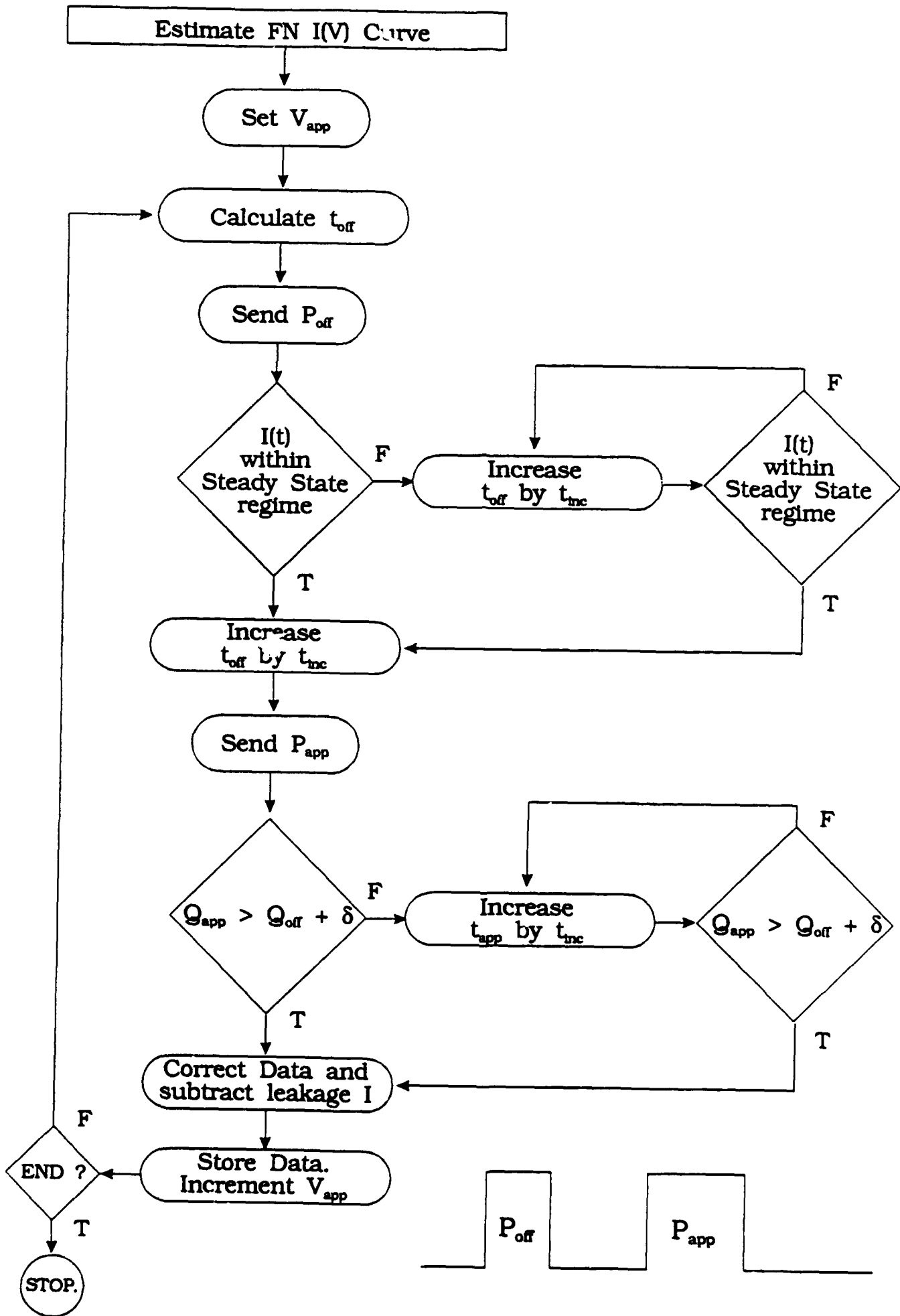
Semiconductor

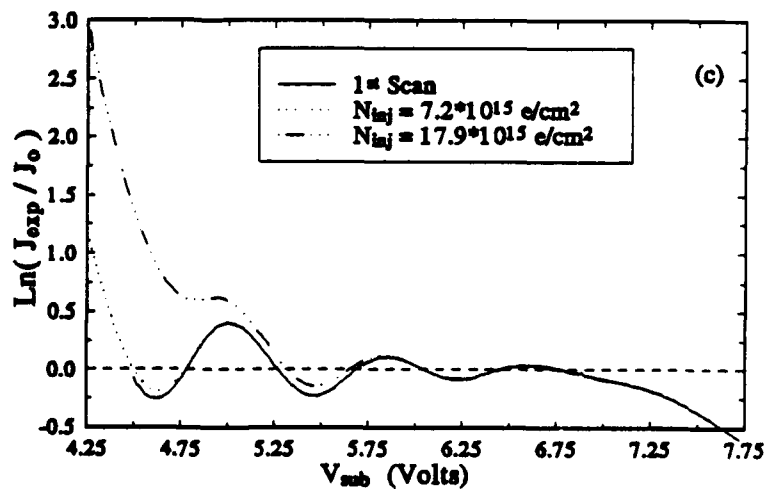
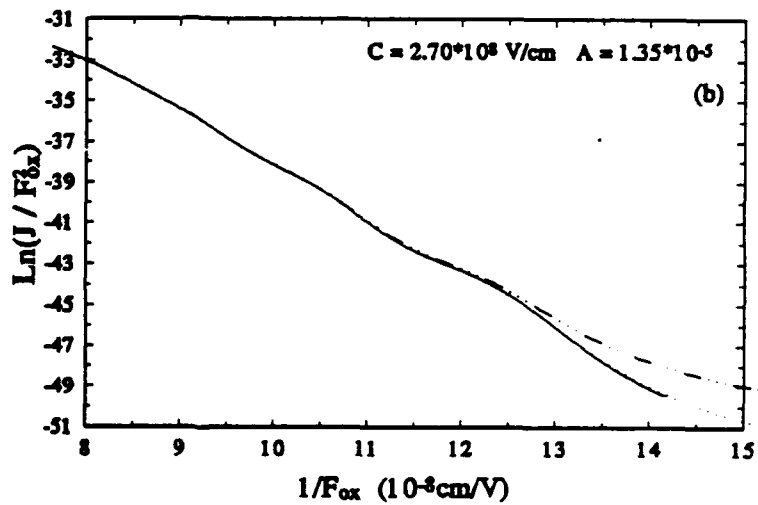
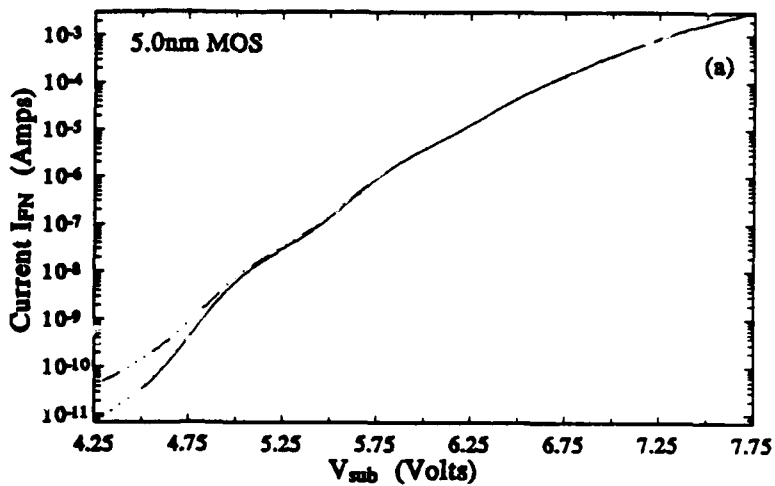


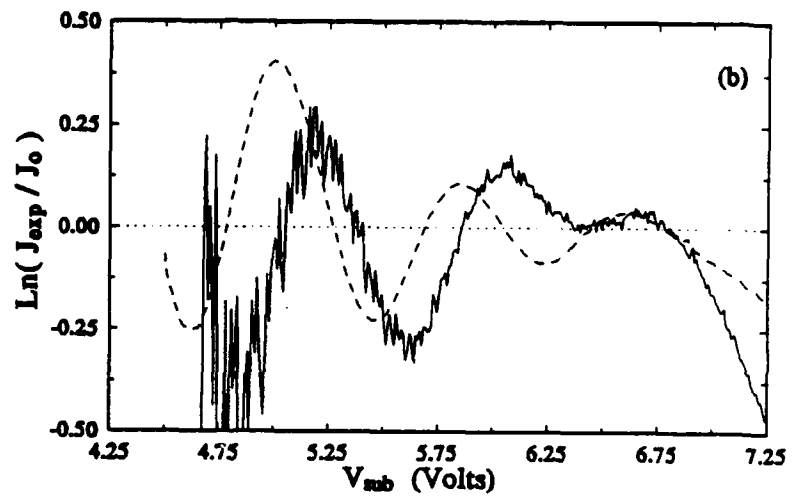
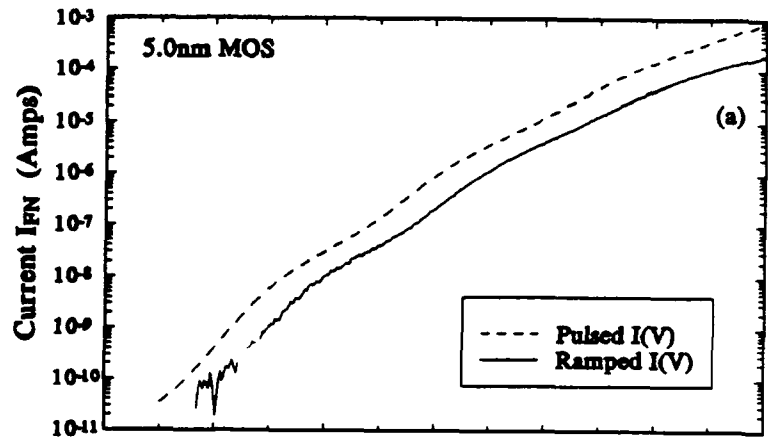












t_i (ms)	t_p (ms)	t_d (ms)	ΔV_i (mV)
300	1	50	0
300	3000	10	18
300	3000	500	18
300	5000	-0	30
500	1	10	20
500	1	10	20
500	10000	100	25
500	10000	10	27

t_i (ms)	t_p (ms)	t_d (ms)	ΔV_i (mV)
500	10000	10	27
500	10000	100	25
500	1	10	20
500	1	10	20
300	5000	-0	30
300	3000	500	18
300	3000	10	18
300	50	1	0