LINK TO TABLE OF CONTENTS

Charge-Coupled Analog Computer Elements and Their Application to Smart Image Sensors

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Eric Roy Fossum

May 1984

ABSTRACT

CHARGE-COUPLED ANALOG COMPUTER ELEMENTS AND THEIR APPLICATION TO SMART IMAGE SENSORS

Eric Roy Fossum

Yale University

1984

Real time machine vision in mobile robots requires pre-processing of images at speeds well in excess of one billion operations per second, depending on the resolution of the image plane. To obtain such high speed in a compact, light-weight, and low-power computing system, alternatives to the standard serial digital processor are currently being explored. The spatially parallel architecture, in which the inter-processor communication structure reflects the topology of the focal plane mosaic of elemental sensors, is organized in a way natural for image pre-processing. The most logical place for the processor array is the imaging focal plane. Such positioning avoids the serial encoding and transmission of the inherently parallel input data, but the available real-estate places a premium on processor simplicity and innovation.

In this dissertation, an analog processor based on the manipulation of discrete charge packets in a semiconductor is advocated. Such a processor shows promise for high density focal plane computing (more than 4000 processors/cm^2). The work focuses on the basic building blocks of the charge-coupled computer, a charge packet differencer/replicator and a charge packet magnitude comparator. The former is implemented in a novel circuit in an inherently linear and compact way through the use of three-dimensional charge coupling. The prototype

circuit was fabricated and measured at Yale, and was shown to operate properly in a manner consistent with its analysis.

Also investigated in the course of this research was the bistable metal/tunnel-oxide/semiconductor (MTOS) junction. The application of this device as a charge packet magnitude comparator was explored by utilizing its hot electron impact ionization internal positive feedback mechanism. It was found that the physical computation time of this device is somewhat longer than was desired in the charge-coupled computer, though it actually performs well as a charge-packet threshold detector.

The control circuitry integrated with the MTOS junction provided a unique opportunity to investigate the internal charging and discharging currents of the thin oxide capacitor. Using a novel charge packet injection technique, the dynamic response of the junction was studied and a measure of the oxide hole transport current and hot tunneling electron induced impact ionization current was made as a function of oxide voltage. The MTOS junction can also be controlled in a steady state manner by utilizing a diode-controlled MOS inversion layer to pin the MTOS junction surface potential. A new voltage-controlled N-type negative resistance was discovered in the course of making basic device measurements. The mechanism of this negative resistance in the charge-coupled MTOS-PN junctions is explained.

The dissertation spans a spectrum from machine vision, through solid-state circuit design to semiconductor device physics. Parochial progress, such as the fabrication of sub-micron interelectrode gaps, suppression of dark current, and the fabrication of a 33 Å gate oxide enhancement mode MOSFET is also reported.

PREFACE
AND
ACKNOWLEDGMENTS

Machines and intelligence are mutually exclusive concepts, at least philosophically. Yet, no one can deny that a machine can be made to appear to be intelligent. Witness the rise of elaborate computer programs which appear to convert a bank of switches into a thinking machine. Indeed, as human technology grows, the definition of human intelligence becomes more elusive. Perhaps there is no such thing as human intelligence in the sense that it is presently and emotionally defined.

It is in this context that 'smart' sensors are pursued. Such sensors can be viewed in terms of an incremental step toward machine intelligence, or as a solution to a practical problem. (Maybe human 'smartness' is also a solution to some Darwinian practical problem.) The norm for approaching the smart sensor problem, which can be considered as reducing the entropy of sensed information, is to adapt a bank of switches for the specific information processing task. Unfortunately, few researchers consider the issue of whether the digital approach is truly optimal for the job. Machine vision is a good example of this.

At the risk of alienating the entire computer science community in one stroke, the following observation is put forth. The problem of entropy reduction of information can be mathematically expressed. It is rarely formulated in Boolean algebra. Yet, the engineering problem of implementing the formulation seems to be routinely solved in a digital

state space. The reasons are historically clear (reduction of total computational error) but philosophically puzzling. Computer scientists take this Boolean algebra based machine, and operate it with a human-esque communication language. It would seem more reasonable to first formulate the problem in a more natural mathematics (e.g. predicate calculus) and second, perform the engineering implementation in a way which more closely resembles the abstract information manipulation.

The thesis presented here pales in comparison to the aspirations of the above paragraph. Perhaps some small step to smart robot vision sensing has been achieved, but the thesis is really a discussion of solid-state circuitry design, fabrication, and measurement. To a lesser extent, it is also a study of solid-state physical phenomena. It has been an interesting investigation and has served well as an educational experience.

The investigation succeeded in converting ideas to practice (the essence of engineering) because of the support of many people. Some support was realized indirectly, such as that provided by Becky Friedkin, my patient wife. Some support was realized directly, such as the proficient and expedient typing of Arlene Vasso, and the graphic arts expertise of Sal Datillo. I would like to thank all those who have enabled me to sit here and write a preface to a Ph.D. dissertation. In particular, I would like to acknowledge:

My parents, who tended a garden with no visible sign of bearing fruit.

An anonymous public school official who in 1969 'pulled my name from a hat' to participate in a special program on Saturdays.

The staff at the Talcott Mtn. Science Center who gave me my first taste of self education, especially Bill Danielson, George Atamian, and Donald LaSalle.

Sandy Stieber, who provided untold inspiration and a professional job.

Charles Miller, Harvey Picker and David Ahlgren and other faculty members at Trinity College who inspired me to finally 'try'.

Richard Barker and Tso-Ping Ma who provided for my professional growth at Yale.

Hans Maurer, Tony Funari, and Colin Whitney who provided for my professional growth at Hughes.

Members of the technical staff and support groups at Hughes who fabricated a test box, provided chip carriers, bonded the devices, photographed the chip and a hundred other things.

Members of the IC Facility at IBM Kingston who in spirit and deed tried to make the collaborative effort work.

My fellow graduate students at Yale who unselfishly gave their time so that I could go to California with chips to test, and particularly those of the Ma Barker gang whom I will always think of as family.

This work was also supported by a Yale University Fellowship (1979/1980) an IBM Fellowhip (1980/1981), a Howard Hughes Ph.D. Fellowship (1981-1984) and a contract with the Hughes Aircraft Company's Missile Systems Group (1982/1983).

Without the assistance of the individuals and institutions listed above, it is a certainty that this dissertation would not exist. To those above and those which I have erroneously omitted, I would also like to extend my sincere gratitude.

Finally, I would like to dedicate this dissertation to the person who pulls a dusty copy off a library shelf, or who views a yellowed micro-fiche, and wonders why someone might spend five years of their life to work on a trivial problem such as robot vision. After all, even children understand the basics of robotics...

TABLE OF CONTENTS

Cartoon at End

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The research described in this thesis concerns the investigation of charge-coupled computer elements for use in discrete time base analog signal processing. The thrust of this work is three fold. First, a charge-coupled computer in an array format is proposed for use in focal plane image processing. Second, the basic building block of the proposed charge-coupled computer, a charge packet differencer/replicator is realized in a novel circuit. Third, the metal/tunnel-oxide/semiconductor (MTOS) bistable junction is used as a charge packet magnitude comparator through the utilization of its thresholding capabilities. The ultimate goal of this line of research is to demonstrate the feasibility of spatially parallel processing on the focal plane. The work reported in this thesis represents a step forward in this direction by proposing and implementing methods of charge-coupled computing which will result in the realization of such a smart vision sensor.

In chapter two, the rationale for choosing a charge-coupled computer approach for focal plane image processing is developed. Real time mobile robot vision requires light-weight, low-power, but enormously high throughout computation. Parallel processing is the adopted method of providing the high throughput but only recently has the light-weight, low-power issue been seriously addressed. The appropriate architecture for the parallel processor is a spatially parallel one [Schaefer, 1982; Schaefer and Strong, 1977; Schaefer and Fisher, 1982]. In this architecture, the image plane topology is reflected in the inter-processor communication structure either logically as in the Massively Parallel Processor (MPP) [Batcher, 1980] or physically, as in the 3D

wafer scale integration of a parallel processor [Grinberg, et al. 1984].

In this work, a spatially parallel architecture based on a discrete time base analog processor is proposed for die size integration. The analog approach provides for significant circuit compaction, yet retains sufficient accuracy for complex image processing. The processor is based on a charge-coupled device building block which performs charge packet differencing, attenuation, or replication in a compact and linear way [Fossum, 1981; Fossum, 1982]. A block diagram and tentative floor plan of the processor is presented in chapter two. A program which causes the single-instruction/multiple-data parallel processor to perform non-uniformity compensation, smoothing, edge detection and thresholding on an image is presented to illustrate the use of this charge-coupled computer concept. A one micron ground rule PMOS implementation of the processor is expected to execute the above program in approximately 58 μsec, which for 100 Hz frame rate imagery leaves 942 μsec for higher order image processing functions. The analog processor can also be programmed to perform A/D conversion of the pre-processed image.

Chapter two concludes with a review of charge-coupled device physics beginning with the deep depleted MOS capacitor. Signal charge input, charge transfer, and signal charge output are discussed. The physics of such devices constrains the charge-coupled computer architecture, yet at the same time permits circuit compaction.

In chapter three, the fabrication process used for manufacturing charge-coupled devices at Yale is presented. Fabricating charge-coupled devices integrated with tunnel oxide devices is a challenge in any state-of-the-art facility and particularly at Yale. An attempt was made to fabricate the devices in a collaborative effort with IBM in Kingston,

New York. Successful bistable tunnel oxide devices were manufactured in a test run and mask plates for prototype charge-coupled computing devices were fabricated. The collaborative effort ended abruptly when the Kingston IC facility was dissolved in a corporate and physical sense in 1982.

In this work, a CCD technology merged with a previously established tunnel oxide technology [Dressendorfer, 1978; Lai, 1979; Dressendorfer, et al. 1980; Teng, 1983] was successfully developed at Yale which overcame two technological hurdles. The first was the fabrication of closely spaced MOS capacitor electrodes (sub-micron spacing) which was achieved through the use of a shadow evaporation/liftoff process [Browne and Perkins, 1976]. The second and more difficult hurdle was the suppression of dark current in the CCD potential wells. Through the slight adjustment of the basic Yale MOSFET process [Ma, 1984; Teng, 1983] the dark current was suppressed by several orders of magnitude. The details of this investigation and subsequent process are described in chapter three. The resultant eight level mask set is the most complex IC process undertaken at Yale to date.

Chapter four discusses the investigation of an MTOS junction based charge packet threshold detector for use as a charge packet magnitude comparator in the charge-coupled computer. The bistable behavior of the MTOS junction and its switching properties with respect to light intensity, gate voltage, and minority carrier injected diffusion current have been investigated by Lai (1979) and Teng (1983) as reported by Hayashi (1980) and Lai, et al. (1981). In this work, the response of the junction to an injected packet of minority carrier charge was addressed. Although the response time of the Al/SiO_2/Si(n) junction appears to be too slow for application in the charge-coupled computer, the integrated circuitry provided a unique opportunity to study the non-

equilibrium behavior of the bistable junction in both a charge controlled manner, and in a surface potential controlled manner [Fossum and Barker 1984a and 1984b]. This investigation, as reported in chapter four, has resulted in the measurement of the non-steady-state internal currents flowing in the junction near the surface arising from impact ionization and oxide hole transport, and has yielded techniques which are applicable in the study of similar device structures constructed of different materials.

In chapter five, the charge-coupled computing investigation continues with the theory, fabrication and operational verification of a prototype circuit which performs charge packet differencing and replication. This novel circuit is analyzed theoretically for both ideal and non-ideal behavior with respect to linearity, gain and speed. The theoretical predictions are compared to the observed operating characteristics of the experimental prototype circuit fabricated and tested at Yale. The prototype circuit is found to work well, with performance properly predicted by the calculations. An improved circuit suitable for fabrication at Yale based on the observed performance of the prototype circuit is proposed and discussed.

Chapter six concludes this work with a summary of the achievements obtained in this research. Observations made in the course of this research which indicate possible future research spin-offs and directions, both major and minor, are highlighted.

In auxillary support of the concepts presented in the course of this thesis, five appendices are included. Appendix A is a list of symbols and terminology. Appendix B discusses a zero-current approximation for determining minority and majority carrier concentrations in non-equilibrium/non-steady-state MIS capacitors. Appendix C is a discussion

of hot tunneling electron induced impact ionization models. Appendix D discusses the operation of the source-follower floating diffusion charge packet output sense amplifier used in the prototype circuits. Finally, appendix E discusses the experimental set-up and apparatus used in making the measurements of the prototype circuits.

CHAPTER 2

CHARGE-COUPLED COMPUTING

2.1 Introduction to Focal Plane Image Processing

Real time vision in mobile robots represents a unique challenge for computer engineering. The system must be fast, light-weight, and low-power. Examples of systems requiring such real time vision include autonomous vehicles which might perform such varied tasks as underwater inspection, farm crop planting, airborne reconnaissance, and space exploration. Future systems may be more consumer oriented such as auto-piloting of automobiles, security watch -'dogs', and conventional popularized robots (R2-D2). In these systems, low cost will be an equally demanding requirement.

Any vision system can be partitioned into sub-systems which sequentially operate on output from lower order functions to perform higher order tasks. This digested data may be passed on to other sub-systems or fed back to improve lower order functioning. The total vision task may be generally divided into an image acquisition sub-task, an image pre-processing sub-task, a feature extraction sub-task, a cognitive sub-task, and finally an action sub-task.

The human eye acquires an image consisting of approximately 125 million picture elements (pixels). The picture elements may be considered as a mosaic 11,000 elements per side with a 3.7 micron pitch. In principle, the human vision acquisition system should be able to resolve two points an inch apart at a distance of 100 yards [Hecht and Zajac, 1974]. A robot eye should decompose an image into at least 1024

x 1024 pixels [Nitzan, 1980], though surprising performance can be obtained with a 32 x 32 array size. The rate at which images must be acquired depends upon the rate at which the robot's environment is changing. If the robot must react in a time T_R, then the image rate should be at least T_R^{-1}. Typical imaging frame rates range from 1 Hz to 100 Hz. In this work, a 100 Hz frame rate is assumed. Thus the image acquisition rate is 100 million picture elements per second. If each picture element is digitally encoded on an eight bit gray scale, serial transmission of the image data stream would require approximately a 1 GHz bandwidth.

The image pre-processing task follows image acquisition. This task includes operations such as smoothing (noise removal), level shifting (fixed pattern 'noise' removal), gain adjustment (sensor non-uniformity compensation or adaptive contrast enhancement), sharpening, edge enhancement, thresholding and region growing. These operations, which act on a particular picture element, require data which generally comes from a local neighborhood of picture elements surrounding the processed pixel. The number of local neighborhood operations is typically 1 to 100 [Danielsson and Levialdi, 1981] requiring 1 to 50 instructions each with accuracy of the order of 1% or seven digital bits. The image pre-processing therefore requires an instruction execution rate on the order of 50 billion instructions per second, or 20 psec/instruction cycle time. Such a processing rate is not anticipated in a serial computer in the near future.

Higher order image processing functions are more difficult to identify. Feature extraction and topology operations typically follow image pre-processing. This information is merged with stored or previously learned data in a cognitive fashion, and an action is decided upon. For example, consider a vehicle-watching task. The vehicle must

be identified in each frame and separated from background objects. If the vehicle moves, its movement relative to its previous location must be ascertained, and a decision of how to move the robot eye to track the vehicle must be reached. To make such high order functional decisions at 100 Hz, the image pre-processing functions should only occupy a small fraction of the frame interval; perhaps no more than 10%. Thus for a 10 msec robotic reaction time, the image pre-processing instruction rate actually needs to be 500 billion instructions per second.

The serial digital computer in its conventional form is ill-suited for such image pre-processing and a computer architecture for 'understanding' images must be employed [Nudd, 1980]. There have been two basic approaches to solve this problem. For low-cost constrained systems, specialized integrated circuits placed in the serial data stream can perform many pre-processing functions such as edge enhancement in a pipe-lined fashion [Nudd, 1977; Fouse, 1980; Nudd 1978; Sternberg, 1983]. The drawback to this approach is that serial transmission of the data is required and the data stream operator must be fast and accurate.

In the long run, parallel computation approaches appear to be a better approach [see the review articles by Danielsson and Levialdi, 1981; and Preston, 1983]. Cellular logic computers have been in development since the 1950's. In the early 1960's parallel processors such as SOLOMON [Slotnick, et al. 1962] and ILLIAC III [McCormick, 1963] were being experimentally constructed. The CLIP series was begun in the mid 1960's [Duff, et al. 1967]. The major problem with these early parallel processors was poor reliability due to the non-integrated circuit technology of the time [Preston, 1983].

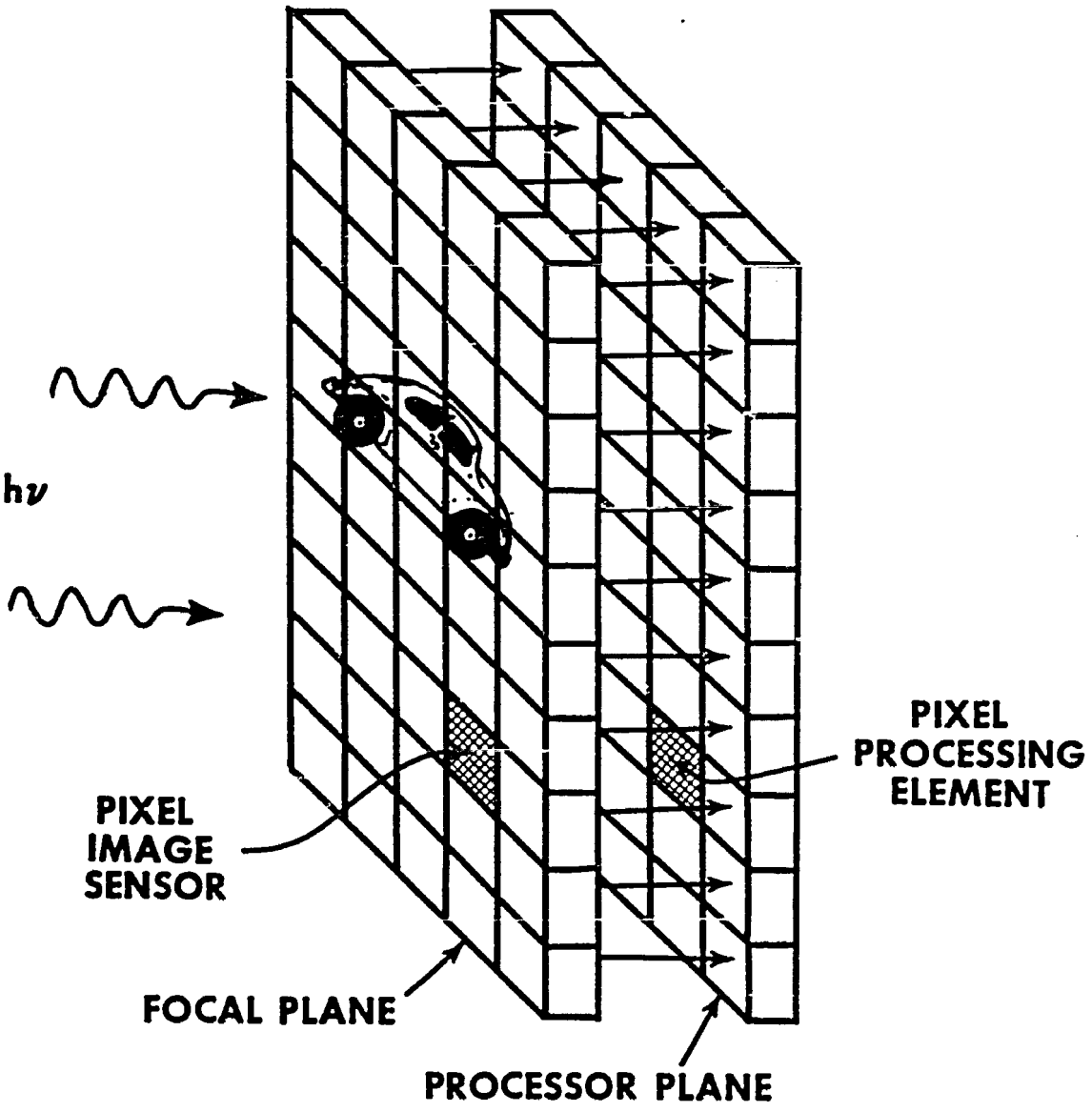
Recent large-scale integrated circuit technology has permitted the

proliferation of research activity in parallel processing to be reduced to practice. Examples of parallel processors for vision applications include;

CLIP IV	[Duff, 1979]
PHP	[Herron, et al. 1982]
GLOPR	[Preston, et al. 1979]
DAP	[Reddaway, 1979]
ZMOB	[Kushner, et al. 1982]
PICAP	[Antonsson, et al. 1982]
TOSPICS	[Kidode, et al.]
FLIP	[Luetjen, et al. 1980]
MPP	[Batcher, 1980]
3D	[Grinberg, et al. 1984]

The architectures of some of these processors are reviewed by Danielsson and Levialdi (1981) or Preston (1983). Some architectures are bus connected microprocessors (ZMOB) and others are three dimensionally integrated on a wafer scale (3D). Only the 3D shows promise for use in a mobile robot vision system since most of the processors occupy volumes measured in cubic meters.

Locating the parallel image pre-processing on, or just behind, the focal plane of the image acquisition systems relieves several problems associated with the above approaches [see Fig. 2.1-1]. First, an image data base acquired in parallel does not need to be converted to a serial data stream for transmission to parallel processor. Second, the problem of connecting a million pixel sensors to a million processors is considerably reduced if the connecting wires are indium bumps or microscopic springs [Grinberg, et al. 1984]. Such an architecture is truly spatially parallel, but requires that each processing element (PE) either occupy a miniscule amount of real-estate or be distributed through several parallel processing planes. In fact, both approaches are reasonable and are being pursued by Hughes, and by Honeywell. [Such information tends to be classified since focal plane image processing is of great interest in many defense projects, and is thereby difficult to obtain in an academic setting].



2.1-1 Schematic illustration of focal plane image processing.

In this work an array size of 32 x 32 PEs is considered for integration on a single processing plane half a centimeter on a side corresponding to a density of 4000 PEs/cm^2, or a pitch of 160 μm. This small amount of real estate is not large enough to permit the layout of a complex digital processor using present day design rules and even a simple digital processor (i.e. bit-slice) would require considerable real estate for the A/D conversion of the analog image sensor signal. In the present study, an analog processor based on discrete charge packet manipulation is proposed.

The charge-coupled computer is a discrete time base analog processor which uses the size of charge packets to represent signal magnitudes. Each charge packet is a continuous valued quantity with a minimum resolvable size of one part in 256 of the maximum charge packet size corresponding to an equivalent eight bit digital accuracy, or -48 dB resolution. Charge-coupled devices (CCDs) typically have dynamic ranges in the 60 dB to 70 dB range. Qualitatively, the reason why a charge-coupled computer can be built in less real estate than a digital processor is that a single digital bit is typically stored as a presence or absence of minority carriers in a semiconductor under an electrode, whereas under the same electrode, a CCD stores a continuum of charge. The noise rejection of the digital bit is much better than in a CCD, but an electrode is required for each bit of accuracy. If twelve bit digital equivalent accuracy was required of the processor, a digital approach would be needed since 0.025 percent accuracy is difficult to achieve in a CCD. But for 1% accuracy, the CCD is more than adequate. In this way, the CCD extends the two dimensionality of the processor into a third dimension of charge magnitude.

Charge-coupled devices have been used as read-out multiplexors for imaging devices almost since their inception [Amelio, et al. 1971].

More recently, the charge packet injected into the charge transfer device has been manipulated to effect simple image pre-processing functions such as non-uniformity compensation and local neighborhood convolutions [Nudd, et al. 1978; Fouse, et al. 1980]. The computation is often hard-wired for specific applications such as star tracking [Wong, et al.].

In the charge-coupled computer proposed in this work, the charge transfer function is not implemented, rather the coupling of charge packets for computation is stressed. The computer is fully programmable and is therefore considerably more flexible in its application, and better suited for more complex processing. It has recently surfaced that Honeywell has pursued a similar line of research and has developed a parallel information processor (PIP) chip [J. Joseph, private communication] based on the same concept, but implemented using different CCD circuits. Such coincident 'parallel' efforts serve to firmly establish the validity of the charge-coupled computer concept presented in this work.

2.2 Charge-Coupled Computer Architecture

The architecture of the charge-coupled computer is driven by real-estate constraints and image pre-processing algorithms. Image pre-processing is discussed in some detail by Rosenfeld and Kak (1976). In most applications we are concerned with the eight nearest neighbors in the rectangular grid image plane which surround the PE at location (i, j). Each neighbor is numbered in a clockwise fashion for convenience as follows:

Neighbor	Location	Neighbor	Location
ϕ_1	(i+1,j-1)	ϕ_5	(i-1,j+1)
ϕ_2	(i+1,j)	ϕ_6	(i-1,j)
ϕ_3	(i+1,j+1)	ϕ_7	(i-1,j-1)
ϕ_4	(i,j+1)	ϕ_8	(i,j-1)

The PE at location (i,j) is referred to as ϕ_0. Some image pre-processing operations are defined in Table 2.2-1. It can be seen from the Table that most of the operations can be executed using a simple arithmetic instruction set. In addition to the arithmetic functions, data management functions must also be implemented. These include both long and short term memory, data transfer between arithmetic operators and memory, and data transfer between nearest neighbors. Data must also be transferred 'normal' to the x-y processing plane, between the image sensor and the PE and between the PE and the system.

The charge-coupled computer proposed in this work can be largely built from two basic building blocks: a charge packet differencer which performs the basic operation $\lfloor \alpha A - B \rfloor$[1], and a MOSFET switch. The first building block is described in detail in chapter 5. It consists of several electrodes in which the device output is controlled through charge subtraction and integration on two of the electrodes. Each electrode acts as a summing node in which charge may be subtracted in multiple packets or continuously as a current. The node may be reset through a pre-charge cycle. The output of the circuit is determined by the difference in charge on the two electrodes during a single (or multiple) generation cycle which forms and stores the output charge packet in a CCD summing bucket. The gain term α is determined by the areal ratio of the two electrodes.

Two arithmetic operators are built using a unity gain balanced circuit. The first performs $\lfloor A - B \rfloor$, and the second performs $\lfloor D - E \rfloor$ only if a conditional switch has been previously triggered. This conditional switch is triggered subsequent to the magnitude comparison

---

[1] $\lfloor x \rfloor = x$ if $x > 0$, else $\lfloor x \rfloor = 0$

Table 2.2-1

Some Image Pre-Processing Operations

Name	Operation
Level Shifting	$\phi_o' = \phi_o - k_o$
Cross Correlation	$\phi_o' = \lvert \phi_o - m_o \rvert$
Gain Adjustment	$\phi_o' = \alpha_o \phi_o$
Thresholding	$\phi_o' = \ell_o$ if $\phi_o < \ell_T$ $= \ell_1$ if $\phi_o \geq \ell_T$
Smoothing	$\phi_o' = \frac{1}{16}[(\phi_1+\phi_3+\phi_5+\phi_7) + 2(\phi_2+\phi_4+\phi_6+\phi_8) + 4\phi_o]$
Sharpening	$\phi_o' = \lvert 5\phi_o - (\phi_2+\phi_4+\phi_6+\phi_8) \rvert$
Edge Detection	$\phi_o' = \lvert(\phi_1+\phi_2+\phi_3) - (\phi_5+\phi_6+\phi_7)\rvert + \lvert(\phi_3+\phi_4+\phi_5) - (\phi_7+\phi_8+\phi_1)\rvert$

of two quantities X and Y. The comparator might be implemented using a bistable MTOS device [Fossum and Barker, 1984a and 1984b] by injecting a charge packet of sufficient magnitude to initiate switching of the junction. Such a circuit investigated in chapter 4 is actually too slow to be useful in this application. A better method is to use an external positive feedback comparator circuit similar to that used by Wong, et al. or Hamilton, et al. (1980), which can resolve a difference of 10^4 charge carriers at 5 MHz.

A third arithmetic operator which generates an output charge packet equal to half the input charge packet is designed by leaving out one input and choosing the layout geometry such that $\alpha = 0.5$. Using this operator, charge packet attenuation may be achieved by single or multiple utilizations of this circuit.

Two identical short term memory (STM) registers are similarly made from this charge packet differencer circuit using only one of the two inputs and unity gain geometry. The signal charge is stored (or summed) on the electrode, and when the register's data is required, a generation cycle is initiated to form the output charge packet. The register is cleared using a pre-charge pulse. The capacity of the STM register can be enlarged by increasing its physical layout dimensions. It does not include the CCD output summing bucket.

Two long term memory (LTM) registers are based on a similar principle but use an MNOS technology to permanently store the electrode charge in a nitride layer. Inputs to the LTM electrodes are included to temporarily subtract from the permanent charge and facilitate level shifting.

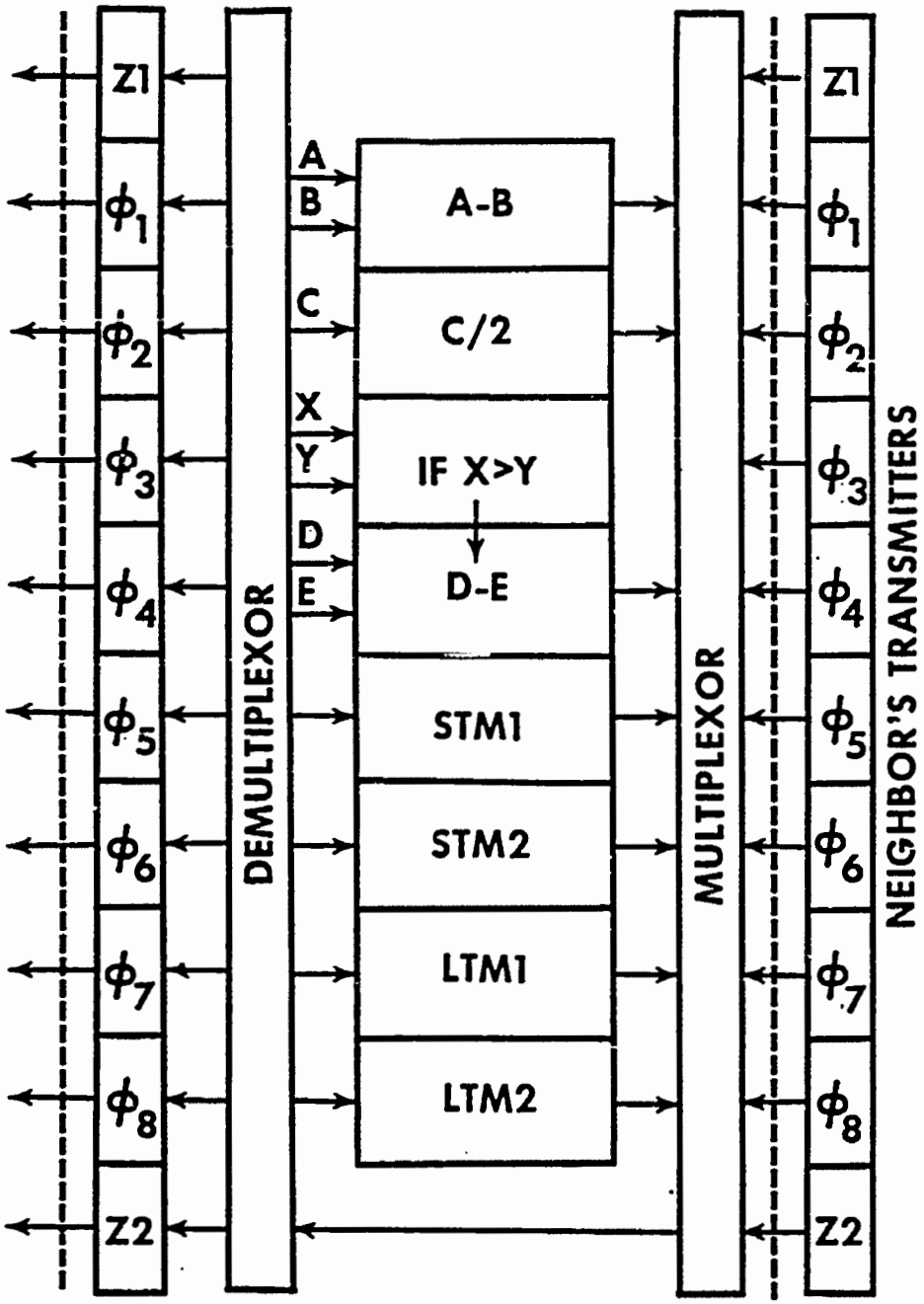
Transmission of data to neighbors takes place through eight buffer memory circuits, and two buffer memory circuits for Z-axis data

transmission. These eight buffer memory circuits are constructed in the same way as the STM registers, except their collective pre-charge and generation circuits are respectively tied in parallel. The neighbor reads out the transmitted charge packet following the generation cycle into one of its arithmetic or memory circuits. Such buffered transmission may appear cumbersome but has utility in its ability to easily re-generate the transmitted data.

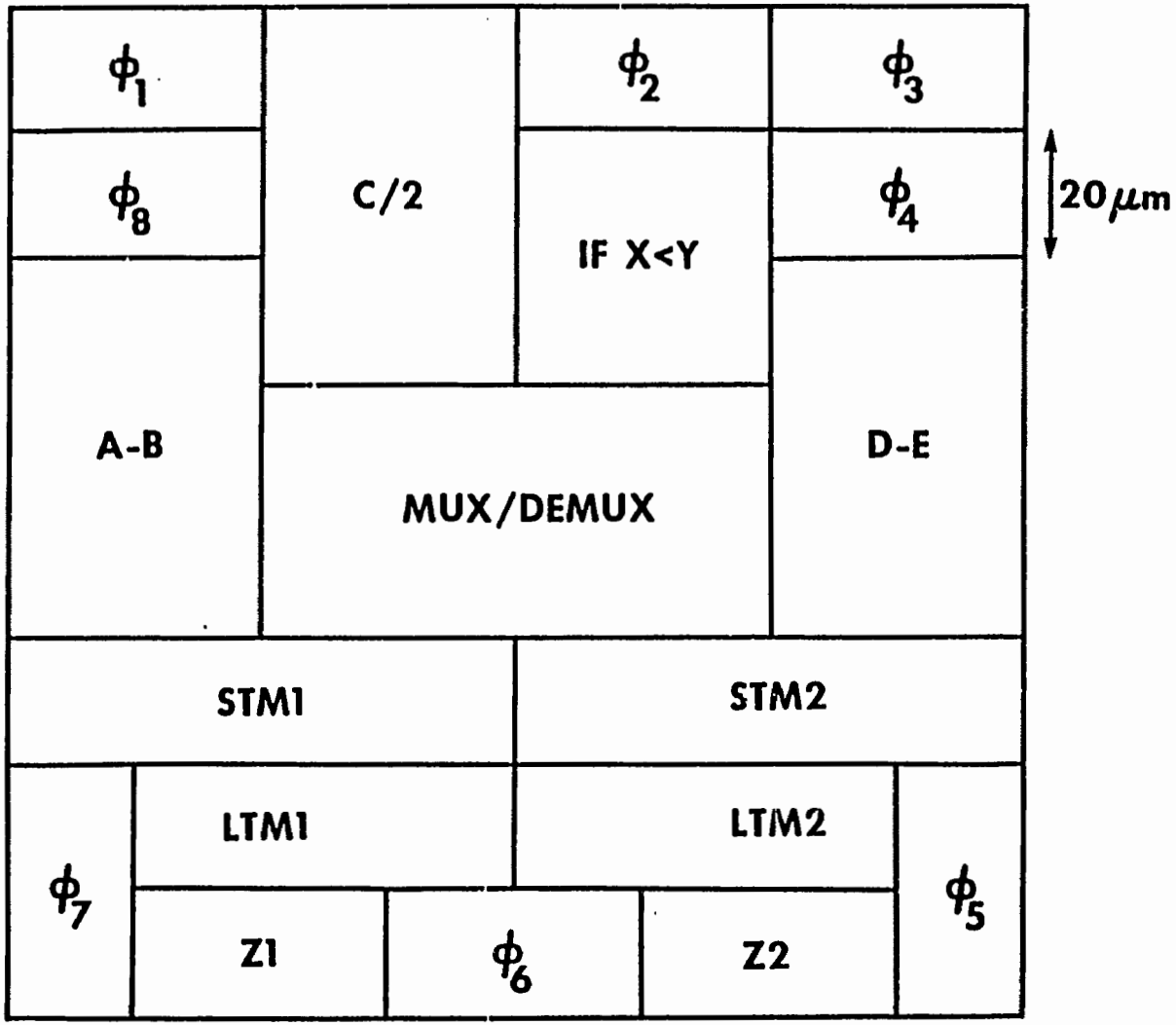
Routing of signal charge occurs through a standard MOSFET switched multiplexor/demultiplexor. This is one of the advantages offered by the charge packet differencer implementation approach taken in this work. Data, represented by minority carriers in a potential well, is converted to majority carriers through a collector diode and subtracted from a precharged electrode. The charge-coupling is thus truly three-dimensional in these circuits. The disadvantage is that wiring capacitance in the mux/demux can lead to non-ideal behavior of the charge packet differencer.

A charge packet multiplier is not implemented in this proposed charge-coupled computer because its utility is low compared to its real estate consumption. However, recent research indicates that a small multiplier may be feasible in future designs in a compact way [P. Nygaard, private communication].

A block diagram of the proposed charge-coupled computer is shown in Fig. 2.2-1 and a tentative floor plan is illustrated in Fig. 2.2-2. Implementation of the functional blocks would be similar to that shown in Fig. 5.3-3, with the primary differencer electrodes in the 8 μm x 40 μm size range, and with a CCD summing bucket appended to the output of arithmetic operators. The overall size including isolation channel stops might be in the 40 μm x 60 μm size range for the arithmetic



2.2-1 Block diagram of the charge-coupled computer.



2.2-2 Tentative floor plan for the charge-coupled computer.

functions and 20 μm x 40 μm size range for the transmission buffers.

As will be discussed in greater detail in ensuing sections, the charge packet differencer operation can be divided into three phases of equal duration: the precharge phase, a fill and spill phase, and an output transfer phase. Using an effective transfer length of 16 μm, the time required for each phase for 99.9% charge transfer efficiency (assuming $\mu_p \approx 150$ cm^2/V sec) is approximately 1 μsec [Lee and Heller, 1972], so that there is time for 1000 operating phases in the course of image pre-processing.

To exercise the charge-coupled computer's architecture, a sample pre-processing task is defined as follows: First, non-uniformity compensation by individualized level shifting is to be performed. Second, the image data is to be smoothed. Third, the edges in the image are to be enhanced through a gradient operation. Finally, the resultant edge intensities are to be compared to a threshold level. If the edge intensity is greater than that level, a system defined high level is to be output, otherwise zero is to be output. This task requires approximately 50 arithmetic operations and 20 I/O operations per pixel.

A sample program for the charge-coupled computer is shown in Table 2.2-2. The program requires 58 phases and at 1 μsec/phase would take 58 μsec to execute with 942 μsec to spare. From a throughput perspective, the charge-coupled computer will be more than adequate for image pre-processing. A close examination of the program reveals that advantage is taken of the architecture so that several arithmetic functions and I/O can occur simultaneously, further extending the degree of parallelism present in the system. This sample program can be written in several ways, some of which may result in slightly faster execution times. However, the sample program reveals that data shuffling consumes

Table 2.2-2

Sample Charge-Coupled Computer Program

INSTR	A-B	C/2	IF X>Y	D-E	Z1	Z2	ϕ_1-ϕ_8	STM1	STM2	LTM1	LTM2	MUX INPUT SELECT	DEMUX OUT SELECT	COMMENTS
1										PC				
2		PC			G					G				OFFSET VALVE IN LTM1
3		G										Z1+LTM1	C/2	
4		—												OFFSET DONE
5		PC												
6		G										C/2	C/2	
7	PC	—												$\phi_o/4$
8	G											C/2	A	
9	—	PC												
10		G										A-B	C/2	
11	G	—					PC							$\phi_o/8$
12	—	G										C/2	ϕ_1	
13	G	—										A-B	ϕ_2	
14	—	G										C/2	ϕ_3	
15	G	—										A-B	ϕ_4	
16	—	G										C/2	ϕ_5	
17	G	—										A-B	ϕ_6	1 2 1
18	—											C/2	ϕ_7	$\frac{1}{8}$ 2 2 ϕ_o
19	G	PC					G					A-B	ϕ_8	1 2 1
20	—											$\phi_1+\phi_2..+\phi_8$		
21	G													
22	—													$\phi_o/2$
23		G										A-B	C/2	
24		—												ϕ_o SMOOTH DONE
25		G												
26	PC	—												$2\phi_o$
27	G											C/2	A	$2\phi_o \rightarrow A$
28	—	G					PC							
29	G	—										A-B	ϕ_2	
30	—	G										C/2	ϕ_1	
31	G	—										A-B	ϕ_4	
32	—	G										C/2	ϕ_2	
33	G	—										A-B	ϕ_6	
34	—	G										C/2	ϕ_3	1 2 1
35		—										A-B	ϕ_8	2 2 ϕ_o
36	PC						G					C/2	ϕ_4	1 2 1

Table 2.2-2

(continued)

INSTR	A-B	C/2	IF X>Y	D-E	71	72	ϕ_1-ϕ_8	STM1	STM2	ATM1	ATM2	MUX INPUT SELECT	DEMUX OUT SELECT	COMMENTS
37												$\phi_1+\phi_2+\phi_3$	A	1 2 1
38	G											$\phi_5+\phi_6+\phi_7$	B	-1 -2 -1
39	|							PC						
40	PC						G					A-B	STM1	
41												$\phi_5+\phi_6+\phi_7$	A	-1 -2 -1
42	G											$\phi_1+\phi_2+\phi_3$	B	1 2 1
43	|													
44	PC						G					A-B	STM1	
45												$\phi_3+\phi_4+\phi_5$	A	-1 1
46	G											$\phi_7+\phi_8+\phi_1$	B	-2 2
47	|													-1 1
48	PC						G					A-B	STM1	
49												$\phi_7+\phi_8+\phi_1$	A	$\frac{1}{2}$ $-\frac{1}{2}$
50	G											$\phi_3+\phi_4+\phi_5$	B	1 -1
51	|													
52		PC	PC									A-B	STM1	EDGE DONE
53								G				72	Y	GET SYST. THRESHOLD
54												STM1	X	
55		G										72	D	COMPARE
56			G											
57			|		PC									GEN DEFAULT VALUE
58												D-E	72	SEND TO SYST.

PC = PRECHARGE DEVICE

G = GENERATE OUTPUT CHARGE PACKET

| = OUTPUT TO SUM BUCKET

a fair amount of the processing time so that the internal I/O may govern the ultimate operating speed of the computer. Implementing an NMOS version of the charge-coupled computer combined with less stringent charge transfer efficiency expectations could significantly increase the operating speed perhaps five times over what is estimated here, though such an increase is not needed for most applications.

2.3 Charge-Coupled Device Background

The term charge-coupled device refers to a broad range of semiconductor devices based on the non-equilibrium MOS capacitor and its coupling to adjacent devices through fringing fields. In this section, the deep depleted MOS capacitor will be discussed with respect to charge, electric field, electrostatic potential, energy bands and equilibrating currents. Inputting charge into the CCD potential well or 'bucket' will be covered in the second sub-section. The surface potential equilibration method will be discussed in some detail since an analogous operation is fundamental to the charge packet differencer circuit disclosed in chapter 5. Charge transfer in the semiconductor will be discussed in the third sub-section. Finally, the output of charge packets is briefly covered in the last sub-section.

Supporting this section is appendix B covering deep depletion carrier concentrations and appendix D, which is an analysis of the source-follower charge packet output sense amplifier used in the prototype circuits. It is assumed the reader has a basic working knowledge of semiconductor device physics. For additional treatment of the subjects touched upon in this section, Sze (1981), Nicollian and Brews (1982), Sequin and Tompsett (1975), Melen and Buss (1977), and Beynon and (1980) may be useful.

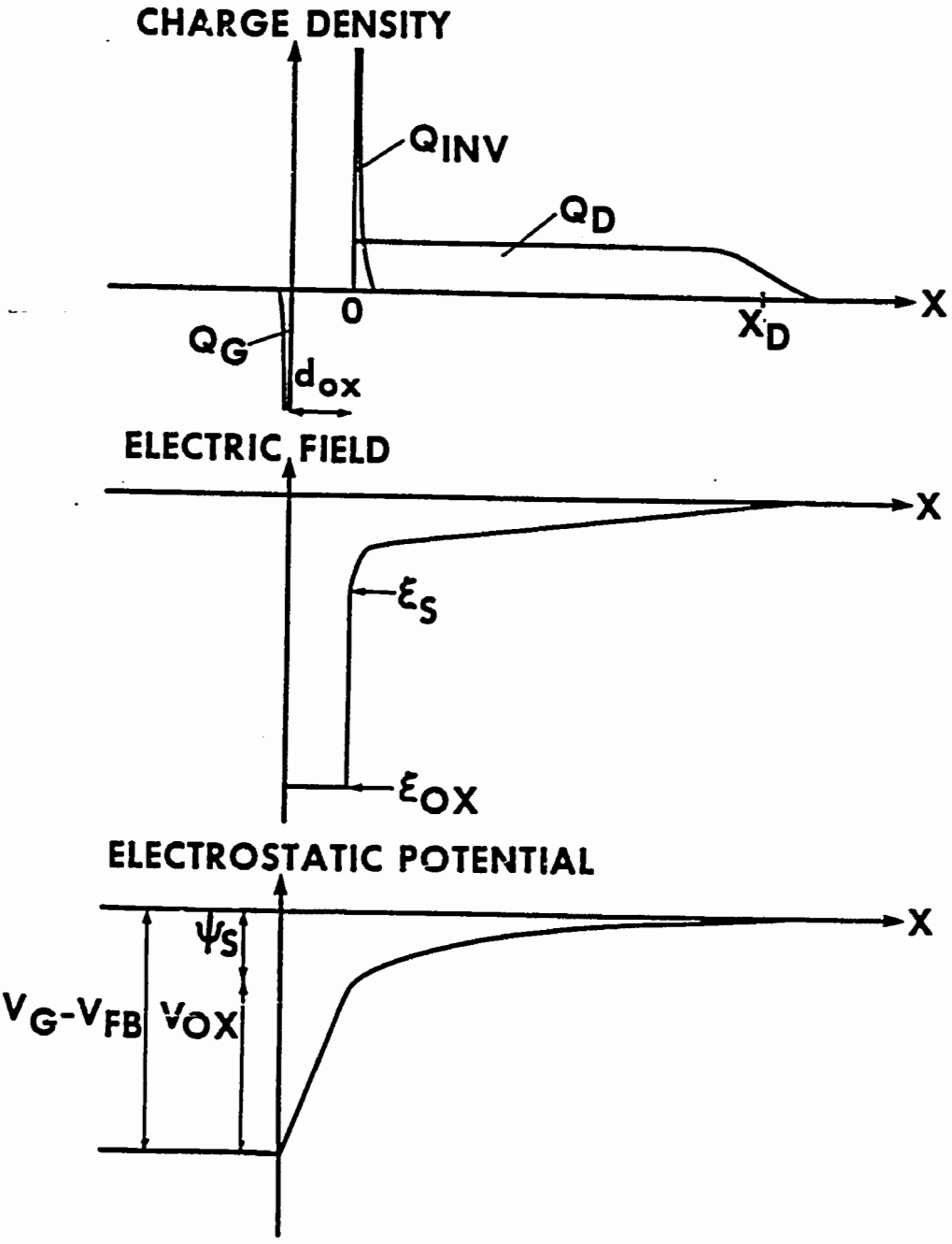
The Deep Depleted MOS Capacitor

A metal/insulator/semiconductor (MIS) structure is considered which is nominally aluminum/silicon dioxide/silicon (n). When a negative bias is applied to the metal with respect to the substrate, the majority carriers (electrons) are repulsed from the semiconductor surface and minority carriers are attracted. In accordance with the zero-current approximation introduced in appendix B, the system balances when the negative gate charge is balanced by an equal amount of positive charge in the semiconductor such that the change in electrostatic potential generated by the charge moment is equal to the applied bias. To simplify the analysis of non-ideal MIS structures, it is assumed that the oxide and interface trapped charge is fixed so that its effect and that of a work function difference between the metal and semiconductor (qW_{MS}) can be lumped into a single offset voltage or flat-band voltage V_{FB} defined as [Nicollian and Brews, 1982]

$$V_{FB} \triangleq W_{MS} + \int_{o}^{-d_{ox}} \frac{\rho_{ox}(x)}{\varepsilon_{ox}\,\varepsilon_{o}} (x + d_{ox})\, dx \qquad 2.3\text{-}1$$

where $\rho_{ox}(x)$ is the oxide and interface trapped charge density, x is measured from zero at the Si/SiO_2 interface, ε_{ox} is the relative dielectric constant of the oxide, ε_o is the permittivity of free space and d_{ox} is the oxide thickness.

In Fig. 2.3-1, charge, electric field and electrostatic potential in the system is shown. As illustrated in Fig. 2.3-1a, the total semiconductor charge Q_s consists of two components, uncompensated donor impurities (Q_D) and minority carriers in the depleted region (Q_{INV}) such that



2.3-1 a. Charge distribution in an MOS capacitor
b. Electric field distribution
c. Electrostatic potential

$$Q_D \triangleq \int_o^\infty q\left[N_D^+(x) - n(x)\right]\,dx \qquad 2.3\text{-}2$$

and

$$Q_{INV} \triangleq \int_o^{X_D} qp(x)dx \qquad 2.3\text{-}3$$

where x_D is the depletion region depth defined by the point where n(x) is half the donor concentration $N_D/2$. For constant doping, the depletion approximation yields

$$Q_D = qN_D x_D \qquad 2.3\text{-}4$$

In Fig. 2.3-1b, the electric field is illustrated, which is calculated as

$$\mathcal{E}(x) = \mathcal{E}_s + \int_o^X \frac{\rho(\hat{x})}{\varepsilon_s \varepsilon_o}\,d\hat{x} \qquad 2.3\text{-}5$$

and $$\rho(x) = q\left[N_D^+(x) - n(x) + p(x)\right] \qquad 2.3\text{-}6$$

and $\mathcal{E}_s$ is the electric field at the semiconductor surface,

$$\mathcal{E}_s = -Q_s/\varepsilon_s\varepsilon_o \qquad 2.3\text{-}7$$

as determined by Gauss'Law.

The electrostatic potential is shown in Fig. 2.3-1c, and is calculated by integrating the electric field

$$\psi(x) = \psi_s - \int_o^{\infty} \mathcal{E}(\hat{x})\, d\hat{x} \qquad 2.3\text{-}8$$

Use of the depletion approximation yields

$$\psi_s = -qN_D x_D^2/2\varepsilon_s\varepsilon_o + \delta\psi \qquad 2.3\text{-}9$$

where $\delta\psi$ is the potential drop across the inversion layer. For low inversion layer charge densities, $\delta\psi$ is negligible (sheet charge approximation). In this case combining Eq. 2.3-4 and Eq. 2.3-9 yields

$$-Q_D = (2q\varepsilon_s\varepsilon_o N_D\psi_s)^{1/2} \qquad 2.3\text{-}10a$$

or $$\psi_s = Q_D^2/2q\varepsilon_s\varepsilon_o N_D \qquad 2.3\text{-}10b$$

From appendix B, a careful calculation yields

$$Q_D + Q_{INV} = \left\{ \frac{2q\varepsilon_s\varepsilon_o}{\beta} \left[n_{no}(e^{\beta\psi_s} - \beta\psi_s + 1) + p_s\left(1 + \frac{p_{no}}{p_s}\beta\psi_s - e^{\beta\psi_s}\right)\right]\right\}^{1/2} \qquad 2.3\text{-}11$$

where $\beta \triangleq q/kT$, $n_{no} = N_D$, $p_{no} = n_i^2/n_{no}$ and p_s is the surface hole concentration which varies as the square of the inversion layer charge density.

In the (charge free) oxide, the electric field is determined from Gauss' Law as

$$\mathcal{E}_{ox} = -Q_s/\varepsilon_{ox}\varepsilon_o \qquad 2.3\text{-}12$$

Integrating this (constant) field across the oxide yields the change in potential across the oxide (V_{ox}) which is

$$V_{ox} = d_{ox}\mathcal{E}_{ox} = -Q_s/C_{ox} \qquad 2.3\text{-}13$$

where C_{ox} is the specific oxide capacitance defined as

$$C_{ox} = \varepsilon_{ox}\varepsilon_o/d_{ox} \qquad 2.3\text{-}14$$

The total potential drop across the MOS system is equal to the applied bias V_G such that

$$V_G - V_{FB} = V_{ox} + \psi_s \qquad 2.3\text{-}15$$

$$\text{or} \quad V_G - V_{FB} = -\left[\frac{Q_{INV}}{C_{ox}} + \frac{Q_D}{C_{ox}} + \frac{Q_D^2}{2q\varepsilon_s\varepsilon_o N_D}\right] \qquad 2.3\text{-}16a$$

$$\text{or} \quad V_G - V_{FB} = -\left[\frac{Q_{INV}}{C_{ox}} + (\phi_o\psi_s)^{1/2}\right] + \psi_s \qquad 2.3\text{-}16b$$

where ϕ_o, the body potential, is defined as

$$\phi_o \triangleq 2q\varepsilon_s\varepsilon_o N_D/C_{ox}^2 \qquad 2.3\text{-}17$$

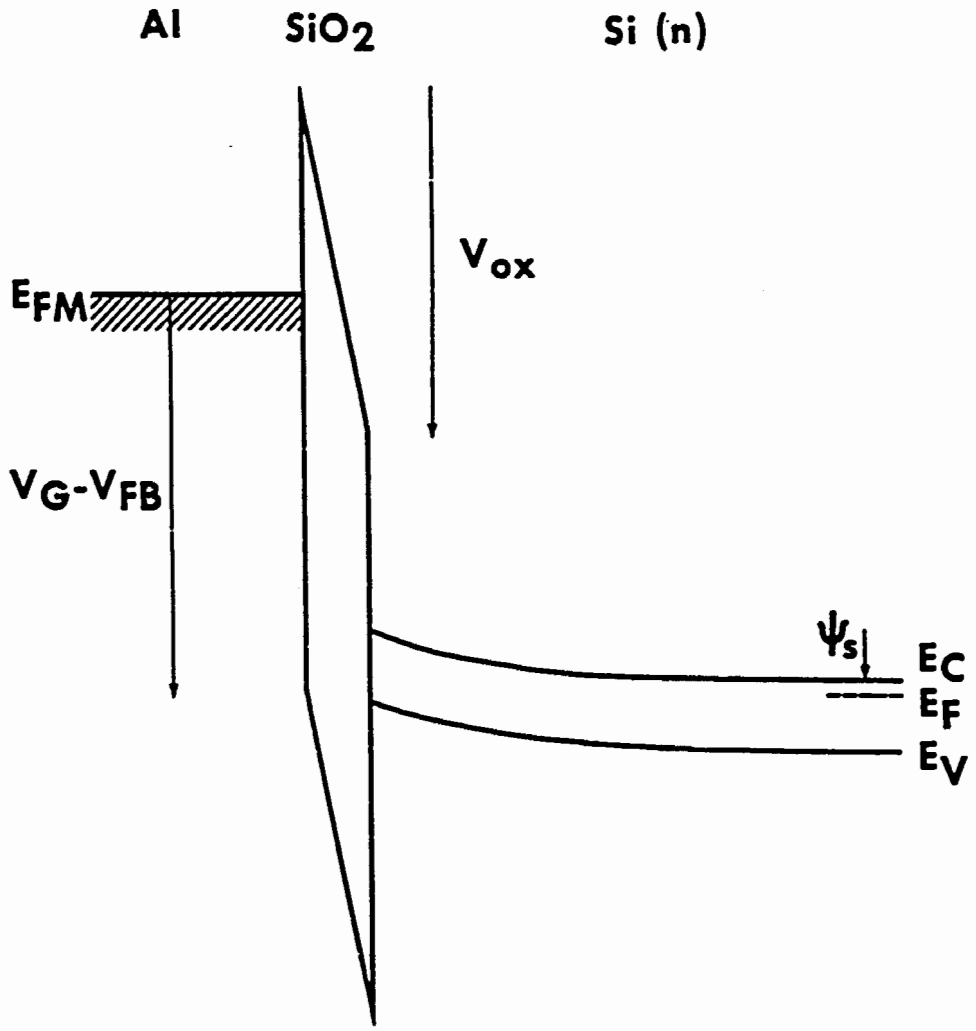
In equilibrium ($pn = n_i^2$), substitution of Eq. 2.3-11 divided by C_{ox} for V_{ox} in Eq. 2.3-15 ($p_s = e^{-\beta\psi_S}$ reveals that the surface potential is pinned at a value

$$\psi_s \simeq -2\phi_n + \delta\psi \qquad 2.3\text{-}18$$

where ϕ_n is defined as

$$\phi_n \triangleq \frac{kT}{q} \ln (N_D/n_i) \qquad 2.3\text{-}19$$

In the sheet charge approximation, Eq. 2.3-16 becomes



2.3-2 Energy band diagram for an MOS capacitor in slight deep depletion.

$$-(V_G - V_{FB}) = \frac{Q_{INV}}{C_{ox}} + (2\phi_n\phi_o)^{1/2} + 2\phi_n \qquad 2.3\text{-}20$$

Solution of Q_{INV} yields

$$-Q_{INV} = C_{ox}\,(V_G - V_T) \qquad 2.3\text{-}21$$

where the inversion layer formation threshold voltage is defined as

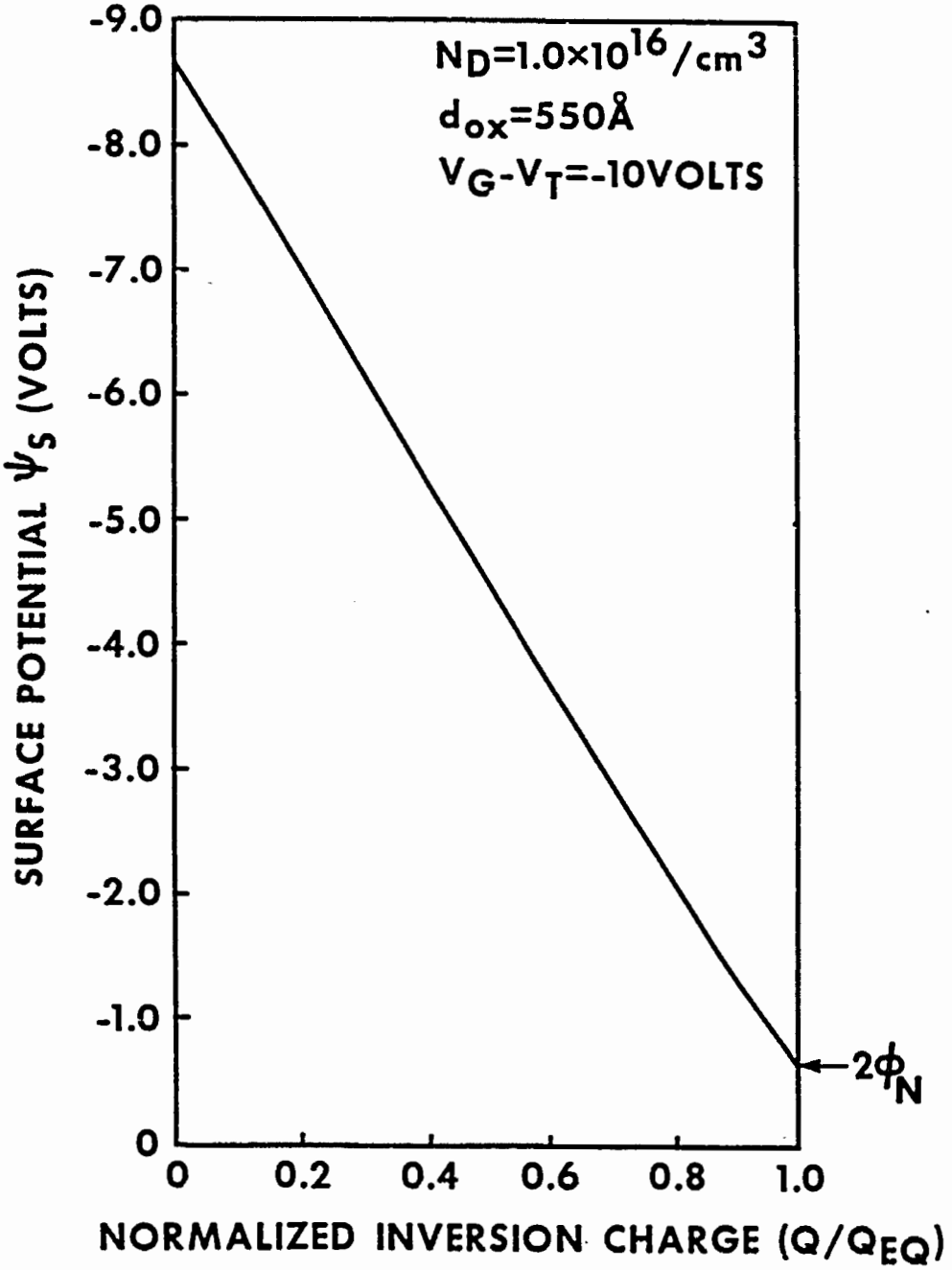
$$V_T \triangleq V_{FB} - (2\phi_n\phi_o)^{1/2} - 2\phi_n \qquad 2.3\text{-}22$$

For biases above threshold ($V_G < V_T < 0$), Q_{INV} less than that given by Eq. 2.3-21 implies that the semiconductor is in a state of deep depletion ($pn < n_i^2$). The surface potential is related to the applied bias and inversion layer charge according to

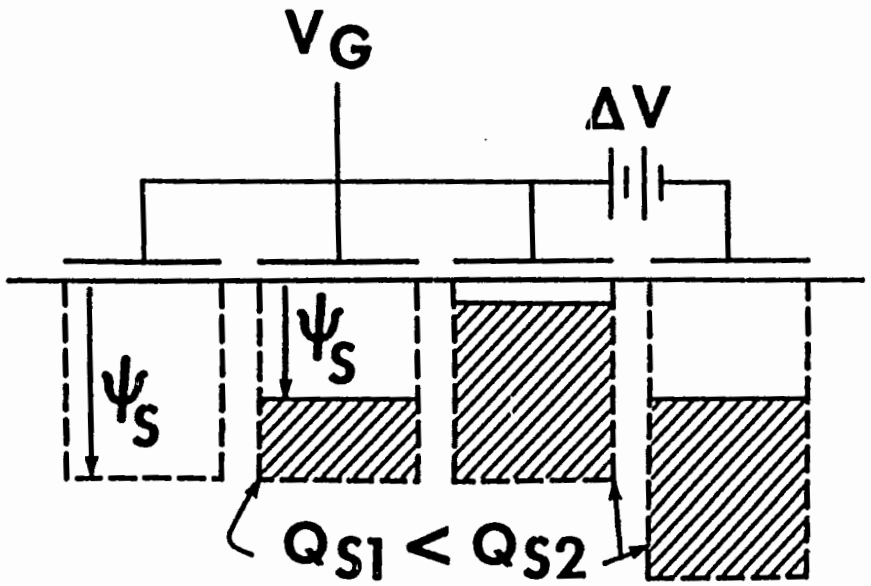
$$\psi_s = V_G - V_{FB} - \frac{\phi_o}{2} + \frac{Q_{INV}}{C_{ox}} + \left[\left(\frac{\phi_o}{2}\right)^2 - \phi_o\left(V_G - V_{FB} + \frac{Q_{INV}}{C_{ox}}\right)\right]^{1/2} \qquad 2.3\text{-}23$$

An energy band diagram for a slightly deep depleted MOS capacitor is shown in Fig. 2.3-2.

In Fig. 2.3-3, a plot of surface potential as a function of inversion layer charge is shown using Eq. 2.3-23. The plot shows a linear relationship in spite of the quadratic appearance of Eq. 2.3-23. This suggests that a 'bucket' diagram may be appropriate for representing the non-equilibrium inversion layer charge. The maximum depth of the bucket corresponds to the surface potential for no minority carrier charge. The extent to which the bucket is filled with charge represents the reduction in surface potential as shown schematically in Fig. 2.3-4.



2.3-3 Surface potential as a function of inversion layer charge normalized by $Q_{EQ} = C_{ox}\ (V_G - V_T)$.



2.3-4 Bucket model of relationship between surface potential and inversion layer charge.

The charge is shown at the bottom of the bucket to create an illusion of a fluid, but the charge is actually a sheet-like layer at the semiconductor surface.

The deep depleted MOS capacitor is in non-equilibrium ($pn < n_i^2$) and there is net thermal generation in the semiconductor. Using the recombination/generation statistical theory of Hall (1952) and Shockley and Read (1952) as written by Sah, et al. (1957), hereafter referred to as the Shockley-Read-Hall (SRH) net recombination rate U, one finds locally,

$$U = \int_{E_v}^{E_c} \frac{\sigma_p \sigma_n v_{th} n_t(E) \left[pn - n_i^2 \right]}{\sigma_n \left[n + n_i \exp\left(\frac{E-E_i}{kT} \right) \right] + \sigma_p \left[p + n_i \exp\left(-\frac{E-E_i}{kT} \right) \right]} dE \qquad 2.3\text{-}24$$

where σ_p and σ_n are the average hole and electron capture cross-sections of the trap density $n_t(E)$, v_{th} is the carrier thermal velocity equal to $(3kT/m^*)^{1/2}$. Equation 2.3-24 is often written with $n_t(E)$ as a delta function

$$n_t(E) = N_t \, \underline{\delta}(E-E_t) \qquad 2.3\text{-}25$$

to consider a monoenergetic set of traps of density N_t. Using this simplification with the traps situated at the optimum energy $E_t = E_i$, and setting $\sigma_p = \sigma_n = \sigma$, the SRH net recombination rate becomes

$$U = \frac{pn - n_i^2}{\tau \, (p + n + 2n_i)} \qquad 2.3\text{-}26$$

where τ is called the minority carrier lifetime and is defined as

$$\tau \triangleq (\sigma \, v_{th} N_t)^{-1} \qquad 2.3\text{-}27$$

In general, τ is a function of position through the trap density $N_t(x)$. In particular, it is useful to write

$$\tau^{-1} = \bar{\tau}^{-1} + S_o\,\underline{\delta}(x) \qquad 2.3\text{-}28$$

where S_o is the surface recombination velocity and is determined by the characteristics of the traps at the Si/SiO_2 interface, and $\bar{\tau}$ is a spatially averaged minority carrier lifetime.

The rate at which the surface re-equilibrates is determined by the net thermal generation rate. In deep depletion such that $pn << n_i^2$, the rate at which the inversion layer grows is determined by the sum of the surface generation rate, the integrated depletion region generation rate, and to a much lesser extent, diffusion of minority carriers to the surface, such that

$$Q_{INV} = J_S + J_D + J_B \qquad 2.3\text{-}29$$

The surface generation current J_S is

$$J_S = -\left.\frac{qS_o\,(pn - n_i^2)}{p + n + 2n_i}\right|_{x=o} \qquad 2.3\text{-}30a$$

$$\approx \frac{qS_o n_i^2}{p_s + 2n_i} \qquad 2.3\text{-}30b$$

The depletion region generation current is

$$J_D = -\int_o^{x_D} \frac{q\,(pn - n_i^2)}{\bar{\tau}\,(p + n + 2n_i)}\,dx \qquad 2.3\text{-}31a$$

$$\approx \frac{qn_i x_D}{2\bar{\tau}} \qquad 2.3\text{-}31b$$

The minority carrier diffusion current J_B is

$$J_B = \frac{qD_p}{L_p} \frac{dp}{dx}\bigg|_{x = x_D} \qquad 2.3\text{-}32a$$

$$= \left(\frac{qkT\mu_p}{\bar{\tau}}\right)^{1/2} p_{no} \qquad 2.3\text{-}32b$$

Lateral diffusion from generation sites surrounding a finite sized MOS capacitor may also contribute to the re-equilibration process [Schroder and Nathanson, 1970]. In Table 2.3-1, the currents have been evaluated for sample parameter values. It can be seen that the depletion current is several orders of magnitude larger than the bulk diffusion current.

The time required for the MOS capacitor to relax to its equilibrium state requires careful modeling of the current components described above [Collins and Churchill, 1975]. However, a simple expression can be obtained if one assumes that the net generation current starts from an initially high value and declines at a constant rate to zero after time Δt. The relaxation time is therefore

$$\Delta t = \frac{4C_{ox}\,(V_G - V_T)}{q\,n_i\,x_D^o}\,\bar{\tau} \qquad 2.3\text{-}33$$

where x_D is the initial depletion layer depth. A typical relaxation time is evaluated in Table 2.3-1 and is typically a million times larger than the minority carrier lifetime. The thermal relaxation time implies that in an analog CCD (continuous signal values), all operations must take place in a time small compared to Δt. Otherwise, the signal charge might be enlarged by an integrated dark current. Thus, the processing time for the charge-coupled computer should be less than 0.1 percent of the capacitor relaxation time calculated above. This signal retention

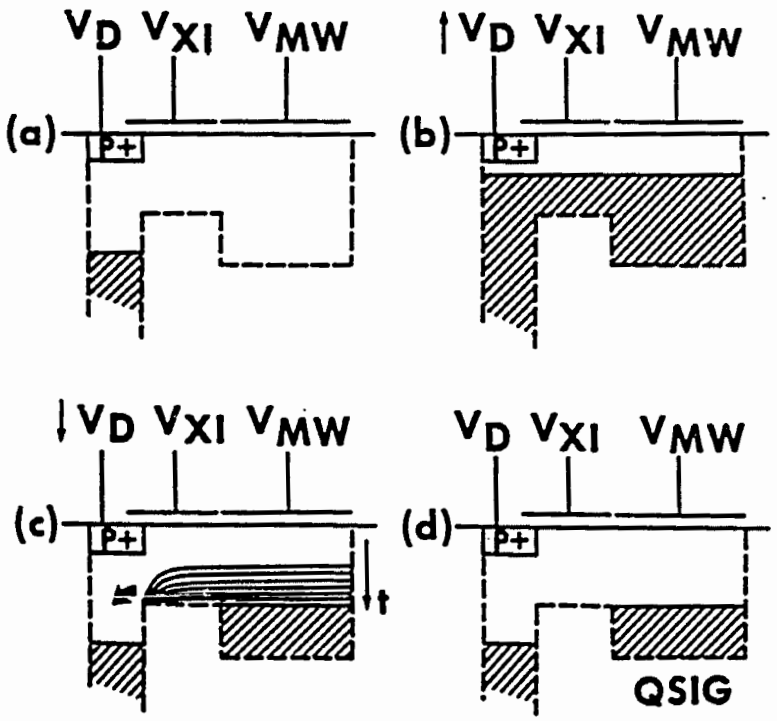
period problem is similar to that encountered in dynamic memory devices where the digital information must be periodically refreshed. Such a refresh cycle is not suitable in an analog circuit unless the dark current is extremely well characterized. This thermal relaxation process also implies that slight cooling of the device can lead to much larger storage times, and conversely, that slightly elevated temperatures can degrade the relaxation time.

Signal Charge Input

For CCDs, the amount of inversion layer charge in the deep depleted MOS capacitor is used as a signal quantity. The signal charge in most applications needs to be carefully metered through an appropriate electrical input circuit (except in image sensors). In the first CCD, the signal charge was created by applying a high voltage pulse to the capacitor causing the avalanche generation of carriers in the CCD bucket via a silicon breakdown process [Amelio, et al. 1970]. Improved performance may be obtained through the use of an input diode and input transfer gate. With an adjacent metering well biased such that the equivalent MOSFET channel under the input transfer gate is pinched off at the metering well end, current can be dynamically injected into the metering well. The amount of charge injected depends upon the injection period, the input diode bias, and on the input transfer gate bias.

However, in most applications, a linear relationship is desired between an input voltage and the signal charge packet. One of the simplest methods for achieving a highly linear relationship is the surface potential equilibration or 'fill and spill' technique [Tompsett, 1975]. This technique is illustrated in Figs. 2.3-5 (a-d). Initially the input diode is reverse biased at voltage V_D. The input transfer gate bias V_{XI} and metering well bias V_{MW} are applied such that

$$V_{MW} < V_{XI} < 0 \qquad 2.3\text{-}34$$



2.3-5 Surface potential equilibration method of setting charge in a CCD

a. Empty bucket is formed by applying biases.
b. Bucket is filled by raising diode voltage to substrate potential.
c. Diode is returned to reverse bias causing spilling of charge from surface.
d. Spillway potential results in charge retained in metering well.

and $V_D < V_{XI} - V_T < 0$ 2.3-35

as shown in Fig. 2.3-5a. In Fig. 2.3-5b, the input diode bias is reduced to V_D' near substrate potential causing an inversion layer to form under the input transfer gate and metering well gate [Grove and Fitzgerald, 1966]. In this case

$$V_{XI} - V_T < V_D' \leq 0 \qquad 2.3\text{-}36$$

The time required to fill the potential wells is approximately

$$t_F = \frac{2L_{MW} L_{XI} \left| V_{MW} - V_T \right|}{\mu_p \left| V_{XI} - V_T \right|^2} \qquad 2.3\text{-}37$$

where L_{MW} is the metering well bucket length, L_{XI} is the input transfer gate length, and μ_p is the effective minority carrier surface mobility, and where it is assumed $V_D' = 0$ and the metering well bucket width is equal to the input transfer gate width. Note that the fill time scales as L^2.

The diode voltage is next returned to its original reverse bias value causing the PN junction to collect minority carriers from the adjacent inverted surface. This charge 'spills' out as shown schematically in Fig. 2.3-5c until the spill channel is completely depleted of minority carriers. The surface potential in this channel is given by Eq. 2.2-23 with $Q_{INV} = 0$. During the spill process (which is diffusion dominated at the pinched-off end), the surface potential under the input transfer gate is equal to the surface potential under the metering well gate if the diffusion potential due to the change in carrier concentration is negligible (approximately equal to $\delta\psi$ in Eq. 2.3-9). Thus, as the conductivity of the spill channel goes to zero, the potentials remain

equal, as shown in Fig. 2.3-5d. Using Eq. 2.3-16 we write

$$V_{MW} - V_{FB} = -\left[\frac{Q_{SIG}}{C_{ox}} + \left(-\phi_o\psi_{SMW}\right)^{1/2}\right] + \psi_{SMW} \qquad 2.3\text{-}38$$

$$\text{and} \quad V_{XI} - V_{FB} = -\left[\left(-\phi_o\psi_{SXI}\right)^{1/2}\right] + \psi_{SXI} \qquad 2.3\text{-}39$$

Subtracting the latter equation from the former and using

$$\psi_{SXI} = \psi_{SMW} \qquad 2.3\text{-}40$$

yields

$$-Q_{SIG} = C_{ox}\,(V_{MW} - V_{XI}) \qquad 2.3\text{-}41$$

provided

$$V_{MW} < V_{XI} < V_T < 0 \qquad 2.3\text{-}42$$

and Q_{SIG} is zero otherwise. Thus, the fill and spill process allows for inherently linear operation with a differential mode as a bonus. Any dark current generated in either bucket results in the premature formation of the signal charge or is collected by the PN junction and does not affect the signal charge magnitude (except in cases of poor charge transfer efficiency). The time required for the spill process is determined by the charge transfer time constants discussed in the next sub-section using a transfer length L equal to the sum of L_{XI} and L_{MW}.

Charge Transfer

To effect lateral charge transfer, adjacent potential wells must be coupled through proximity so that their depletion regions overlap. This requires that the MOS capacitor electrodes be closely spaced. The spacing for good charge transfer efficiency should not exceed a few

microns [Amelio, 1972; Suzuki and Yanai, 1974]. Charge transfer in charge-coupled devices has been studied extensively and there are three basic mechanisms which control the transfer rate. The first is thermal diffusion [Kim and Lenzlinger, 1971] in which the charge packet remaining in the well decays according to

$$Q_{SIG}(t) = Q_{SIG}(0) \cdot \exp\left[-\pi^2 D_p t/4L^2\right] \qquad 2.3\text{-}43$$

where D_p is the minority carrier surface diffusion constant, and L is the potential well length. The decay constant scales as L^2.

The second mechanism is self-induced drift [Carnes, et al. 1972], which causes the charge packet to decay according to

$$Q_{SIG}(t) = Q_{SIG}(0)\ (1 + t/t_o)^{-1} \qquad 2.3\text{-}44$$

where the characteristic time t_o is given by

$$t_o = L^2 C_{ox}/1.57\ \mu_p Q_{SIG}(0) \qquad 2.3\text{-}45$$

and scales as L^2. This self-induced drift dominates the transfer process for long channel lengths until

$$\bar{Q}_{INV}/C_{ox} = kT/q \qquad 2.3\text{-}46$$

where $\bar{Q}_{INV}$ is the average inversion layer density across the channel. Thermal diffusion controls the transfer process for long channel lengths after a time period equal to approximately

$$t = (1.6)4L^2/\pi^2 D_p \qquad 2.3\text{-}47$$

has elapsed.

The third process is transfer assisted by the fringing field from the adjacent electrode [Carnes, et al. 1971]. Such a fringing field is

significant only for short channel lengths (i.e. less than 10 μm) and lower doping concentrations (i.e. N_D less than $10^{15}/cm^3$). The single carrier transit time in a fringing field is given by

$$t_{TR} = L^3 \ (1 + L/5x_D)^4/6.5\mu_p d_{ox} V_G \qquad 2.3\text{-}48$$

This transit time scales as L^7 so that the effect is dramatic for short channel lengths. Actual calculation of the signal charge transfer rate is complicated by the three-dimensional nature of the problem and is usually performed using numerical methods [for example, Lee and Heller, 1972]. Carnes, et al. (1972) calculate that a three order of magnitude increase in transfer speed is possible for a three fold decrease in channel length from 12 μm to 4 μm (for an n-channel device doped at $N_A = 10^{15}/cm^3$). Such a dramatic increase in speed encourages device layout with large W/L channel dimension ratios.

Signal Charge Output

Following CCD signal processing functions, the resultant signal charge packets must be sensed and transduced to current or voltage. One method is to collapse the CCD potential well and measure the bulk recombination current through the substrate [Amelio, et al. 1970]. The charge packet may also be collected by a reverse biased PN junction and the transient current detected. A much improved sensing technique is to integrate the charge packet on a floating capacitor. The change in capacitor voltage is used to modulate an output MOSFET [Kosonocky and Carnes, 1973]. This technique is discussed in detail in appendix D.

It is often useful to non-destructively sense the signal charge. This can be achieved by using a floating gate above the channel which again results in a change in effective capacitor voltage [Engeler, et al. 1970]. This technique tends to be non-linear due to the non-linear

depletion layer capacitance of the MOS structure. Improved sensitivity is obtained by using a series of floating gates in a distributed manner [Wen, et al. 1975] as described by Sequin and Tompsett (1975).

Table 2.3-1

Sample Values of MIS-CCD Quantities
(Aluminum/SiO_2/Silicon-n)

Work function voltage W_{MS}	0.24 volts
Oxide thickness d_{ox}	550 $\mathring{A}$
Relative dielectric constant ε_{ox}	3.9
Silicon doping concentration N_D	$1 \times 10^{16}/cm^3$
Silicon relative dielectric constant ε_s	11.7
Average minority carrier lifetime $\bar{\tau}$	10 μsec
Surface recombination velocity S_o	10 cm/sec
Bulk minority carrier mobility μ_p	480 $cm^2/v \cdot sec$
Surface minority carrier mobility μ_p	150 $cm^2/v \cdot sec$
Intrinsic carrier concentration n_i	$1.45 \times 10^{10}/cm^3$
Bulk minority carrier concentration P_{no}	$2.1 \times 10^4/cm^3$
Oxide capacitance C_{ox}	0.063 $\mu F/cm^2$
Silicon built-in potential ϕ_n	0.35 volts
Body potential ϕ_o	0.84 volts
Threshold voltage V_T	-2.0 volts
Gate voltage V_G	-12.0 volts
Equilibrium inversion layer charge Q_{INV}	0.630 $\mu C/cm^2$
Equilibrium depletion layer charge Q_D	0.048 $\mu C/cm^2$
Equilibrium depletion layer depth X_D	0.30 μm
Equilibrium surface potential ψ_s	-0.70 volts
Maximum depletion layer charge (Q_{INV}=0)	0.170 $\mu C/cm^2$
Maximum depletion layer depth (Q_{INV}=0)	1.06 μm
Maximum surface potential (Q_{INV}=0)	-8.75 volts
Deep depletion generation rate - U	$7.25 \times 10^{14}/cm^3$-sec
Maximum depletion region dark current J_D	12.3 nA/cm^2
Maximum surface generation current J_S (p_S=0)	11.0 nA/cm^2
Maximum bulk diffusion current J_B	3.7 pA/cm^2
Capacitor relaxation time Δt	102 seconds
Maximum CCD processing time	102 msec
Metering well voltage V_{MW}	-14 volts
Input transfer gate voltage V_{XI}	-4 volts
Metering well charge	0.630 $\mu C/cm^2$
Metering well length	10 mils

Table 2.3-1 (cont.)

Input transfer gate length	2 mils
Fill time	5.2 μsec
Diffusion transfer time constant	97 μsec
Self-induced drift time constant	0.3 μsec
Spill time (99.8% CTE)	~150 μsec

CHAPTER 3

FABRICATION OF PROTOTYPE CIRCUITS

3.1 Introduction

The fabrication of charge-coupled devices requires stringent control over semiconductor processing quality. Semiconductor defects, impurities, and surface traps can dramatically alter the performance characteristics of CCDs. Dark current, originating from bulk and surface traps shortens the storage time of deep depleted MOS capacitors. Bulk traps also contribute to the reverse leakage current of PN junctions. Surface traps can lead to charge transfer inefficiency and reduced dynamic range.

The processing of semiconductor devices at Yale has historically been limited to MOS capacitors and occasionally MOSFETs with the primary emphasis on (tunnel) oxide and interface properties rather than on semiconductor quality. [Low temperature, long channel MOSFET research excepted]. There has also been no work on the charge-coupling of MOS devices. To successfully couple two deep depleted MOS capacitors, the capacitor electrode spacing must be less than a few microns, depending upon the oxide thickness. The charge transfer efficiency degrades rapidly with increasing spacing. Thus the fabrication challenge was to improve the standard Yale MOSFET process [Ma, 1984] to reduce dark current and make sub-micron structures in a mil-scale facility.

The major reason why the Yale facility is presently limited to design rules in the one mil range is not because of photolithographic limits (5 microns) or particulate contamination (class 100 air), rather it is due to the inability to make multiple level mask sets which register to an accuracy much better than $\pm$ 0.5 mils over the die width. This

registration error source is the rotational error during the frame mounting of 10x reticles.

During the process development, a literature search revealed a particularly elegant method for making sub-micron gaps in the Yale environment. This process worked well and was found to be reliable. It is discussed in detail in the next section. The dark current remained enigmatic for some time primarily because of our own inexperience in dealing with this processing aspect, though it continues to commercially be a topical issue for dynamic memory circuits. The dark current was reduced in the Yale 8205 process through the use of a lower wet oxidation temperature and through the gettering properties of heavily (phosphorous) doped silicon. The tunnel oxide growth process developed by Dressendorfer (1978) and Lai (1979), and employed by Teng (1983) was retained without significant modification. However, due to a plumbing problem in the dry oxidation furnace, the first bistable MTOS junction grown in air (800 C) may have been inadvertently fabricated. Once corrected, the tunnel oxidation process worked well. A defective dopant source (Borosilicafilm) which has a short shelf life led to some initially false conclusions regarding the optimization of PN junction leakage current. This too was corrected by changing the dopant source (Borofilm) and the process was brought under control.

On the same die, charge-coupled devices with inter-electrode spacings less than one micron and low dark current, low leakage current PN junctions, and high quality bistable tunnel oxide junctions have been successfully fabricated. As it was once remarked, 'It's amazing what one can make with equipment industry discarded ten years ago; imagine what one could do with state-of-the-art equipment ...' [P. Santhanam, private communication].

3.2 Sub-Micron Gap Fabrication

The coupling of MOS capacitors to effect minority carrier charge transfer requires that the electrode spacing be quite small, typically less than a few microns [Amelio, 1972; Suzuki and Yanai, 1974]. As described in a review article by Kosonocky (1974), there have been several fabricated structures which satisfy this requirement. Single level metal electrodes [Amelio et al. 1970] are attractive from a structural simplicity point of view, but have suffered from two problems. First, creation of the gap requires micron scale photolithography followed by etching. Although the former is possible in a modern facility, etching of the gap remains difficult due to wetting problems with wet etches, and oxide/interface damage problems associated with ion milling or reactive ion etching process. Such difficulties were overcome using a shadow evaporation technique [Browne and Perkins, 1976]. However, a stability problem remains for the exposed oxide surface in between the two electrodes. For commercial devices, this latter effect has lead to the development of sealed structures.

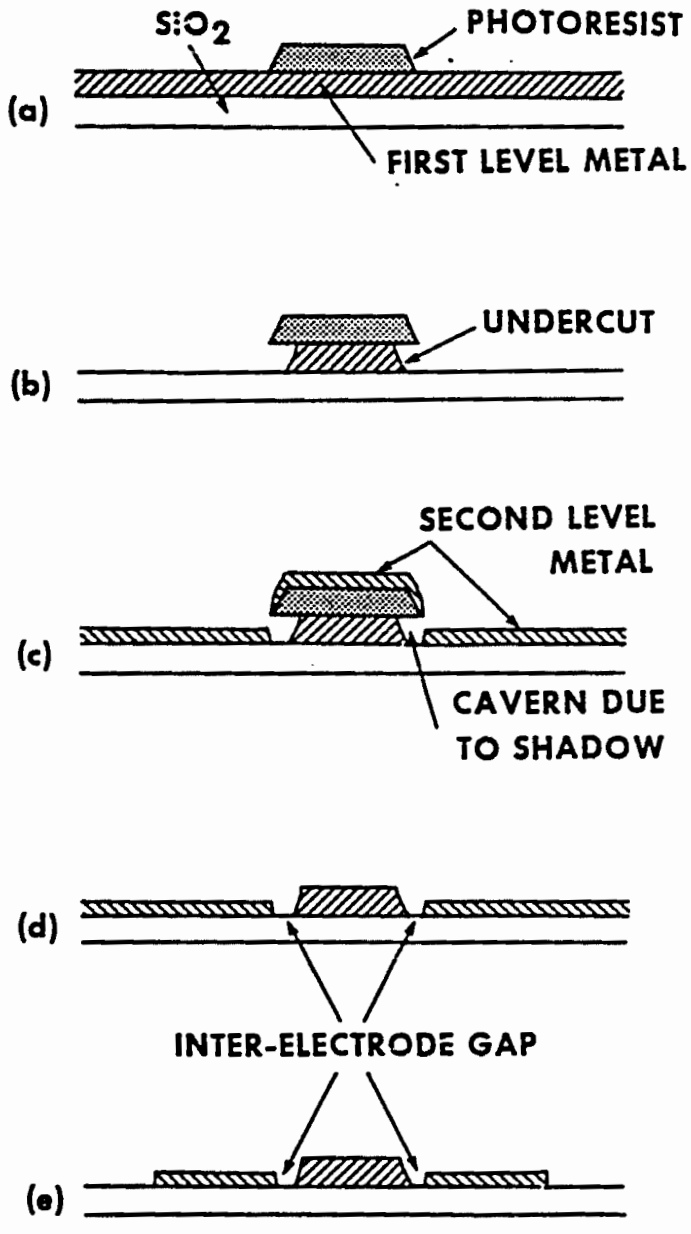
For example, a single-level polysilicon gate technology can replace the metal described above [Kim and Snow, 1972]. The polysilicon gates have the advantage of being defined by a subsequent implant process and being passivated by an upper level of oxide. Unfortunately the gap size is still limited by photolithography and the implant process.

The use of two levels of electrodes can reduce the effective gap width to zero if they overlap. The first level might be polysilicon and the second aluminum [Kosonocky and Carnes, 1973; Patrin, 1973]. The drawbacks of this technique are that alternating electrodes have a different oxide capacitance requiring higher clocking voltages so that shorts between overlapping electrodes can completely shut off a channel.

An improved overlapping gate technology employing three levels of polysilicon [Sequin, et al. 1973]. In this technology the primary polysilicon levels are isoplanar, with one electrode built up on one side to be able to overlap the adjacent electrode. An anodized aluminum structure can be used in a similar fashion [Collins, et al. 1974]. The advantage of the aluminum structure is the improved conductivity of the metallic gate, though Al_2O_2 is not as good an insulator as SiO_2. Other metal-insulator systems have also been investigated [Engeler, et al. 1970].

It was elected to use the Browne and Perkins (1976) method at Yale for several reasons. First, the charge-coupled differencer is best implemented using an isoplanar electrode structure (see chapter 5). Second, the Yale facility does not include a polysilicon deposition system. Third, the photolithographic resolution is limited to the five micron range on a wafer wide basis. Thus, the Browne and Perkins method was ideally suited for our requirements and capabilities.

The basic fabrication technique is illustrated in Figs. 3.2-1 (a-e). In Fig. 3.2-1a, a 3500 Å layer of aluminum has been coated with photoresist. The photoresist has been subsequently patterned by exposure and development in the usual way (see section 3.5). In Fig. 3.2-1b, the aluminum has been overetched resulting in a photoresist overhang. Following the overetch, a second layer of aluminum (2000 Å) is deposited without prior removal of the photoresist as shown in Fig. 3.2-1c. Note that the photoresist overhang shadows the aluminum evaporation resulting in a cavern. In Fig. 3.2-1d, the aluminum which was deposited on the photoresist has been removed using a liftoff process. The liftoff is effected by soaking the wafer in acetone for an hour. Although the liftoff is facilitated through the simultaneous use of ultrasonic agitation, such agitation causes degradation of tunnel



3.2-1 Sub-micron inter-electrode gap formation process.

a. Photoresist patterned
b. Overetch of aluminum results in undercut
c. Second metallization
d. Lift-off removes photoresist and metal cap
e. Final metal patterning complete

oxide durability and was not used in the Yale 8205 process. Following liftoff, the aluminum is patterned to its final form as shown in Fig. 3.2-1e.

In Figs. 3.2-2 (a-c) a series of scanning electron microscope (SEM) pictures are displayed which illustrate the gap formation process. In Fig. 3.2-2a, the photoresist overhang can be seen. Note the one micron size bar at the base of the picture. In Fig. 3.2-2b, the structure has been coated with 2000 Å of aluminum. The cavern formed by the photoresist overhang can be discerned. Finally, in Fig. 3.2-2c, the liftoff process has yielded a clean gap between the two metal electrodes. The gap size is slightly under one micron in this photograph.

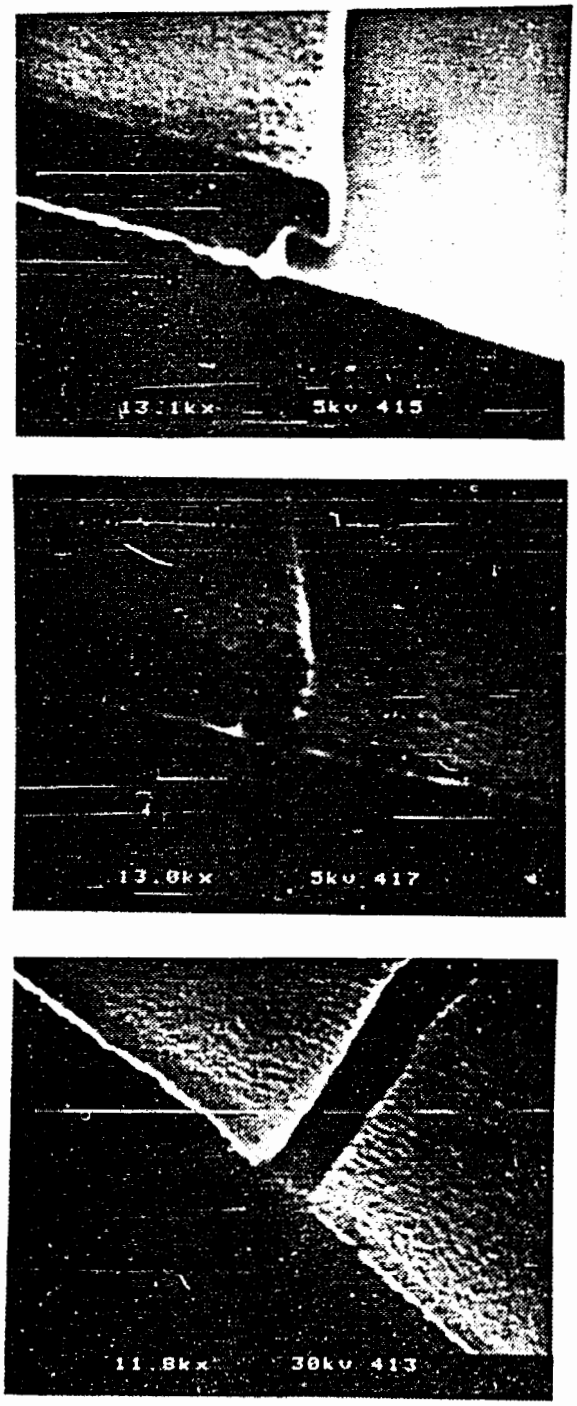
The size of the gap depends upon the thickness of the first layer of aluminum and the overetch time. To simplistically model the wet chemical aluminum etch process, it is assumed that the etching proceeds at a fixed rate normal to the aluminum surface. A series of contours depicting this model is sketched in Fig. 3.2-3. To calculate the gap width using Fig. 3.2-3, let t_d be the time it takes to etch the aluminum down to the SiO_2 (which is experimentally straight-forward to observe) and let t_w be the total etch time. As measured from the corner of the photoresist, a right triangle is formed by the vertical drop d and the horizontal gap w (assuming the subsequent shadowing process is clean and leaves a sharp shadow under only the photoresist). The hypotenuse h assuming a constant etch rate is

$$h = d^2 + w^2 \qquad 3.2\text{-}1$$

or

$$h = (d/t_d)\ t_w \qquad 3.2\text{-}2$$

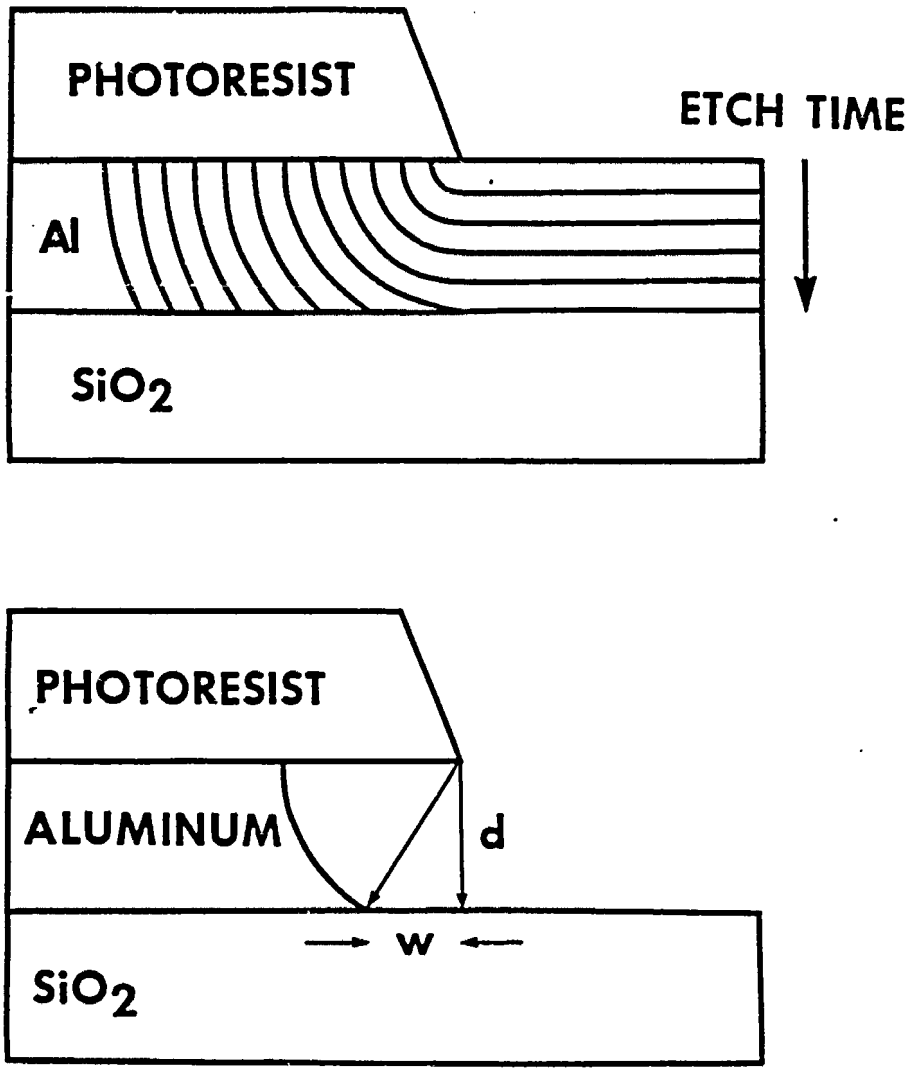
Solving for the total etch time yields



3.2-2 SEM photographs showing gap formation process

a. Photoresist overhang/aluminum undercut
b. Second metallization leaves cavern structure
c. Lift-off leaves clean gap structure (≈ 1 μm)

(Photographs by R. Wisnieff)



3.2-3 Simple model of etch process profiles

$$t_w = t_d \left[(w/d)^2 + 1\right]^{1/2} \qquad 3.2\text{-}3$$

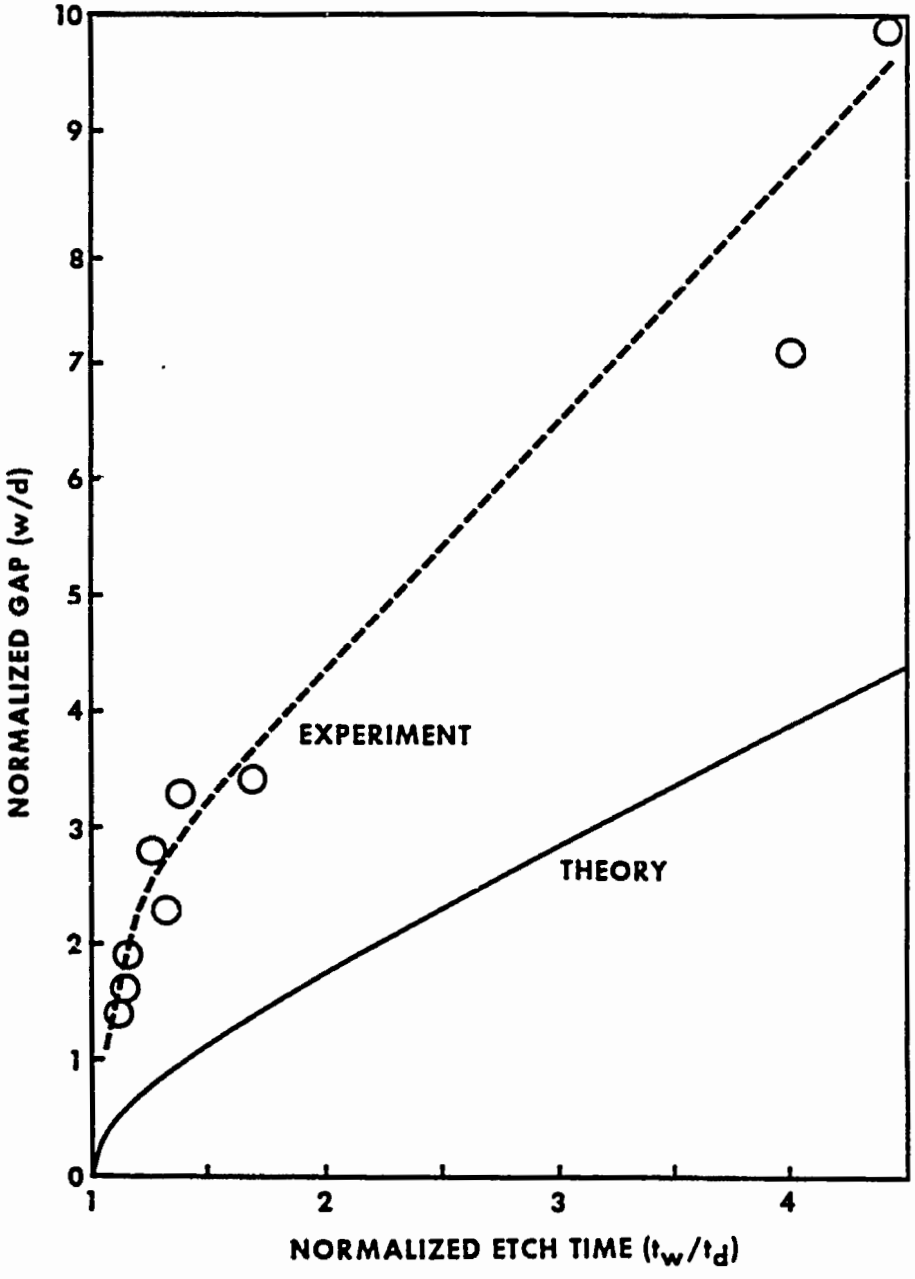
This equation is a useful experimental predictor if the etch rate is unknown (due to etch temperature fluctuations, for example). If the aluminum thickness d is measured during the evaporation process, then the total etch time according to the simple model may be determined. For example, suppose d = 3500 Å and w was to be one micron. From Eq. 3.2-3 it follows that the total etch time should be three times as long as the time required to etch down to the SiO_2.

Unfortunately, this simple model fails to accurately determine the size of the gap. In Fig. 3.2-4, several experimental data points measuring the ratio w/d as a function of t_w/t_d are plotted. By rewriting Eqs. 3.2-3,

$$w/d = \left[(t_w/t_d)^2 - 1\right]^{1/2} \qquad 3.2\text{-}4$$

This theoretical curve is also plotted in Fig. 3.2-4, indicating the failure of the model to predict the gap size. Indeed, the theory underestimates the gap size by a factor of two, though it seems to describe the shape of the experimental curve reasonably well. The discrepancy cannot be explained in terms of an aluminum oxide layer initially delaying the etch process. Although the principle is reasonable, the delay time must be nearly equal to the entire time it takes to etch down to the SiO_2 in order to cause the data to match such a hypothesis. Neither can the discrepancy be explained in terms of a broader than expected shadow. Such an effect should yield a nearly constant offset between the two curves of Fig. 3.2-4 since the aluminum thickness was nearly always 3500 Å.

Although there is no strong a priori reason to support this explanation, the experimental evidence suggests an etch anisotropy such that the horizontal etch rate exceeds the vertical etch rate. Such

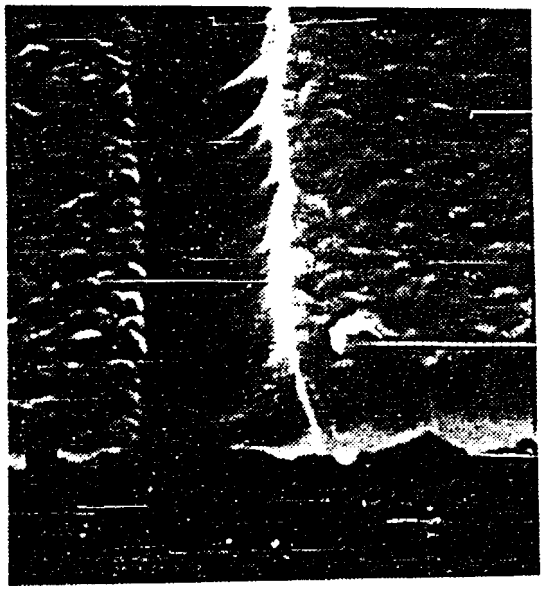


3.2-4 Measured gap width as a function of overetch time.

anisotropy may be due to an interaction between the photoresist and etching solution, or due to polycrystalline aluminum grain boundaries facilitating horizontal etching. In partial support of the anisotropy hypothesis, two scanning electron microscope photographs taken early in the sub-micron gap process development are displayed in Figs. 3.2-5 (a,b). In the first photograph, the basic profile illustrated in Fig. 3.2-3 can be readily discerned, though it appears to be elongated in the horizontal direction. In Fig. 3.2-5b, the gap is shown from an angled viewpoint following photoresist removal. The irregularity of the undercut edge (since improved, see Fig. 3.2-2c) is of roughly the same size as the observed grain size, supporting the suggested effect of grain boundaries on the etch process.

This latter concept was put to work in subsequent processing runs. From the preceeding statement, it follows that the smaller the grain size, the less irregular the gap will be. To achieve a smaller aluminum grain size, the evaporation rate was increased to a higher level. Of course a very large grain size should also yield improved gaps (as well as establish the role of grain boundaries in the observed anisotropy) but this will have to wait for future investigation. It appears that the worst grain size is one which is the same size as the gap itself.

Mask design for the gaps is straight forward if done properly. There is a subtle pitfall though, which is associated with slight mask misalignment. Such a pitfall resulted in the redesign of the first metal level mask. There are two ground rules which are useful for the design of the metal level masks. First, one should try to leave as much of the initially evaporated metal on the wafer as possible. This metal can be used to protect tunnel oxide devices and the gate oxide from contamination and physical damage. Second, the dimensions of the first metal mask in the direction parallel to the gap should be exaggerated to



3.2-5 SEM photographs showing (early) gaps

a. Undercut profile suggesting etch anisotropy
b. Top view showing enhanced etching along grain boundaries

(Photographs by A. Pooley)

prevent misalignment shorts subsequent to the second level metal definition.

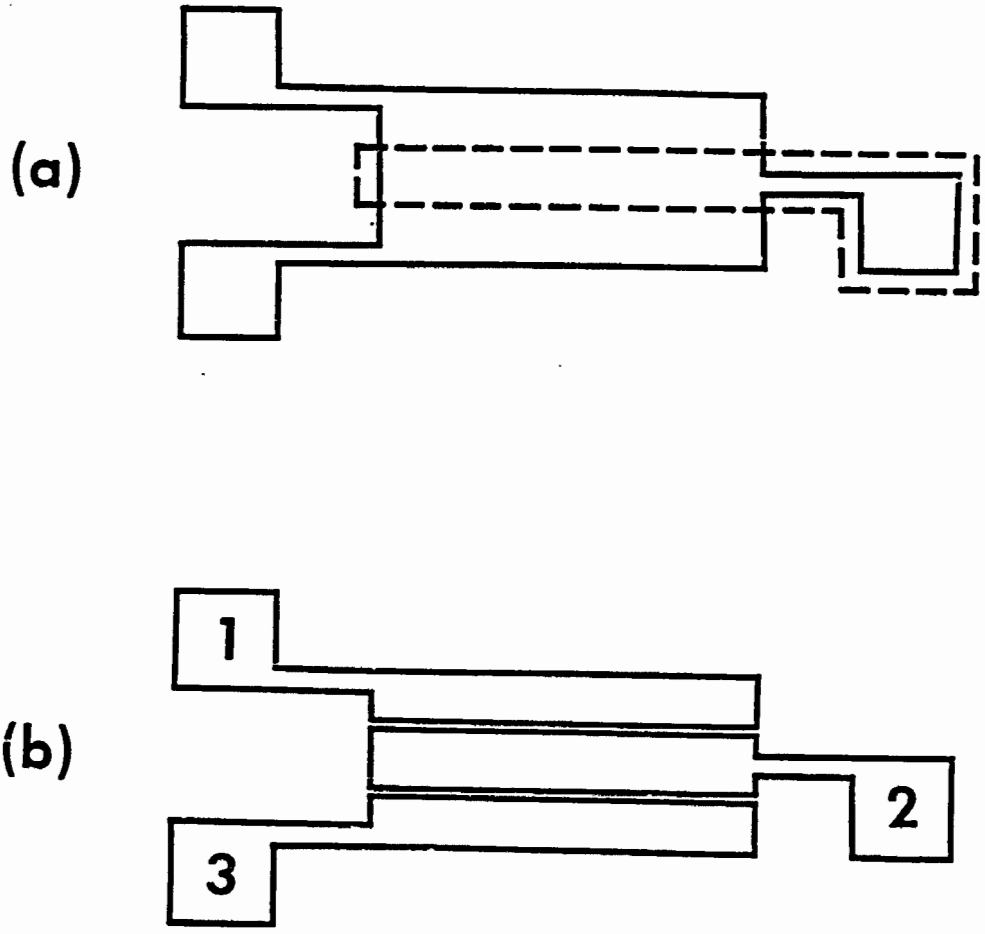
In Fig. 3.2-6a, an example of first and second level masking patterns is shown for making three electrodes and two gaps. The dashed line shows the border of the first level metal etch pattern. The metal within the border will be removed during the course of the overetch process. Following evaporation and lift off, all metal outside of the border defined by the solid line will be removed. The dashed line within this border will be an inter-electrode gap as shown in Fig. 3.2-6b. Slight misalignment of two patterns will still result in the electrode pattern of Fig. 3.2-6b except that the relative areas of electrodes 1 and 3 will be altered (though electrode 2 will not be affected. Note that this layout represents the reverse pattern of that illustrated in Fig. 3.2-1.

The inter-electrode gap formation process has been found to work reliably. Perhaps its only drawback is that the topology of the process makes gap intersections (such as a 'T' intersection) nearly impossible to fabricate.

3.3 Dark Current Reduction

Causes and Effects

A charge-coupled device is a non-equilibrium device and its key to success is in the length of time it takes to re-equilibrate. For the MOS structure in deep depletion, re-equilibration involves the generation of a minority carrier inversion layer. As discussed in chapter 2, this rate of change of inversion layer charge $\dot{Q}_{INV}$ due to minority carrier generation at the oxide-silicon interface, generation in the depleted silicon, and minority carrier extraction from the bulk is referred to as dark current. The bulk extraction current is usually



3.2-6 a. First level (dashed border) and second level (solid border) mask patterns for metal electrodes

b. Resultant electrode pattern (gap size exaggerated)

negligible so that interface generation and depletion region generation constitute the dark current, and depending upon processing conditions, can exchange dominating roles.

Dark current due to interface traps and dark current due to depletion region generation are nearly independently determined by the processing conditions, but the physical mechanisms describing the generation processes are similar. Generation/recombination is most easily described statistically in terms of rates and concentrations. In this work, the generation rate is of primary concern. According to the Shockley-Read-Hall recombination process, the local generation rate is proportional to the local trap density (see chapter 2). This is true at the interface and in the depletion region so that the key to dark current reduction in charge-coupled devices operating at room temperature is to reduce the number of active traps.

The interface contributes dark current in two ways. First, the interface traps act as stepping stones for the excitation of electrons into the conduction band. The ability of these interface traps to act as stepping stones depends upon their occupancy. It works out that the type of trap and its energy distribution results in a decreased number of active stepping stones for increased inversion layer charge during the re-equilibration cycle. This is the reason for using the so-called 'fat-zero' [Carnes and Kosonocky, 1972] signal charge in surface channel CCDs to reduce the number of active interface traps and reduce dark current effects, as well as improve charge transfer efficiency.

The second type of dark current arises from minority carrier emission from interface traps, as opposed to the electron-hole pair generation mediated by a trap. In a CCD, the MOS capacitor is repeatedly pulsed from deep depletion to flat-band with varying amounts of charge in the

bucket. Suppose a large signal charge resident in the potential well filled all the trap states with minority carriers. When that charge packet is transferred out and an empty well left behind, those filled interface traps will tend to emit those captured carriers.

Quantitatively, the mean number of carriers emitted in time t per unit area from a set of traps initially full at time t = 0, N(t), is given by [Tompsett, 1973]

$$N(t) = \int_0^{E_{G/2}} D_{it}(E) \cdot \left\{ 1 - \exp\left[-t\sigma_{it}(E) v_{th} N_C \exp\left(-\frac{E}{kT} \right) \right] \right\} dE \qquad 3.3\text{-}1$$

where D_{it} is the interface trap areal density and σ_{it} the trap cross-section. Assuming D_{it} and σ_{it} are constant, this can be simplified to

$$N(t) = kTD_{it} \ln (\sigma_{it} v_{th} N_C t) \qquad 3.3\text{-}2$$

where $(\sigma_{it}\ v_{th}\ N_c)^{-1} << t << \tau_{mb}$ and τ_{mb} is the midband interface trap time constant. The time derivative of this latter expression is the dark current due to interface trap emission. If the emission time constant is close to the clocking period of the CCD, this dark current has the worst effect on charge transfer efficiency.

The salient point of the preceeding discussion is two-fold. First, the elimination or passivation of interface traps is quintessential to the suppression of dark current. Although in MOS capacitor analysis, order of magnitude changes in interface trap densities are routinely observed, even a factor of two can dramatically affect the viability of a CCD device. The second point is that the interface trap contribution is history dependent with a memory dictated by the interface trap time constants.

The reduction of interface traps has been a major part of research and development in the semiconductor industry and the processing techniques used to reduce interface traps has become imbedded in the normal MOS process at Yale. The problem, as will be discussed in section 3.5, is that these techniques may be incompatible with other constraints in the process.

Dark current originating in the deep depleted silicon is bulk trap dependent. From the Shockley-Read-Hall equation, it is clear that the local trap density and local carrier concentration both control the local thermal generation rate. Bulk traps can be physically realized in many ways. Metallic impurities such as gold provide efficient mid-gap traps for carrier generation [Bullis, 1966]. Silicon crystal defects also provide traps for carrier generation [Tanikawa, et al. 1976; Unter, et al. 1977]. In this manner, unpaired silicon bonds in the defect probably behave similarly to unpaired silicon bonds at the surface. These defects may be present in off-the-shelf wafers but they are usually introduced by the fabrication process. One of the more serious defects is the oxidation induced stacking fault [Ravi and Varker, 1974]. This defect is nucleated during the oxidation process by local excess oxygen diffusing into regions of excess vacancy concentration in the crystal. These planar defects usually extend from the silicon surface to several microns in the bulk and are 1-2 microns in diameter. The elastic strain field surrounding these defects can lead to a precipitation of metallic and other impurities around the defect, contributing a double punch to the dark current. The elastic strain field plus impurity decoration can result in a locally enhanced generation zone 20 microns in diameter [Varker and Ravi, 1974].

The gettering effect associated with silicon defects has recently been exploited by creating silicon defects deep in the substrate by

intentional processing. This intrinsic gettering process aids in the removal of impurities from active regions of the semiconductor [Nelson, et al. 1983; Swaroop, 1984].

The key to dark current suppression in the deep depleted silicon is to first avoid the introduction of both metallic impurities and silicon defects into the substrate, and second, to getter any impurities which are introduced. This discussion will be continued following the next sub-section.

Experimental Characterization of Dark Current

Determination of the dark current is closely related to the determination of minority carrier lifetime, but simpler. The minority carrier lifetime is a local parameter describing the local generation/recombination zone, and if the zones are related to silicon defects in a nearly defect free substrate, it is inappropriate to describe the semiconductor by a single lifetime parameter. The dark current is a lumped parameter which includes surface trap generation as well as bulk trap generation and is the natural parameter to describe a CCD. The problem with using the dark current is that the dark current is history dependent and voltage dependent. To a lesser extent, the dark current is also a local parameter and will vary from capacitor to capacitor, but will only vary dramatically in a few pathological capacitors. Such dark current 'spikes' have long plagued the CCD industry and are accepted as a matter of course as long as the density is low.

To measure the dark current, a pulsed capacitance measurement (C-t) is used. The capacitor is biased at -5.0 volts (inversion) through a Boonton model 72B capacitance meter. The capacitor is exposed to light momentarily to guarantee that the surface is inverted. For devices with

low dark currents (long 'lifetime') it was observed that when illumination is removed, the capacitance decreases rapidly at first, but would then decrease slowly to its steady state value in the 10-30 second range. Such a decrease must be due to an increase in depletion layer depth corresponding in this constant bias condition to a decrease in surface charge, suggesting slow surface recombination or diffusion of the optically generated super-inversion layer.

Once the device has returned to equilibrium (as determined by a constant capacitance), the capacitor bias is stepped to -10.0 volts, resulting in deep depletion. The capacitance decreases immediately, and then slowly returns to nearly its initial value as the semiconductor re-equilibrates. The time it takes for the semiconductor to return to equilibrium is referred to as the capacitor relaxation time as shown in Fig. 3.3-1.

The dark current may be determined using the Zerbst relation [Zerbst, 1966] as described by Nicollian and Brews (1982, pg. 411). The oxide voltage is written as

$$V_G - V_{FB} - \psi_s(t) = (Q_{INV} + Q_D)/C_{ox} \qquad 3.3\text{-}3$$

Differentiating with respect to time yields

$$- \dot{\psi}_s = (\dot{Q}_{INV} + \dot{Q}_D)/C_{ox} \qquad 3.3\text{-}4$$

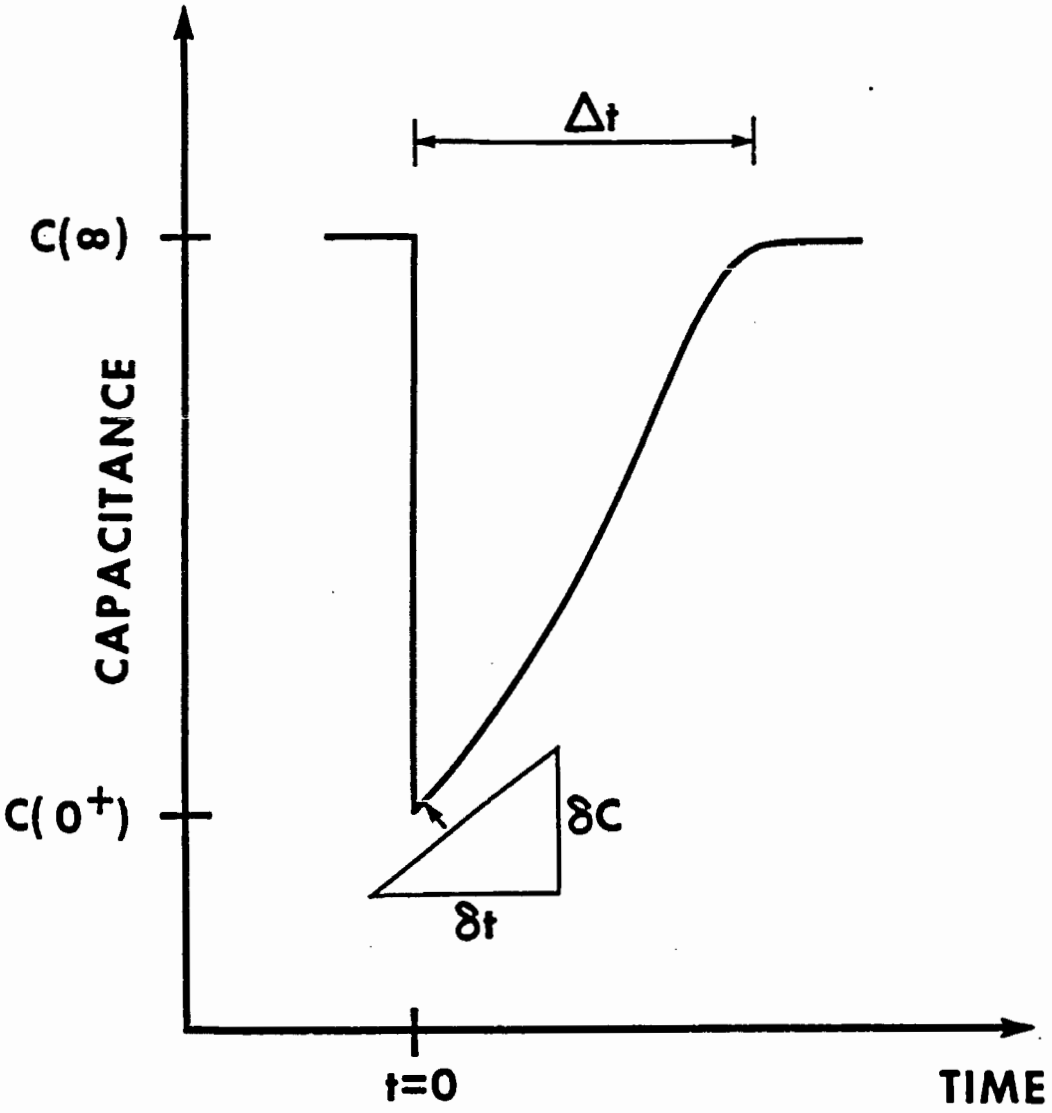
Rearranging gives

$$- \dot{Q}_{INV} = C_{ox} \dot{\psi}_s - \dot{Q}_D \qquad 3.3\text{-}5$$

For an arbitrary doping profile, it is true that

$$\dot{Q}_D = qN_D\dot{x}_D \qquad 3.3\text{-}6$$

and that if the voltage drop across the inversion layer is ignored,



3.3-1 Experimental capacitance as a function of time in response to a five volt step bias.

$$-\dot{\psi}_s = (qN_D/\varepsilon_s\varepsilon_o)x_D\dot{x}_D \qquad 3.3\text{-}7$$

Substituting these latter equations into Eq. 3.3-5 results in

$$-\dot{Q}_{INV} = qN_D(x_DC_{ox}/\varepsilon_s\varepsilon_o + 1)\dot{x}_D \qquad 3.3\text{-}8$$

The inversion layer depth x_D is accurately probed by the high frequency capacitance according to

$$x_D = \varepsilon_s\varepsilon_o\ (1/C - 1/C_{ox}) \qquad 3.3\text{-}9$$

so that Eq. 3.3-8 becomes

$$-\dot{Q}_{INV} = (q\varepsilon_s\varepsilon_oN_DC_{ox}/C)\ \frac{d}{dt}\ (\frac{1}{C}) \qquad 3.3\text{-}10$$

or finally,

$$\dot{Q}_{INV} = (q\varepsilon_s\varepsilon_oN_DC_{ox}/C^3)\ (dC/dt) \qquad 3.3\text{-}11$$

In this work, the Zerbst relation (Eq. 3.3-11) is evaluated just after the step voltage is applied. Thus, the dark current is consistently defined for a five volt bucket.

The capacitor relaxation time Δt can also be used to obtain an average value of the dark current during the re-equilibration process. The incremental charge collected in a re-equilibrated CCD bucket ΔQ_{INV} is proportional to the step voltage ΔV_G such that

$$\Delta Q_{INV} = C_{ox}\ \Delta V_G \qquad 3.3\text{-}12$$

The average dark current over the interval Δt is

$$Q_{INV} = C_{ox}\ \Delta V_G/\Delta t \qquad 3.3\text{-}13$$

If the dark current decreased at a constant rate to zero during the re-equilibration period, one would expect that the average dark current (Eq. 3.3-13) would be half the initial instantaneous dark current (Eq.

3.3-11).

If desired, an average minority carrier lifetime $\bar{\tau}_p$ may be determined by assuming that the thermal generation rate in the incremental depletion region is given by the Shockley-Read-Hall expression (chapter 2) using $p,n << n_i$. Such an assumption is specious both in terms of the actual Shockley-Read-Hall rate as determined by the actual carrier concentrations [Collins and Churchill, 1975] and in terms of the local variations in generation centers [Varker and Ravi, 1974], but nevertheless gives a reasonable lumped estimate. The dark current is therefore

$$\dot{Q}_{INV} = \frac{qn_i}{2\bar{\tau}_p} \left[x_D^o - x_D^\infty\right] \qquad 3.3\text{-}14$$

Employing Eq. 3.3-9 and solving for $\bar{\tau}_p$ gives

$$\bar{\tau}_p = \frac{q\varepsilon_s\varepsilon_o n_i}{2\dot{Q}_{INV}} \left[(1/C)\Big|_{t=0^+} - (1/C)\Big|_{t=\infty}\right] \qquad 3.3\text{-}15$$

Figure 3.3-1 shows an actual C-t plot as measured for the active and parallel channel metering wells of device 9A13 (see Fig. 4.3-2). The device area is 203 mils^2 with a stray capacitance estimated to be 3 pF. The effective oxide capacitance is 81 pF. At $t = 0^+$, the capacitance drops from 27 pF to 12 pF, and then increases at the initial rate of 0.11 pF/second. The capacitor relaxation time Δt is 87 seconds.

Using Eq. 3.3-11 with N_D equal to $1.0 \times 10^{16}/cm^3$, the five volt bucket instantaneous dark current is 11 nA/cm^2. The average dark current using Eq. 3.3-13 is 3.8 nA/cm^2. Using the approximate calculation of Eq. 3.3-15, the lifetime appears to be 6.6 μsec. Using a

more sophisticated approach[1], lifetime analysis yields a value of 8.1 μsec.

Dark Current Reduction Experiments

The device dark current was not initially thought to be a major fabrication problem. However, following the fabrication of the second test lot of 8205 wafers (which were nearly structurally identical to the final lot of wafers) it was discovered that the dark current exceeded 3000 nA/cm^2. To estimate the acceptable dark current level, assume that no more than a fraction η of a maximum signal charge can be generated by dark current in a maximum integration period T_I. The dark current should therefore not exceed

$$\dot{Q}_{INV}^{MAX} \leq \eta \, C_{ox} \, (V_G - V_T)/T_I \qquad 3.3\text{-}16$$

Setting η to one part in 2^9, and using a maximum integration period of 50 milliseconds, the dark current in a five volt bucket should be less than 12 nA/cm^2. Thus the challenge was to reduce the dark current introduced by the basic Yale MOSFET process [Ma, 1984], channel stop diffusion, tunnel oxidation and sub-micron gap formation by three hundred fold.

The first experiment was to resolve the issue of why other researchers in our group obtained long capacitor relaxation times ($\sim$ 60 seconds) yet wafer 2A had a relaxation time measured in tenths of seconds. In collaboration with V. Zekeriya, three wafers were cleaned by the standard process [Dressendorfer, 1978] and simultaneously given a dry oxidation at 1000 C for 60 min. Two wafers were p-type and one was n-type. One of the two p-type wafers had been wet oxidized at 1100 C

---

[1]A more sophisticated approach assumes the entire depletion region initially contributes to the dark current

for 60 min. and then stripped of the SiO_2 prior to dry oxidation. As summarized in Table 3.3-1 (expt. 1), the wafer which had undergone prior steam oxidation had a dark current greater than 2000 nA/cm^2 whereas the two wafers which had only had the subsequent dry oxidation had currents of 120 nA/cm^2 (n) and 18 nA/cm^2 (p). Clearly, something happened to the silicon during the steam oxidation which it 'remembered' even after the wet oxide had been stripped.

The second experiment attempted to resolve a contamination problem in the steam oxidation furnace. Two wafers were scribed in half and cleaned. One half of each wafer was dry oxidized in the steam furnace tube, and the other half was dry oxidized in the dry oxidation furnace tube. Each was oxidized at 1050 C for 60 min. followed by an in-situ N_2 anneal for 30 min. as summarized in Table 3.3-1 (expt. 2). The wafers were given a 15 minute post metallization N_2 anneal (PMNA) at 420 C. The dark currents ranged from 1 to 30 nA/cm^2, and failed to indicate that either furnace was contaminated, except that a bias-temperature stress capacitance check revealed the presence of mobile ions (perhaps sodium) through a half volt C-V shift.

The furnaces were cleaned and the experiment retried (Table 3.3-1, expt. 3). The results were nearly identical (including sodium contamination) except that the wafer half closest to the furnace mouth in the dry oxidation tube showed an order of magnitude higher dark current than the other three wafer halves. The wafer in the same position in expt. 2 was similarly afflicted, even though the wafer brand had been subsequently swapped. The only difference between the two furnaces after cleaning was the presence of a platinum-rhodium thermocouple in the dry oxide furnace. The thermocouple was installed inside a hollow quartz tube and in principle should not have been a problem. However, it is possible for heavy metals to diffuse through

quartz [Schmidt, 1983] so that the thermocouple was removed to observe any effect it may have.

In the fourth experiment, the same procedure was used except the wafer cleaning process was slightly modified to try to improve the heavy metal gettering of the cleaning process (see section 3.5). Also, the dry oxidation temperature was reduced to 1000 C. The effect was to improve the device and wafer to wafer consistency, and the overall dark current. Because three 'parameters' were changed between experiments 3 and 4, only the lumped effect was observed and the role each individual change played not firmly established. It was decided to retain all three changes since the overall effect was beneficial.

The fifth experiment was designed to examine the effect of wet oxidation and post-metallization forming gas annealing (PMFGA; forming gas is a 95% N_2 + 5% H_2 mixture) since the dry oxidation MOS capacitor relaxation times had been successfully extended to the ten minute range for a five volt bucket. Four wafers were fabricated as summarized in Table 3.3-1 (expt. 5). All four were cleaned using the improved technique. Wafer T4 received a 45 min. wet oxidation at 1000 C which was then stripped. Then, T3 and T4 underwent a 45 min. wet oxidation at 1000 C which was also stripped. Finally, all four wafers were dry oxidized at 1000 C for 60 min. followed by an in-situ N_2 anneal. The devices were characterized immediately following metallization. Wafer T1 had a dark current of 10 nA/cm^2, wafer T3 had 480 nA/cm^2, and wafer T4 more than 8000 nA/cm^2. Wet oxidation clearly has an influential effect on dark current and the silicon truly remembers its processing history. Following a 10 minute PMFGA, the wafers' dark current improved to 3.4, 94 and 828 nA/cm^2 respectively. Adding an additional 30 minutes of PMFGA further improved the dark current to 0.6, 20 and 100 nA/cm^2 respectively.

The forming gas anneal significantly reduced the dark current. Wafer T4 recovered in a dramatic fashion from 8000 nA/cm^2 after 40 minutes of PMFGA at 420 C. Varker and Ravi (1974) saw similar effects in PN junction reverse currents following low temperature heat treatments (400 C), which they suggest are due to thermal relaxation of the elastic strain field associated with an oxidation induced stacking fault. They speculate that the strain field itself results in a locally enhanced generation rate, in addition to the contribution made by impurity precipitates and silicon defect.

Wafer T4 may have recovered in a different manner than Ravi-Varker strain field relaxation model. Forming gas decreases the number of interface traps [Deal, et al. 1969], so that one might speculate that wafer T4 originally had an excessive number of interface traps due to surface termination of oxidation induced stacking faults perhaps resulting in excess surface strain fields or excess dangling bonds. The long forming gas anneal may have reduced this interface trap generation source, and decreased the dark current. Such a hypothesis might be easily tested by comparative annealing in an inert atmosphere (e.g. PMNA) with a non-aluminum gate. Careful measurement of (near mid-gap) interface states should also be performed. Although the circumstances were not controlled as indicated above, long PMNA cycles in expt. 11 (Table 3.3-1) resulted in increased dark current rather than decreased dark current which, in the least, does not contradict the oxidation-induced stacking fault induced interface trap generation hypothesis. Support for this hypothesis is also obtained with wafer 02B of expt. 7. The wafer was tested after POFGA and after PMFGA. The POFGA decreased the dark current by a factor of 10, whereas the PMFGA improved it by a factor of 40. The presence of a metal gate should only affect the interface, not bulk strain fields. A careful study should be made of this phenomenon to make any substantial conclusions.

The seemingly miraculous healing powers of post-metallization forming gas annealing is unfortunately not a panacea for dark current. For some devices, PMFGA has virtually no effect on the capacitor relaxation time indicating that the dark current is most likely dominated by lattice defects and impurities in the depleted bulk. More significantly, extended PMFGA may be incompatible with the fabrication of tunnel oxide devices. Dressendorfer (1978) indicates that high quality tunnel oxides may survive a ten minute PMFGA. However, Teng (1983) found that a PMFGA of more than three minutes degrades the MTOS junction performance. In the present work it was found that MTOS junctions can tolerate at least 60 minutes of PMNA (400 C) and at least 10 minutes of PMFGA with only improved device performance. This will be elaborated upon later in this chapter.

In the sixth experiment, the effect of using a dry-wet-dry oxidation process was investigated following a suggestion from T-P. Ma. Three wafers were tested as summarized in Table 3.3-1 (expt. 6). Wafer W4 was dry oxidized at 1000 C for 90 minutes. Wafer W5 was oxidized and annealed at 1000 C in a dry(25 min.), wet (45 min.), dry (25 min.) and N_2 (20 min.) sequence. Wafer W6 was given a 45 min. wet oxidation. All three wafers were stripped of their oxide and then dry oxidized at 1000 C for 90 min. followed by an in-situ N_2 anneal for 30 min. The three wafers were given a 60 min. PMFGA. The dark currents for wafers W4, W5 and W6 were 90, 300, and 20 nA/cm^2 respectively. The conclusion was that the dry-wet-dry process did not improve dark current, and if it had any effect, it was to make the dark current worse.

It is suggested by Nicollian and Brews (1982, pg. 721) that among many techniques to decrease oxidation induced stacking faults, high temperature wet oxidation works well. In the seventh experiment, the effect of wet oxidation temperature was investigated. As indicated in

Table 3.3-1 (expt. 7), four wafers were wet oxidized at three different periods and temperatures yielding roughly equal oxide thicknesses. The oxide was subsequently stripped and a dry oxide grown at 1100 C for 35 min. One of the wafers was given a low temperature post-oxidation (pre-metallization) FGA for 60 min. The dark current densities following a PMFGA clearly indicate that the low temperature (900 C) wet oxidation is preferable, but also that 1100 C is better than 1050 C. Also indicated is that post oxidation FGA improves the dark current.

In a quick experiment in collaboration with C-C. Wei, an RF plasma anneal [Ma and Ma, 1978; Chin 1981] was attempted at 500 watts for 20 min. The high power level was chosen due to previous success with RIE damaged wafers. Unfortunately, the RF plasma anneal annihilated any observable capacitor relaxation time and distorted the C-V characteristic, suggesting the plasma caused the creation of an enormous number of interface trap generation centers. Following these disastrous results, no further investigation into RF plasma annealing was performed. Perhaps a lower power setting might yield more satisfactory effects.

In the eighth experiment, four different oxides which might be suitable for diffusion masking and a field oxide were tested. In collaboration with T-C. Chen, a plasma enhanced chemical vapor deposition (PECVD) process was used to form a 4700 Å thick oxide. This oxide was densified at 1000 C for 90 min. in N_2. A spin-on SiO_2 film was applied to the second wafer and similarly densified. A dry oxide grown at 1000 C for 120 min. was used on the third wafer, and a wet oxide grown at 900 C for 120 min. was used on the fourth wafer. The oxide mask layer was stripped on all four wafers and a dry oxide was grown at 1000 C for 120 min. The devices were characterized following a 60 min. PMFGA. The PECVD wafer showed a wide range of dark currents

from 35 nA/cm^2 (very best) to well over 2000 nA/cm^2. These devices also exhibited poor C-V characteristics indicating bulk damage or contamination by the plasma. The spin-on wafer yielded consistently poor results at over 2000 nA/cm^2 indicating probable bulk contamination during the densification process. The dry oxide wafer gave an almost reasonable dark current level of 23 nA/cm^2 but the low temperature (900 C) wet oxide wafer gave the best results at about 10 nA/cm^2.

Having established a promising wet oxidation process, the effect of dopant diffusion was examined in the ninth experiment. The basic MOSFET process was employed and for two wafers, the channel stop diffusion was included. The first wafer (X1) was subjected to all masking oxidations and gate oxidations but the actual spin-on/drive-in process was omitted. The second wafer (X2) was processed as was X1, but a borosilicate film was spun-on and driven-in. The third and fourth (X3 and X4) wafers, like X2, were subjected to the p^+ drive-in but were also given the n^+ channel stop diffusion. The first three wafers were stripped of their oxide layers and a new field oxide grown at 900 C for 120 mn. Gate oxidation windows were opened in all four wafers and dry oxidation at 1000 C for 95 min. and in-situ N_2 annealing cycles were carried out. All four wafers were given a 60 min. pre-metallization FGA at 420 C, and a 10 min. PMNA. The results summarized in Table 3.3-1 (expt. 9) show that following the p^+ diffusion, the dark current increased significantly, but was reduced after the channel stop n^+ diffusion. However, regrowing the field oxide resulted in increased dark current.

A probable explanation involves the temperature dependent solid solubility of impurity contaminants in heavily doped silicon, moderately doped silicon, and in silicon defect elastic strain fields. The gettering property of heavily phosphorous doped silicon was noted by Goetzberger and Shockley (1960) and studied by Lambert and Reese (1968).

Impurities were probably introduced into the silicon by the borosilicate and phosphorosilicate spin-on films. Due to the differentially higher solid solubility of the n^+ regions, the impurities were gettered during the n^+ diffusion process. During the 120 min. wet oxidation subsequent to the n^+ drive-in, the impurities were drawn out of the n^+ diffused regions and perhaps collected by higher solid solubility elastic strain field zones caused by the wet oxidation, and could not be re-gettered during the gate oxidation process. Thus, avoiding subsequent wet oxidation after the n^+ drive-in yields significantly lower dark current. This practice was adopted in the Yale 8205 process.

A satisfactory lot of wafers was fabricated (lot 7) and as indicated by the tenth experiment of Table 3.3-1, a wafer which underwent the MOSFET portion of the Yale 8205 process and channel stop diffusion had excellent characteristics. The MOS dark current was 1.4 nA corresponding to a seven minute capacitor relaxation time (5.0 volt bucket) and the PN junction reverse current density (at -5.0 volts) was 15 nA/cm^2.

The PN junction reverse current is another measure of semiconductor quality, though it also includes the quality of the diffusion itself. Following an exasperating series of experiments which showed that the borosilicate film spin-on process was unreliable, a borofilm dopant source was employed. In the final lot of devices made (lot 9), the PN junction reverse current density (measured at -5.0 volts) was 1.5 nA/cm^2 following the p^+ drive-in. After n^+ diffusion and gate oxidation it increased to 7.7 nA/cm^2, and following tunnel oxidation, it rose to 37 nA/cm^2.

The MOS dark current in wafer 9A was 11 nA/cm^2 following a 10 min. PBNA. For increased PBNA cycle times, the dark current increased.

However, when a 10 min. PMFGA was subsequently performed, the dark current improved to its initial low level or better. For the lot 9 packaged devices, only the initial 10 min. PMNA was performed to insure the integrity of the MTOS junctions.

Summary of Dark Current Reduction Findings

It was observed that the reduction of wet oxidation temperature had a major impact on the dark current. Such an effect is ascribed to the reduction of oxidation rate and subsequent reduction of oxidation induced stacking fault nucleation due to precipitation of local excess of diffused oxygen [Ravi and Varker, 1974].

Forming gas atmosphere low temperature (420 C) heat treatments also contributed to the reduction of dark current. Besides the normal effect of PMFGA [Deal, et al. 1969], the heat treatment may also relax elastic strain in the silicon lattice due to defects and shrink strain field induced excess generation zones in the silicon [Varker and Ravi, 1974]. Post-gate oxidation/pre-metallization forming gas annealing (420 C) was also found to reduce dark current, and is more compatible with tunnel oxide device fabrication.

Phosphorous doped silicon channel stop n^+ diffusions also significantly enhanced dark current reduction. This effect is due to the impurity gettering properties of heavily doped silicon through enhanced contaminant solid solubility [Lambert and Reese, 1968]. The gettering effect was observed to depend on the processing sequence, and seemed to work best when the n^+ diffusion was the final high temperature (> 800 C) processing step. It is expected that heavily doping the wafer's backside will assist in the gettering action, but two attempts to try this resulted in a severe autodoping problem during subsequent processing.

To a lesser extent, it was observed that a slightly improved wafer cleaning process showed a positive effect on the dark current. Such wafer cleaning helps remove metallic impurities from the wafer surface prior to high temperature processing.

In future studies, it is hoped that the development of an intrinsic gettering process [Swaroop, 1984] might help control dark current. Success might also be achieved with a combination of the techniques referred to by Nicollian and Brews (1982) such as HCl gettering [Young and Osburn, 1983], but it seems that dark current reduction is a processing art which varies from laboratory to laboratory. One commercial manufacturer revealed that dark current in the 10 nA/cm^2 range was achieved only on good days. Thus, the suppression of dark current from 3000 nA/cm^2 on wafer 2A to 11 nA/cm^2 on wafer 9A through process development in the Yale facility can be considered satisfactory.

Table 3.3-1

Summary of Dark Current Reduction Experiments

Expt.	ID	Wafer Type*	Process	Typ. Dark Current (nA/cm^2)
--	2A	M	Non-optimized 8205	> 3000
1	M0A	M	Dry 1050-60; N_2 1050-30; PMNA 420-15	120
	M0B	M	(p) Dry 1050-60; N_2 1050-30; PMNA 420-15	20
	M0C	M	(p) Wet 1100; Strip; Dry 1050-60; N_2 1050-30; PMNA 420-15	> 2000
2	M1A	M	Dry 1050-60 (in dry ox tube); N_2 1050-30; PMNA 420-15	1.5
	M1B	M	Dry 1050-60 (in wet ox tube); N_2 1050-30; PMNA 420-15	3
	M2A	SM	Dry 1050-60 (in dry ox tube); N_2 1050-30; PMNA 420-15	30
	M2B	SM	Dry 1050-60 (in wet ox tube); N_2 1050-30; PMNA 420-15	15
3	M3A	M	Dry 1050-60 (in dry ox tube); N_2 1050-30; PMNA 420-15	60
	M3B	M	Dry 1050-60 (in wet ox tube); N_2 1050-30; PMNA 420-15	4
	M4A	SM	Dry 1050-60 (in dry ox tube); N_2 1050-30; PMNA 420-15	5
	M4B	SM	Dry 1050-60 (in wet ox tube); N_2 1050-30; PMNA 420-15	2
4	M5A	M	Dry 1000-60 (in dry ox tube); N_2 1000-30; PMNA 420-15	1
	M5B	M	Dry 1000-60 (in wet ox tube); N_2 1000-30; PMNA 420-15	6
	M6A	SM	Dry 1000-60 (in dry ox tube); N_2 1000-30; PMNA 420-15	2
	M6B	SM	Dry 1000-60 (in wet ox tube); N_2 1000-30; PMNA 420-15	2

Table 3.3-1 (cont.)

Expt.	Wafer ID	Type	Process	Typ. Dark Current (nA/cm^2)
5	W1	W	Dry 1000-60; N_2 1000-30;	15
			+ PMFGA 420-10	7
			+ PMFGA 420-30	2
	T1	SM	Dry 1000-60; N_2 1000-30	10
			+ PMFGA 420-10	3
			+ PMFGA 420-30	0.6
	T3	SM	Wet 1000-45; Strip; Dry 1000-60; N_2 1000-30;	480
			+ PMFGA 420-10	94
			+ PMFGA 420-30	20
	T4	SM	Wet 1000-45; Strip; Wet 1000-45; Strip; Dry 1000-60; N_2 1000-30;	> 4000
			+ PMFGA 420-10	800
			+ PMFGA 420-30	100
6	W4	W	Dry 1000-60; PMFGA 420-60	92
	W5	W	Dry 1000-20; Wet 1000-45; Dry 1000-20; N_2 1000-20; Strip; Dry 1000-60; PMFGA 420-60	300
	W6	W	Wet 1000-45; Strip; Dry 1000-60; PMFGA 420-60	22
7	01A	SM	Wet 900-120; Strip; Dry 1100-35; PMFGA 420-60	35
	01B	SM	Wet 1050-35; Strip; Dry 1100-35; PMFGA 420-60	75
	02A	SM	Wet 1100-24; Strip; Dry 1100-35; PMFGA 420-60	70
	02B	SM	Wet 1100-24; Strip; Dry 1100-35; POFGA 420-60 PMFGA 420-60	40

Table 3.3-1 (cont.)

Expt.	Wafer ID	Wafer Type	Process	Typ. Dark Current (nA/cm^2)
8	W8A	W	PECVD; N_2 1000-90; Strip; Dry 1000-120; PMFGA 420-60	35-2000
	W8B	W	Spin-on; N_2 1000-90; Strip; Dry 1000-120; PMFGA 420-60	> 2000
	W9A	W	Dry 1000-120; Strip; Dry 1000-120; PMFGA 420-60	23
	W9B	W	Wet 900-120; Strip; Dry 1000-120; PMFGA 420-60	11
9	X1	M	MOSFET but no diffusions; Regrow field ox; Dry 1000-95; N_2 1000-60; POFGA 420-60; PMNA 420-10	550
	X2	M	MOSFET; Regrow field ox; Dry 1000-95; N_2 1000-60; POFGA 420-60; PMNA 420-10	> 5000
	X3	M	MOSFET + channel stop; Regrow field ox; Dry 1000-95; N_2 1000-60; POFGA 420-60; PMNA 420-10	60
	X4	M	MOSFET + channel stop; Dry 1000-95; N_2 1000-60; POFGA 420-60; PMNA 420-10	1.8
10	7C	M	MOSFET + channel stop; Dry 1000-95; N_2 1000-60; POFGA 420-60; PMNA 420-10	1.4
11	9A	M	Yale 8205	11
			+ PMNA 420-15	23
			+ PMNA 420-35	45
			+ PMFGA 420-10	10

*Wafer Types: M = Monsanto C7 (100) 1-2 Ω-cm 2" diam (n-type)
SM = Semimetal C7 (100) 3-5 Ω-cm 2" diam (n-type)
W = Wacker F7 (100) 3-5 Ω-cm 2" diam (n-type)

3.4 Mask Making

Using the present facilities at Yale to manufacture a set of mask plates is an unpleasant process. Computer aided layout of masks with mask fabrication occuring with a modern E-beam or optical pattern generator is the preferred route to take in the future. Perhaps an intermediate solution might be to employ the new SEM to generate 10x reticles by exposing aluminum and photoresist coated glass plates in a manner similar to that recently demonstrated by M. Rooks of Yale University to write ultra-small patterns. In this section, the current recipe for making masks is presented for completeness, since change often comes slowly.

To aid in the mask making (and subsequent fabrication) process, the mask layout should include a border line encompassing the chip. Such a border aids in 10x reticle mounting and mask alignment. Each mask level should also include a series of fine structures (e.g. fine lines) to assist in focusing and determining photoresist exposures. Such a resolution pattern was found to be useful in the Yale 8205 mask set. Level identifications are also suggested. An alignment key or registration mark is required for each level. A simple cross suffices. The best alignment method seems to make a solid cross on the first level. The second level has a translucent cross of slightly larger dimensions (e.g. 0.2 mils on the final plate). Alignment is done by centering the fabricated cross in the slightly larger cross-shaped window. Alignment accuracies of better than $\pm$ 2 μm can be achieved with this simple technique if the 10x reticle rotation problem is solved. Note that the cross may be reused for the third level if the second level process doesn't obliterate it, otherwise a new cross must be made in an adjacent location by the second level.

The Yale 8205 process requires eight mask levels. After cutting each

rubylith (100x), they must be photographically reduced by a factor of ten, resulting in a 10x reticle. The rubylith is mounted on a light box and photographed at the highest possible f-number (largest depth of focus) after being focused at the lowest available f-number. Focusing was facilitated when ground glass system was modified during the course of the present work, so that 'what you see is what you get'. The exposed plate is developed as indicated by Table 3.4-1. The developer mixture ratios and development times may be varied depending on the outcome and exposure time.

Once the 10x reticles have been made, they must be mounted on frames for the step and repeat camera. Mounting is secured using Krazy Glue™ with the reticle's emulsion side facing into the frame. All reticles must be mounted with the same amount of rotation. In the present work, a technique was developed to assist in 'square' mounting. Two glass slides are glued to graph paper at right angles. The frame is snugly pressed against the glass slides. The 10x reticle is mounted on the frame by making the chip border (discussed above) square with the graph paper lines by eyeballing the set-up. To achieve $\pm$ 0.5 mil error or less on the final mask plates, the frame border must be parallel to the graph paper line within $\pm$ 0.1 mm. This challenge can be met by patience. Should the final result be undesirable, the reticle can be separated from the frame by soaking the pair in acetone for three hours using ultrasonic agitation. The acetone will not affect the reticle.

The correct use of the step and repeat camera requires similar patience and concentration. The utmost care must be taken to reproduce x and y steps between each run to within $\pm$0.5 mils. For the three barrel step and repeat camera at Yale, a minimum of three runs must be made to make the eight master mask plates. Development is as described in Table 3.4-1.

Table 3.4-1

HRP Development Process

Mixtures

Clearing Bath:	Dacomatic DN-3/DR-5: H_2O :: 1:3
Developer:	HRP Developer: H_2O :: 1:16
Fixer:	Kodak Rapid Fixer as is
50% Methanol:	Methanol: H_2O :: 1:1
75% Methanol:	Methanol: H_2O :: 3:1
100% Methanol:	Methanol

All baths 22°C - 24°C. Gently agitate plate in each bath.

Process

1.	Clearing bath to remove anti-halation backing	30-60 sec
2.	Rinse in running water	60 sec
3.	Developer	~ 300 sec
4.	Rinse in running water	60 sec
5.	Fixer	60 sec
6.	Rinse in running water	300 sec
7.	50% Methanol	15 sec
8.	75% Methanol	15 sec
9.	100% Methanol	15 sec
10.	Blow dry with N_2	

To copy a working mask plate from the master mask plate, an emulsion to emulsion contact exposure is used. Subsequent development results in a negative polarity mirror image of the master plate. To obtain a faithful copy of the master plate, the intermediate copy must be copied. However, to obtain a positive polarity mirror image, a reversal development process is used. The basic process is to develop the emulsion to emulsion exposed plate in a solution which does not dissolve unexposed emulsion, but turns exposed emulsions black. This blackened emulsion is then removed using a bleach bath followed by a clearing bath. The previously unexposed remaining emulsion is then given a blanket exposure and developed turning it black. This emulsion is then fixed and hardened in a fixer bath in the usual way. Thus, the initially unexposed emulsion shadowed by a black master plate emulsion is blackened, resulting in a positive polarity mirror image. The resolution of this process is nearly as good as the negative process and more than adequate for the Yale 8205 plates. The process is described in Table 3.4-2.

A photograph showing the completed (and mounted) 10x reticles is displayed in Fig. 3.4-1.

Table 3.4-2

Reversal Process

Mixtures

Dacomatic Clearing Bath	DN-3/DR-5: H_2O :: 1:3
D-8 Developer (Kodak)	Stock: H_2O :: 1:3
R-9 Bleach Bath	4 liters H_2O
	+ 38 grams Kodak Potassium Dichromate
	+ 48 mL H_2SO_4
CB-6 Clearing Bath	4 liters H_2O
	+ 60 grams Kodak Sodium Bisulfite
	+ 2 grams Calgon (water softener)
HRP Developer (Kodak)	HRP Dev. : H_2O :: 1:8
Kodak Rapid Fixer	as is
50% Methanol	Methanol: H_2O :: 1:1
75% Methanol	Methanol: H_2O :: 3:1
100% Methanol	Methanol

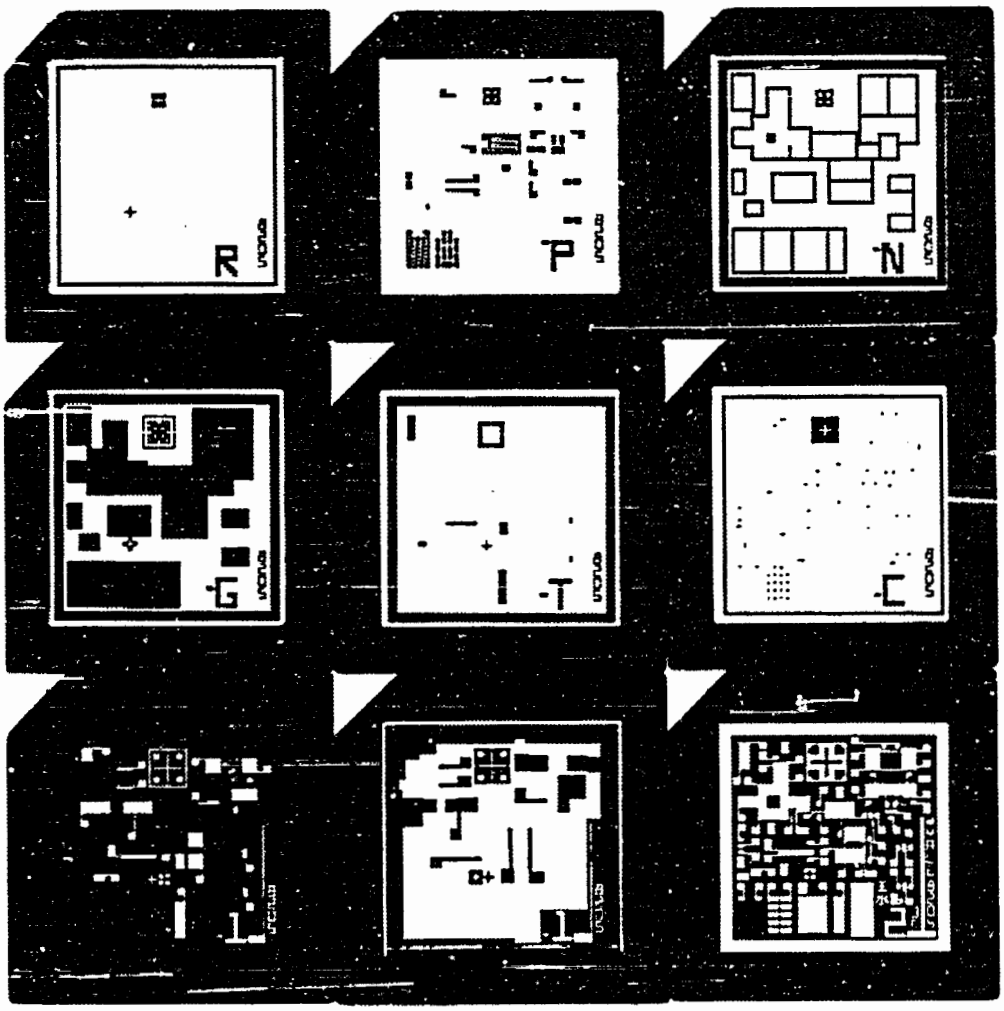
Baths 22°C - 24°C. Gently agitate plates in each bath.

Process

1.	Clearing bath Dacomatic	30-60 sec
2.	Rinse in running water	60 sec
3.	Develop in D-8 Developer until black	~ 120 sec
4.	Rinse in running water	60 sec
5.	Bleach in R-9 bath until transparent	~ 120 sec
6.	Rinse in running water	60 sec
7.	Clearing bath CB-6	180 sec
8.	Rinse in running water	60 sec
9.	Presoak in HRP developer	20 sec
10.	Rinse in running water	60 sec
11.	Blanket exposure	~ 120 sec
12.	Develop in HRP developer until black	60-120 sec

Table 3.4-2 (cont.)

13. Rinse in running water	30 sec
14. Fix in Rapid Fixer	120 sec
15. Rinse in running water	300 sec
16. Dehydrate in 50% Methanol	15 sec
17. Dehydrate in 75% Methanol	15 sec
18. Dehydrate in 100% Methanol	15 sec
19. Blow dry with N_2	



3.4-1 10X reticles for Yale 8205 chip. Level IDs in lower right-hand corner are as follows:

R:	Registration	C:	Contact vias
P:	Boron diffusion	1:	First level metal
N:	Channel stop diffusion	1:	Improved first level metal
G:	Gate oxide windows	2:	Second level metal
T:	Tunnel oxide windows		

(Note: Polarity reversal from final mask plates)

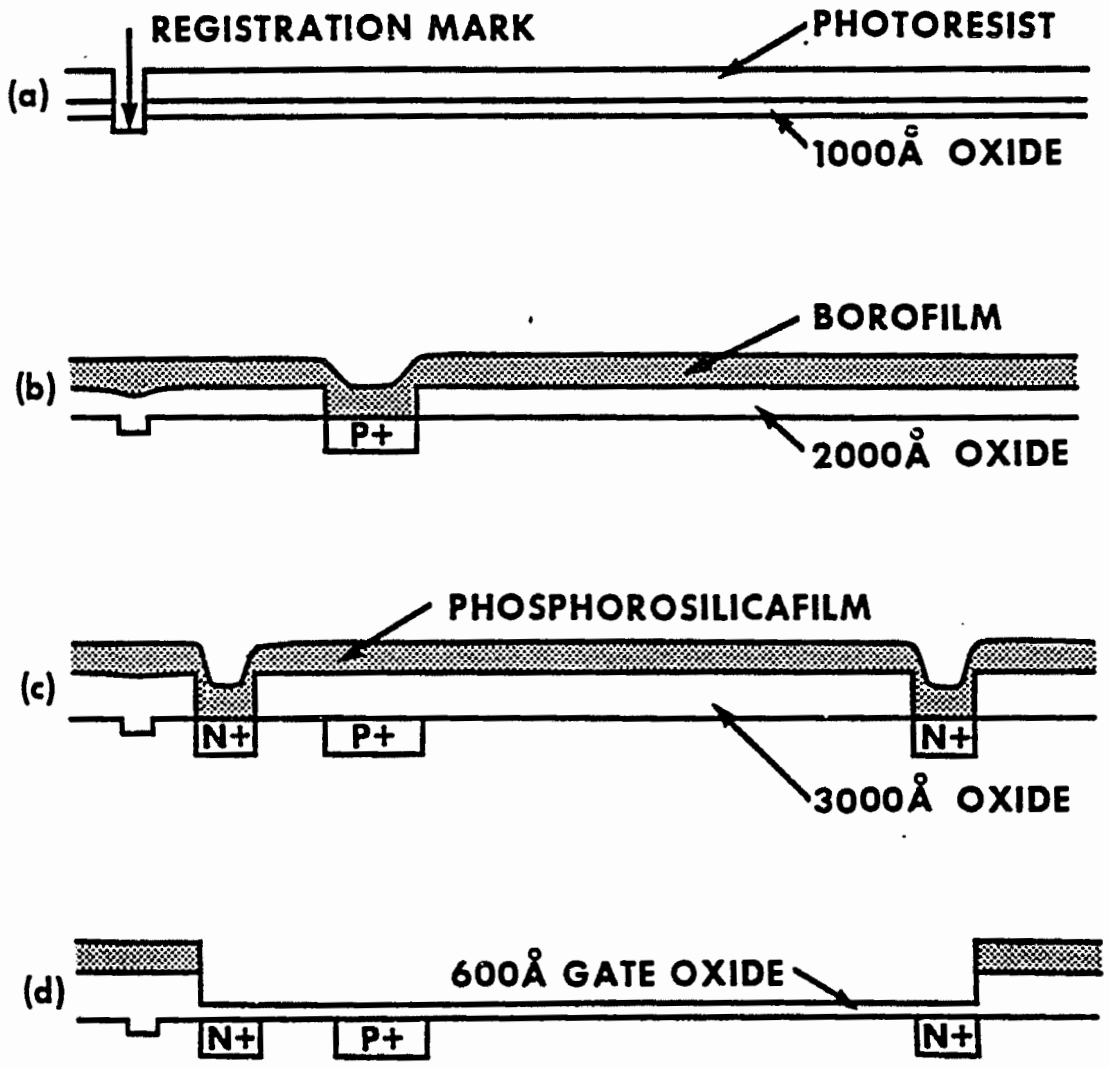
3.5 Fabrication Process

The Yale 8205 process is quite similar to the processes used by past and present investigators, and only departs significantly by the reduction of wet oxidation temperature to 900 C, the use of post-gate-oxidation forming gas annealing (420 C), the formation of source-drain contacts, and of course in the fabrication of sub-micron inter-electrode gaps. The process is interlocked beyond what is revealed by a casual inspection. Slight changes in parameters or sequences can dramatically affect device behavior and yield. The best advice offered in this thesis is to not 'mess' with the recipe.

A series of device cross-sections illustrating the basic fabrication sequence is shown in Figs. 3.5-1(a-h). The device fabricated in this sequence is the charge packet initiated MTOS junction switching experiment described in detail in chapter 4 and illustrated in Fig. 4.3-3.

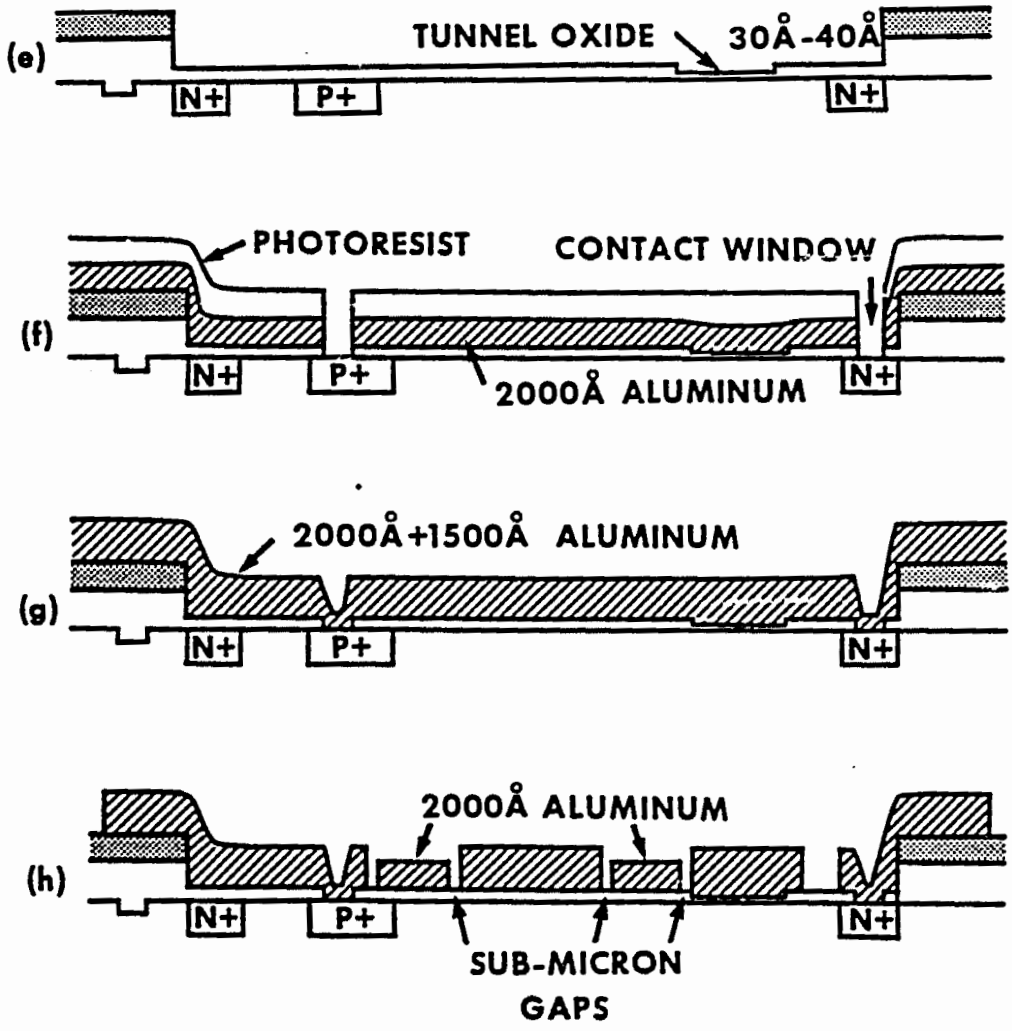
The wafers are selected and cleaned. Best results have so far been obtained with Monsanto n-type wafers (100) 2'' in diameter with 2·Ω cm resistivity (CZ). In retrospect. more lightly doped wafers may have been preferable (10-20 Ω·cm). Attempts to use Wacker n-type wafers (100) 2'' in diameter with 5 Ω·cm resistivity (float zone) were disappointing. The dark current was usually worse with these wafers and the MTOS junctions seldom worked. The latter may be related to the quality of surface polishing, which was done for the Wacker wafers by a small business in Massachusetts.

A wet oxidation is used to grow a 1000 Å oxide mask for the silicon etching of registration marks (R-level). Since photoresist does not adhere well to SiO_2 exposed to Si etch, the SiO_2 is stripped and a 2000 Å wet oxide regrown. This oxide serves as the borofilm drive-in mask



3.5-1 Yale 8205 process summary cross-sections

a. Registration marks
b. Borofilm diode formation
c. Phosphorosilicafilm channel stop formation
d. Gate oxide window and growth
e. Tunnel oxide window and growth
f. Contact via etch
g. First level metal
h. Final cross-section


(e)
TUNNEL OXIDE
30Å-40Å
N+
P+
N+
(f)
PHOTORESIST
CONTACT WINDOW
N+
P+
2000Å ALUMINUM
N+
(g)
2000Å+1500Å ALUMINUM
N+
P+
N+
(h)
2000Å ALUMINUM
N+
P+
N+
SUB-MICRON
GAPS

(P-level). The borofilm (a boron doped polymer) and oxide mask are stripped using 10% HF.

A final wet oxide is grown to a thickness of 3000 Å to serve as a phosphorosilica film drive-in mask and field oxide (N-level). Gate oxide windows are opened in the sandwich structure and the gate oxide grown. No autodoping was observed during the gate oxide growth from adjacent phosphorosilicafilm. The gate oxide is given a post-oxidation, low temperature heat treatment (420 C) in a forming gas atmosphere. Although not as effective as a post-metallization forming gas anneal (which is incompatible with tunnel oxides for long periods of time), it does reduce dark current by an order of magnitude.

Tunnel oxidation windows are opened in the gate oxide and the tunnel oxide is grown as described by Dressendorfer (1978), Lai (1979) and Teng (1983). The oxidation is at 800 C for 15 minutes followed by an in-situ anneal of 15 minutes. The final cleaning of the windows prior to tunnel oxide growth was found to have a significant influence on tunnel oxide thickness and quality. A few drops of HF in the final H_2O rinse bath (preferably chilled) seems to yield the best results [T-C. Chen, private communication]. It was also observed that wafers which had undergone significant processing prior to tunnel oxide growth had thinner tunnel oxides than near virgin wafers oxidized for the same amount of time. Aluminum is immediately evaporated on the tunnel oxide to a thickness of 2000 Å.

Contact windows are opened to p^+ and n^+ regions by a double etch process. First, aluminum etch is used to etch through the 2000 Å metal layer. Then, oxide etchant is used to etch through the gate oxide layer. Unfortunately, BOE attacks aluminum so that unless the process is carefully controlled, the aluminum surrounding the contact window is

bevelled over several microns. Although actually a side benefit assisting step coverage in the Yale 8205 process, it looks ugly and cannot be used in smaller geometries. Other oxide etch techniques (i.e. gas phase) may be more fruitful in constrained layouts. The contact made by this new double-etch technique has relatively low resistance, and for example, allows the fabrication of high quality tunnel oxide MOSFETs.

Once the contact windows have been opened and the photoresist removed, aluminum (1500 Å) is evaporated. The first level metal mask is used to form the photoresist overhang for the subsequent aluminum (shadow) evaporation. This 2000 Å aluminum layer and the previous 3500 Å of aluminum are patterned following lift-off using the second level metal mask. Backside evaporation of aluminum is followed by a 10 minute post-metallization N_2 anneal.

Following tunnel oxidation, it has been found that the use of ultrasonic agitation degrades tunnel oxide durability. Teng (1983) used ultrasonic agitation and observed MTOS junction degradation for annealing periods exceeding a few (i.e. 3) minutes. This was confirmed by more recent experiments in which ultrasonic agitation was inadvertently used [C-C. Wei, private communication]. It was found in this work that without ultrasonic agitation, low temperature (400 C) post-metallization N_2 anneals could be performed for at least 60 minutes, and PMFGA for at least 10 minutes. The MTOS junction performance as measured by OFF state current (ambiguous as discussed in chapter 4) improved steadily (lower current).

To package the chips, they are coated with a protective layer of photoresist and carefully scribed and split apart. The resist is removed in an acetone bath. The chips are mounted in packages using a

conductive epoxy which is cured by an air bake at 150 C for 60 min. Bonding is done with 1 mil gold wire and an ultrasonic bonder set at a low power and time setting. The chip/package is heated to 150 C as a ball bond is formed on the chip bonding pad and a wedge bond on the package. The package leads are trimmed if necessary. The Yale 8205 chips were packaged in both 24 pin dual-in-line packages (DIPs) and in 44 pin leadless carriers. The MTOS packaging yield was 100% as determined by measuring device characteristics before and after packaging.

The Yale 8205 process is given in recipe form in Table 3.5-1 and a photograph of a completed device after packaging in a 44 pin leadless carrier is shown in Fig. 3.5-2.

Table 3.5-1

Yale 8205 Fabrication Process

1. Wafer Initialization

   A. Select wafers

      - best results have been obtained with Monsanto n-type (100) 2" diam. 1-2 Ω-cm one side polished

   B. Scribe IDs into backside

      - place front side on texwipe and use diamond scribe on backside near edge

   C. Blow off wafer with N_2 to remove particles

2. Wafer Cleaning

   A. Preheat TCE and acetone to ~ 50°C in ultrasonic agitator. Cover to keep from evaporating.

   B. TCE ultrasonic > 5 min

   C. Acetone ultrasonic > 2 min

   D. DI rinse > 1 min

   E. Methanol ultrasonic > 5 min

      - prepare "super-mickey" ingredients
        40 mL H_2O_2 into SM beaker
        120 mL H_2SO_4 into grad. cyl.

   F. DI rinse > 2 min

      - mix SM; pour H_2SO_4 into SM beaker slowly

   G. SM ultrasonic 10 min

   H. DI rinse > 2 min

   I. 10% HF ~ 1 min (49% HF: H_2O :: 1:1)

   J. Blow dry w/N_2 out of 10% HF - should be hydroscopic if clean

   K. Transport to furnace in clean petri dish

Table 3.5-1 (cont.)

3. Initial wet oxidation (900°C, 40 min)
   (This is the standard furnace procedure with a slow pull)
   A. Load wafers into furnace
      i. Pull out hot boat with cool rod
      ii. Blow dust off wafer with N_2
      iii. Load wafers onto boat facing mouth
      iv. Push in 1-2 in/sec
      v. Put on end cap
   B. Oxidize 40 min.
   C. Unload wafers
      i. Pull boat to furnace mouth 1 in/sec
      ii. Replace end cap
      iii. Wait 5 min.
      iv. Pull out boat
      v. Blow wafers cool with N_2
      vi. Place in clean petri dish

4. Photolithography R-level
   (This is the standard photolithography procedure)
   A. Prepare AZ1370 photoresist (room temp.)
   B. Blow dust off pipette and bulb
   C. Mix MF312 developer: H_2O :: 1:1 (room temp.)
   D. Spin on AZ1370 3000 RPM 60 secs
   E. Bake 90°C in clean air 10 min.
   F. Align and expose ~ 65 secs
   G. Develop until developed
      - one-at-a-time ~ 40 secs
   H. DI rinse > 2 min
   I. Spin dry ~ 1 min
   J. Inspect and check alignment
   K. Bake 90°C in clean air 10 min

5. BOE etch SiO_2 until clear + 1 min ~ 3 min

Table 3.5-1 (cont.)

6.	Strip PR (This is the standard photoresist strip process)	
	A. DI rinse	> 2 min
	B. Acetone (ultrasonic)	> 3 min
	C. Fresh acetone (ultrasonic)	> 2 min
	D. DI rinse	> 1 min
	E. Methanol (ultrasonic)	> 5 min
	F. Blow dry with N_2	
7.	Silicon Etch	
	A. Mixed etchant - 70% HNO_3: 49% HF: 99.9% Acetic :: 3:1:1 - etch individually	~ 1 sec
	B. DI rinse immediately	> 3 min
8.	BOE strip SiO_2	5 min
9.	DI rinse	> 2 min
10.	Methanol (ultrasonic)	> 1 min
11.	Blow dry with N_2 and transport to furnace	
12.	Wet oxidation, 900°C 90 min slow pull	> 2000 Å
13.	Photolithography P-level	
14.	BOE etch SiO_2 until clear + 1 min	~ 5 min
15.	Strip PR	
16.	Nitric Acid (70% HNO_3) -need hydrophilic surface	10 min
17.	DI rinse	> 2 min
18.	Methanol (ultrasonic)	> 1 min

Table 3.5-1 (cont.).

19.	Blow dry with N_2	
20.	Bake out 90°C in clean air	> 5 min
21.	Spin on Borofilm 100, 3000 RPM	60 secs
22.	Bake 90°C	> 5 min
23.	Bake 150°C - 200°C	> 5 min
24.	Drive in p^+ 1050°C, N_2 @ 1.6 slpm, O_2 @ 0.07 slpm	
	A. Load into boat face-to-face, back-to-back	
	B. Load boat in furnace mouth, put on end cap	
	C. Prebake	10 min
	D. Push into furnace 1 in/sec	
	E. Replace end cap, drive-in	75-90 min
	F. Pull to mouth to cool	10 min
	G. Unload	
25.	Strip oxide and borofilm (10% HF) until clear	~ 8 min
26.	Wet oxidation, 900°C 3 hours, slow pull	> 3000 Å
27.	Photolithography N-level	
28.	BOE etch SiO_2 until clear + 1 min - record time	~ 7 min
29.	Strip PR	
30.	Bakeout 90°C	> 5 min
31.	Spin on Phosphorosilica film 3000 RPM - "$C_o \approx 3 \times 10^{20}/cm^3$"	60 secs
32.	Bake 90°C in air	> 5 min

Table 3.5-1 (cont.)

33. Bake 150°C - 200°C in air > 5 min

34. Drive-in n^+ 1000°C N_2 @ 2.0 slpm, O_2 @ 0.2 slpm

    A. Load wafers close packed front-to-back

    B. Proceed as for p^+ drive-in (step 24)

35. Photolithography G-level

36. BOE etch SiO_2 until kerf clears + time from step 28 + 1 min

37. Strip PR

38. Dry O_2 should flow in furnace > 60 min

39. Dry oxidation 1000°C O_2 @ 2.0 slpm 90 min

40. N_2 in-situ anneal cool down

    A. Switch to N_2 @ 2.0 slpm

    B. Reset furnace temperature to 800°C

    C. Let cool down 60 min

    D. Slow pull

41. Post-oxidation FGA 420°C
    (5% H_2, 95% N_2) 60 min

42. Photolithography T-level

43. BOE etch SiO_2 until clear + 1 min
    record time ~ 3 min

44. Strip PR

45. Dip etch in chilled H_2O with a drop
    or two of HF 10 sec

46. Blow dry with N_2 (or spin dry) and transport to furnace

47. Dry oxidation 800°C O_2 @ 2.0 slpm 10-15 min

Table 3.5-1 (cont.)

47. cont.

A. Pull hot boat to mouth

B. Load wafer face up at mouth

C. Push in and replace end cap

48. Switch to N_2 @ 2.0 slpm in-situ anneal — 15 min

49. Evaporate 2000 Å Al @ 10 Å/sec

50. Photolithography C-level

51. Al etch until clear + 1 min — ~ 5 min

52. DI rinse — > 2 min

53. Spin dry

54. BOE etch SiO_2 (use time from step 43) — ~ 3 min

55. DI rinse — > 2 min

56. Strip PR (no ultrasonic)

57. Evaporate 1500 Å Al @ 10 Å/sec

58. Photolithography M1-level - align carefully

59. Al etch until clear $\pm$ 3 min overetch — ~ 10 min

60. DI rinse — > 2 min

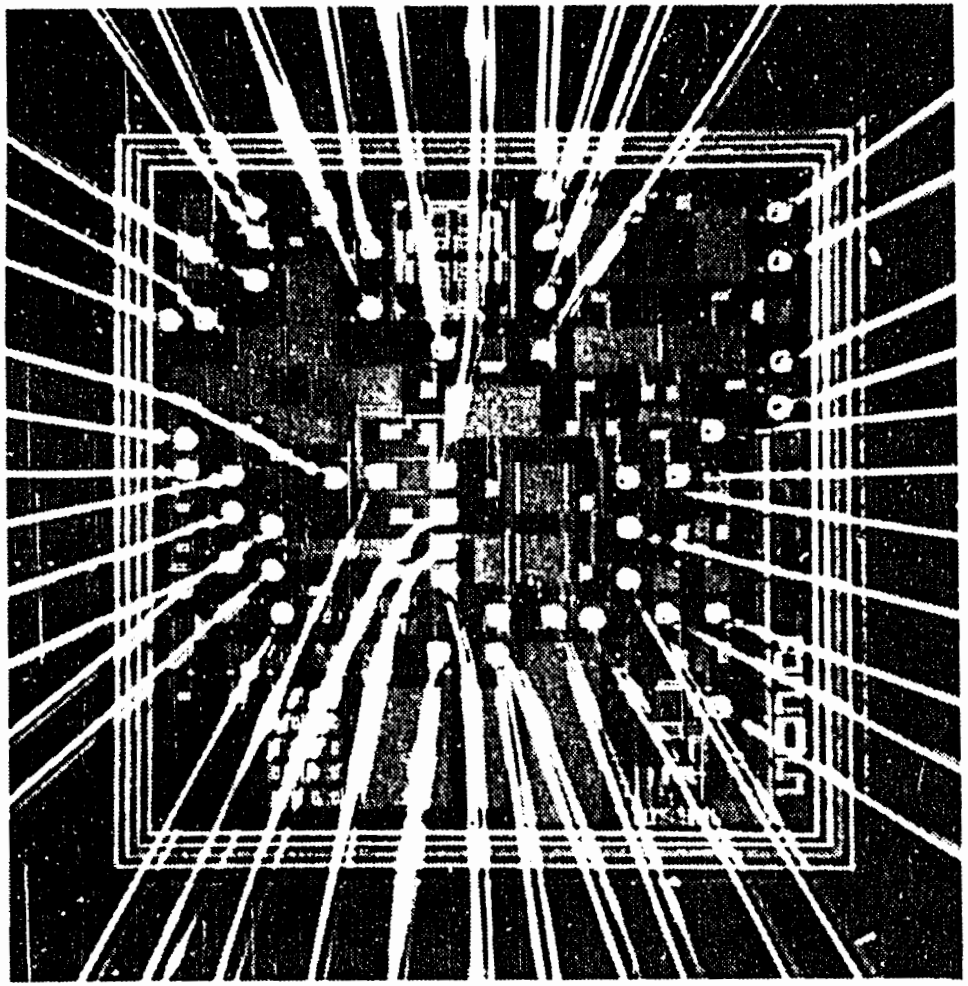
61. Spin dry

62. Evaporate 2000 Å Al @ 10 Å/sec
-use point-like source

63. Lift-off in acetone (no ultrasonic) — 60 min

Table 3.5-1 (cont.)

64.	DI rinse	>	2 min
65.	Spin dry		
66.	Photolithography M2-level		
67.	Al etch until clear + 1 min - don't stop when 2000 Å Al clears	~	8 min
68.	Strip PR (no ultrasonic)		
69.	Backside swab etch BOE		
70.	DI rinse	>	2 min
71.	Methanol (no ultrasonic)	>	1 min
72.	Blow dry with N_2		
73.	Backside Al evaporation	~	3000 Å
74.	PMNA 400°C		10 min
75.	Measure and weep		



3.5-2 Photograph of bonded chip
(N. Ferguson, Hughes Aircraft Co.)

3.6 Process Verification

In this section, the completely fabricated device characteristics are reported as measured using the process verification structures. The data is reported for wafers 7B and 9A which were those diced and packaged and used in the subsequent experimental testing of the prototype circuits.

The p^+ boron diffusion process can be checked using diffused resistor and contact string and evaluated in terms of the diffused layer sheet resistance and PN junction leakage current. The sheet resistance r_D and contact via (1 mil x 2 mils) resistance r_C can be determined from the diffused resistor resistance R_D and contact string resistance R_C in combination with the layout geometry. From the layout,

$$R_D = 172\ r_D + 2\ r_C \qquad 3.6\text{-}1$$

and

$$R_C = 9\ r_D + 18\ r_C \qquad 3.6\text{-}2$$

so that

$$r_D = R_D/171 - R_C/1539 \qquad 3.6\text{-}3$$

and

$$r_C = 9.55\ R_C/171 - R_D/342 \qquad 3.6\text{-}4$$

For wafer 7B, a defective dopant source resulted in r_D = 2.1 k Ω/□ and r_c = 450 Ω. In the improved wafer 9A process, the sheet resistance was reduced to 16 Ω/□ and r_C to 4 Ω. The PN junction leakage current measured at -5.0 volts was approximately 40 nA/cm^2, though wafer 7B diodes suffered from low voltage soft breakdown above 10 to 12 volts reverse bias, undoubtedly attributable to the defective dopant source. The diodes fabricated show a two to three order of magnitude improvement in leakage current over those fabricated by Teng (1983), attributable to the reduction in bulk traps achieved by the improved process.

The n^+ phosphorous diffusion yielded sheet resistances well under 1 Ω/□. The field oxide sandwich was in the 5000 Å to 6000 Å thickness

range. This gives a bonding pad stray capacitance close to 1 pF. The gate oxide was 590 Å thick and 550 Å thick for 7B and 9A devices respectively as determined using the accumulated gate oxide capacitance corrected for stray capacitance. The gate oxide interface trap density measured close to inversion using the high/low frequency capacitance technique of Wagner and Berglund (1972) was in the 2-3 x $10^{10}/cm^2$ eV range for both wafers. In that method,

$$D_{it} = \frac{C_{LF} - C_{HF}}{qA} \left(1 - \frac{C_{LF}}{C_{ox}}\right)^{-1} \left(1 - \frac{C_{HF}}{C_{ox}}\right) \qquad 3.6\text{-}5$$

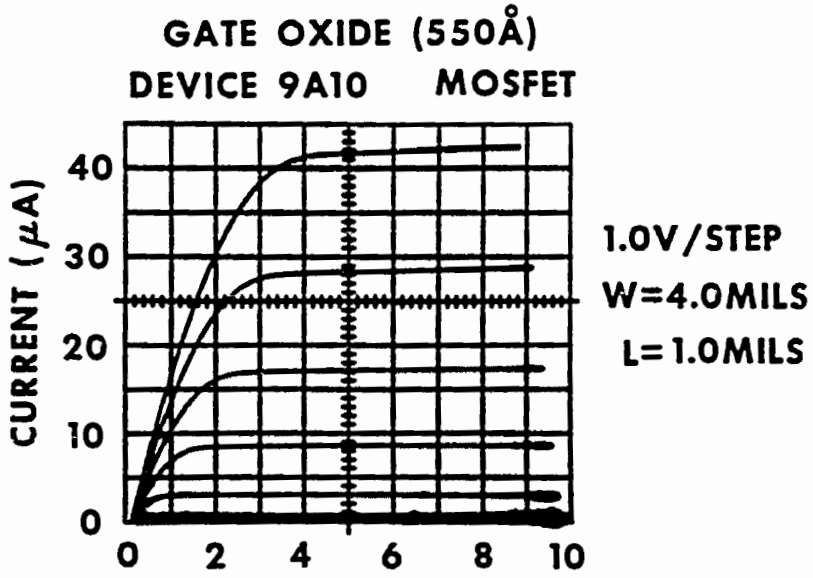
where C_{LF} is the low frequency capacitance (quasi-static) at some voltage, and C_{HF} the high frequency (1 MHz) capacitance at the same voltage. No hysteresis was observed in the room temperature HF C-V characteristic, though it is expected that there were unactivated mobile ions present due to the aluminum evaporation process. Fortunately, almost all biases used experimentally were negative, thus retaining positive mobile ions (i.e. N_a^+) at the Al-SiO_2 interface.

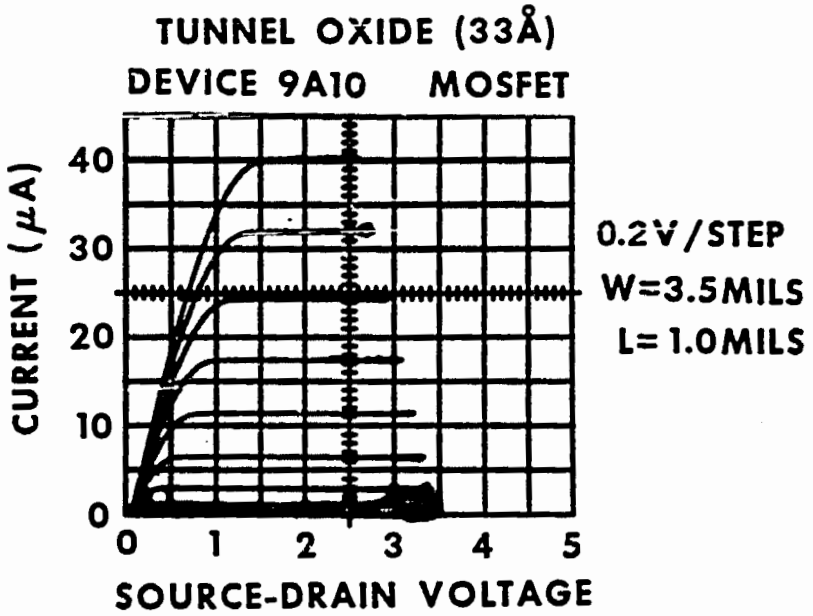
The typical breakdown strength of the gate oxide for wafer 7B was 7 x 10^6 V/cm). Anomalous low voltage breakdown (1-2 x 10^6 V/cm) of the wafer 9A gate oxide was observed in many parts, and was the major cause of low operational yield on this wafer. This breakdown was not systematically studied but may be related to the PMFGA of wafer 9A following gate oxidation, metallization and electrode definition prior to tunnel oxidation, etc. Such an inserted sequence was designed to preview device behavior prior to investing unwarranted processing time for tunnel oxidation and sub-micron gap formation. Such a preview process sequence was not performed on wafer 7B. Post-oxidation FGA prior to tunnel oxide window definition has not exhibited any problem (lot X) so that PMFGA in the preview sequence is most likely the culprit.

The tunnel oxide was determined to be in the 31 Å to 32 Å thickness range on wafer 7B, and in the 33 Å to 34 Å thickness range on wafer 9A as determined using a modified [Dressendorfer, 1978] C-V extrapolation technique [McNutt and Sah, 1975], with $\varepsilon_{ox} = 3.9$. The flat-band voltage was -0.19 volts $\pm$ 0.02 volts where the uncertainty is due to uncertainties in oxide thickness and stray capacitance values. The OFF state current (for wafer 9A) biased at -3.5 volts was typically 3.5 $\mu A/cm^2$, although this current can be reduced by extending the PMNA cycle. The ON state current was typically 100 times larger at this bias, and increases exponentially (see chapter 4) with bias voltage up to a breakdown voltage of -4.1 volts at a current density of 2.5 mA/cm^2. Operating the device in this range is a risky proposition since breakdown is a costly event with only two dozen devices packaged. The ON state holding voltage was -3.2 volts so that most devices were operated at -3.5 volts.

The MTOS durability test site consists of two adjacent MTOS junctions (40 $mils^2$ each) separated by a sub-micron gap. One junction is formed by etching off the aluminum covering the tunnel oxide during the overetch/ undercut process, and then recoating the tunnel oxide with aluminum. The MTOS junction survived the etch/recoat process without significant change in characteristics. For example, the OFF current in the recoated junction was 3.1 $\mu A/cm^2$ compared to 2.1 $\mu A/cm^2$ in the adjacent junction (-3.5 volts). The ON current (also at -3.5 volts) was 217 $\mu A/cm^2$ in the recoated junction and 201 $\mu A/cm^2$ in the adjacent device. The ability to fabricate such closely spaced MTOS junctions leads to intriguing opportunities for future investigation.

The gate oxide MOSFET test site on wafer 9A was measured and showed excellent MOSFET characteristics, as shown in Fig. 3.6-1a. The saturated MOSFET transfer characteristic was carefully measured over the





3.6-1 MOSFET characteristics

a. Gate oxide (550 Å)
b. Tunnel oxide (33 Å)

range zero to 0.5 mA. Plotting the square root of current as a function of gate voltage gave a very straight line indicating a threshold voltage of -2.1 volts. The slope of the line is related to an effective surface mobility μ_p' using a simple equation for saturation current I_s [Streetman, 1972]

$$I_s = (\mu_p' C_{ox} W/2L) (V_G - V_T)^2 \qquad 3.6\text{-}6$$

so that

$$\frac{d\sqrt{I_s}}{d V_G} = (\mu_p' C_{ox} W/2L)^{1/2} \qquad 3.6\text{-}7$$

Using $W = 4$ mils, $L = 1$ mil and $C_{ox} = 62 nF/cm^2$ yields an effective surface mobility of 146 cm^2/V sec.

This result can be used for reference in discussing similar measurements performed on the tunnel oxide MOSFET test site with similar geometry. The characteristics of this 33 Å oxide MOSFET can be seen to appear normal in Fig. 3.6-1b. The saturated MOSFET current was carefully measured over the same range of zero to 0.5 mA and the square root of the saturated current as a function of gate voltage plotted. A very straight line was obtained yielding a threshold voltage of -0.8 volts and effective mobility of 75 cm^2/V sec. This mobility reduction may be due to a higher inversion layer charge density. If the saturated channel charge is written as

$$Q_C = \frac{2WL}{3} C_{ox} (V_G - V_T) \qquad 3.6\text{-}8$$

then the ratio of channel charge for the tunnel oxide MOSFET to channel charge in the gate oxide MOSFET at equal saturation currents is

approximately

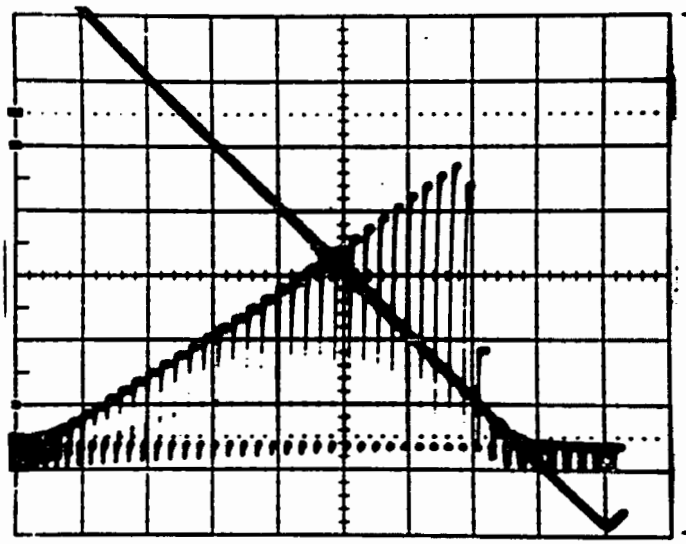
$$\frac{Q_C^T}{Q_C^G} = \left(\frac{d_{ox}^G}{d_{ox}^T}\right)^{1/2} \qquad 3.6\text{-}9$$

assuming equal mobilities and geometries. Thus, the tunnel oxide MOSFET has approximately 20 times higher channel charge. The mobility reduction may also be due to differences in interface trap density or may even be fundamental in nature, such as dimensionally enhanced surface scattering or silicon surface tensile stress reduction. Further studies to understand this phenomenon are presently underway.

The sub-micron gaps were fabricated with excellent yield on wafer 7B and less so on wafer 9A. However, in most cases, any residual short circuits between adjacent electrodes could be burned out by passing a few hundred miliamperes of current through the short, without observable damage to the devices. Typically the resistance between gaps exceeded 10^{12} Ω.

To test the basic functionality of the fabricated CCDs, the parallel channel of the MTOS junction charge packet initiated switching experiment (Fig. 4.3-2) was used to create a charge packet using the surface potential equilibration method, and then dump it to the charge packet output sense amplifier (appendix D). The output of the source-follower is shown as a function of metering well voltage in Fig. 3.6-2. This photograph demonstrates the functionality of the CCD structures fabricated on the Yale 8205 chip.

DEVICE 7B9 PARALLEL CHANNEL



R_L=500kΩ 2MSEC/DIV

V_{OA}=-12VOLTS

V_{DA}=-12VOLTS

V_{MW}=-8VOLTS

3.6-2 CCD input/output verification

CHAPTER 4

MTOS CHARGE PACKET THRESHOLD DETECTOR

4.1 Introduction

For the charge-coupled computer, one needs a method of comparing the magnitude of a quantity (represented as a charge packet) to a reference quantity. The MTOS bistable junction was investigated as a possible device for performing such an operation. As discussed in chapter 2, it was believed that insertion of a charge packet into a deeply depleted MTOS junction would cause the MTOS device to switch to its ON state if the inserted charge packet exceeded some critical value, as determined by the physics of the MTOS junction. The switching threshold was thought to be adjustable by the pre-insertion of a charge packet smaller than the critical value. In this way, the MTOS device would perform the necessary magnitude comparison operation required in the charge-coupled computer.

Although the layout simplicity of such an MTOS charge packet threshold detector was a compelling reason for investigating its utility as a quantity magnitude comparator, an equally compelling reason was the history of the Yale research group in investigating the properties of ultra-thin tunneling oxides [Grant 1970, Gruodis 1971, Ma 1974, Iskarlatos 1974, Dressendorfer 1978, Lai 1979, Teng 1983]. Thus, significant momentum in the preparation, measurement and understanding of such thin oxides had been acquired.

As will be seen in this chapter, the MTOS junction can serve as a charge packet threshold detector, but the physical 'computation' time is long; typically 10 to 100 milliseconds. In retrospect, this is not surprising. As R.C. Barker is fond of saying, the positive feedback

loop has an integrator in it, with an R determined by tunneling physics, and C determined by a pair of capacitor plates spaced 35 Å apart. Under such circumstances, the operation of the device is interesting, but not very useful for the applications designated at the outset of this investigation.

However, the device structure which was fabricated to demonstrate the thresholding operation is very utile for probing the physics of the switching phenomena. Using this structure, it is possible to separate and measure electron and hole oxide transport currents as a function of oxide voltage by two different techniques and obtain consistent results. In addition, it is possible to estimate the impact ionization hole generation current by measuring the transient device current as it switches into its ON stable state.

To this end, the investigation of the MTOS charge packet threshold detector has been quite fruitful. Some previously outstanding questions regarding the device switching speed, and the quantitative roles various current components play in the switching process have been answered. The techniques developed in the course of this investigation will be useful for probing the behavior of thinner and thicker (Fowler-Nordheim) tunnel junctions, and for studying the current transport mechanisms of other dielectric thin films, such as silicon-rich SiO_2 and silicon nitride.

4.2. Review and Preview of MTOS Junction Device Physics

The MTOS junction as a bistable device has been investigated by Lai (1979) and Teng (1983) and in the present work. The tunneling physics has been covered by Dressendorfer (1978). In this section, a tutorial review of the device physics will be given, without unduly repetition of the fine work presented by the above authors. Device behavior which has

been probed in the present work will be included so as to lay a proper foundation for the discussion of experimental results later in this chapter.

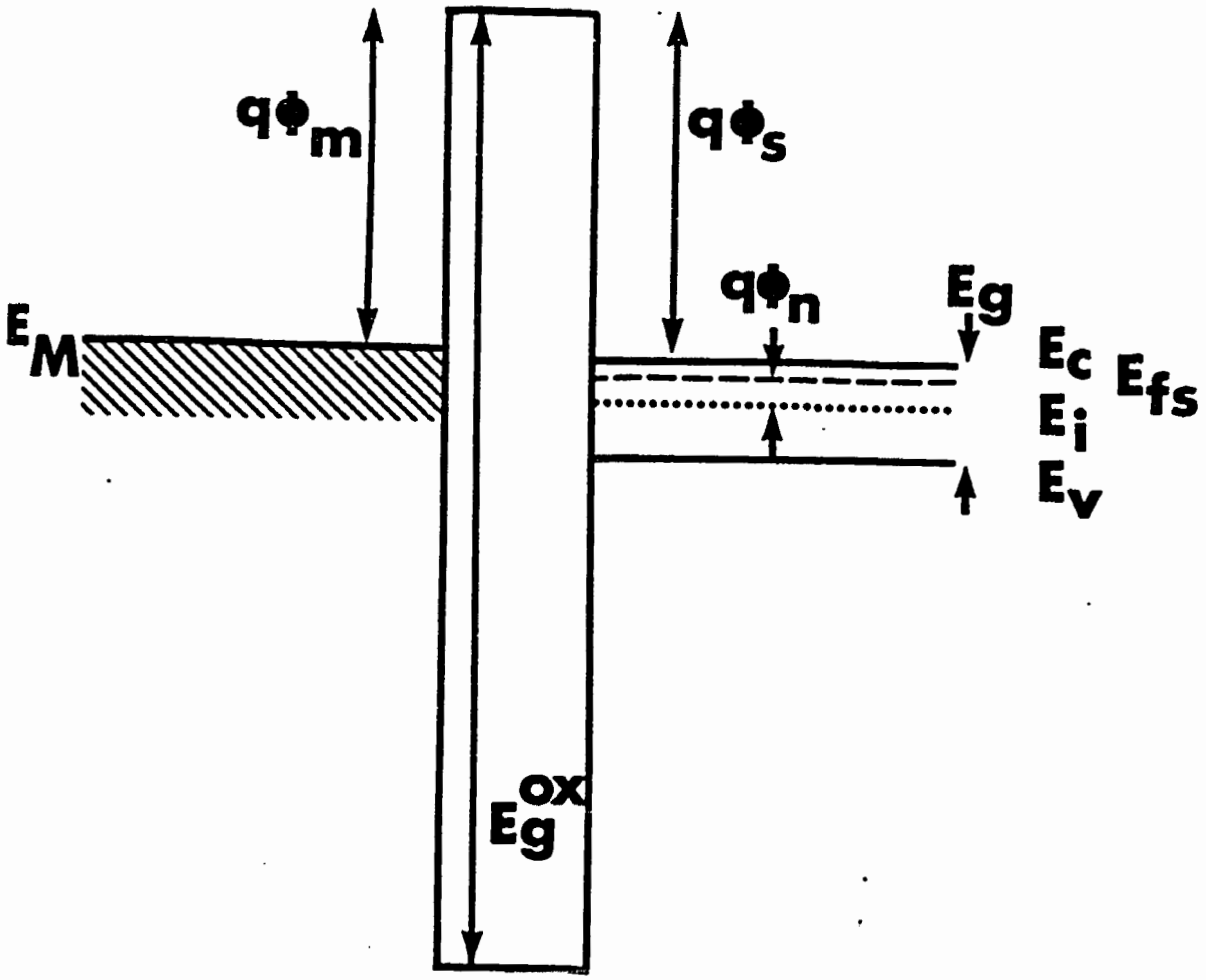
Energy Bands and Space Charge

An energy band diagram of the junction under the flat band condition is shown in Fig. 4.2-1. In the discussion of the junction, an aluminum/SiO_2/n-type Si structure with $N_D = 1 \times 10^{16}/cm^3$ and a tunnel oxide thickness of 33 Å will be assumed. For such a structure, $\phi_n = E_f - E_i$, is 0.35 volts at room temperature. The theoretical flat band voltage V_{FB} is given by

$$V_{FB} = \phi_{Al} - \phi_{Si} - E_g/2q + \phi_n \quad .$$
$$V_{FB} = -0.24 \text{ volts} \qquad 4.2\text{-}1$$

using $\phi_{Al} = 3.2$ volts, $\phi_{Si} = 3.23$ volts, and $E_g = 1.12$ eV. It should be noted that the first two voltages are the standard textbook [Sze, 1981] values of these quantities. However, as Dressendorfer (1978) points out, quoted values for ϕ_{Si} vary from 3.03 to 3.33 volts, and ϕ_{Al} shares a similar situation. The theoretical flat band voltage may vary a few hundred millivolts from the above calculated value, without any charge trapped in the oxide.

When a bias voltage V_{TO} is applied to the aluminum gate (substrate grounded), the semiconductor majority and minority carriers respond to form a space-charge region. In this thesis, attenuation is confined to negative values of V_{TO} so that the semiconductor surface is depleted of majority carriers. (Most of the tunneling calculations referred to above assume an accumulated state). The semiconductor space-charge region is then described by the calculations in appendix B. The applied bias is distributed between the oxide and the semiconductor in normal MOS fashion according to



4.2-1 MOS energy band diagram under flat-band conditions.

$$V_{TO} - V_{FB} = V_{OX} + \psi_s \qquad 4.2\text{-}2$$

where

$$V_{OX} = Q_D/C_{OX} + Q_{INV}/C_{OX}$$
$$V_{OX} = Q_S/C_{OX} \qquad 4.2\text{-}3$$

and Q_S is the total semiconductor space charge partitioned into an ionized donor component Q_D and an uncompensated minority carrier hole component Q_{INV}. For the device current range of interest, the semiconductor space-charge is given by

$$Q_S^2 = \frac{2q\varepsilon_s\varepsilon_o}{\beta}\left[n_{no}(e^{\beta\psi_s} - \beta\psi_s - 1) - p_s\,(e^{\beta\psi_s} - \frac{p_{no}}{p_s}\,\beta\psi_s - 1)\right] \qquad 4.2\text{-}4$$

In the appendix, it is shown that for $Q_{INV} \gg Q_D$ (typical for MTOS devices), the surface hole concentration is related to the total hole charge by

$$Q_{INV} = (2\varepsilon_s\varepsilon_o kT p_s)^{1/2} \qquad 4.2\text{-}5$$

The results of exercising the computer model for determining carrier concentrations is shown in appendix B.

From the above equations, it can be shown that the electrostatic potential, charge and field equations are a system of relationships with only two independent parameters. For most of this discussion in this chapter, the two most convenient parameters are applied bias V_{TO} and inversion layer charge Q_{INV}. Given these two values, quantities such as V_{OX}, ψ_s, p_s, Q_D and C can be calculated. In addition, device currents can also be computed. Thus, the parameters V_{TO} and Q_{INV} can be used define the state of the system. The MTOS states form a continuum which can be conveniently divided into eight regimes, based upon particular relationships among device current components. Prior to discussing each

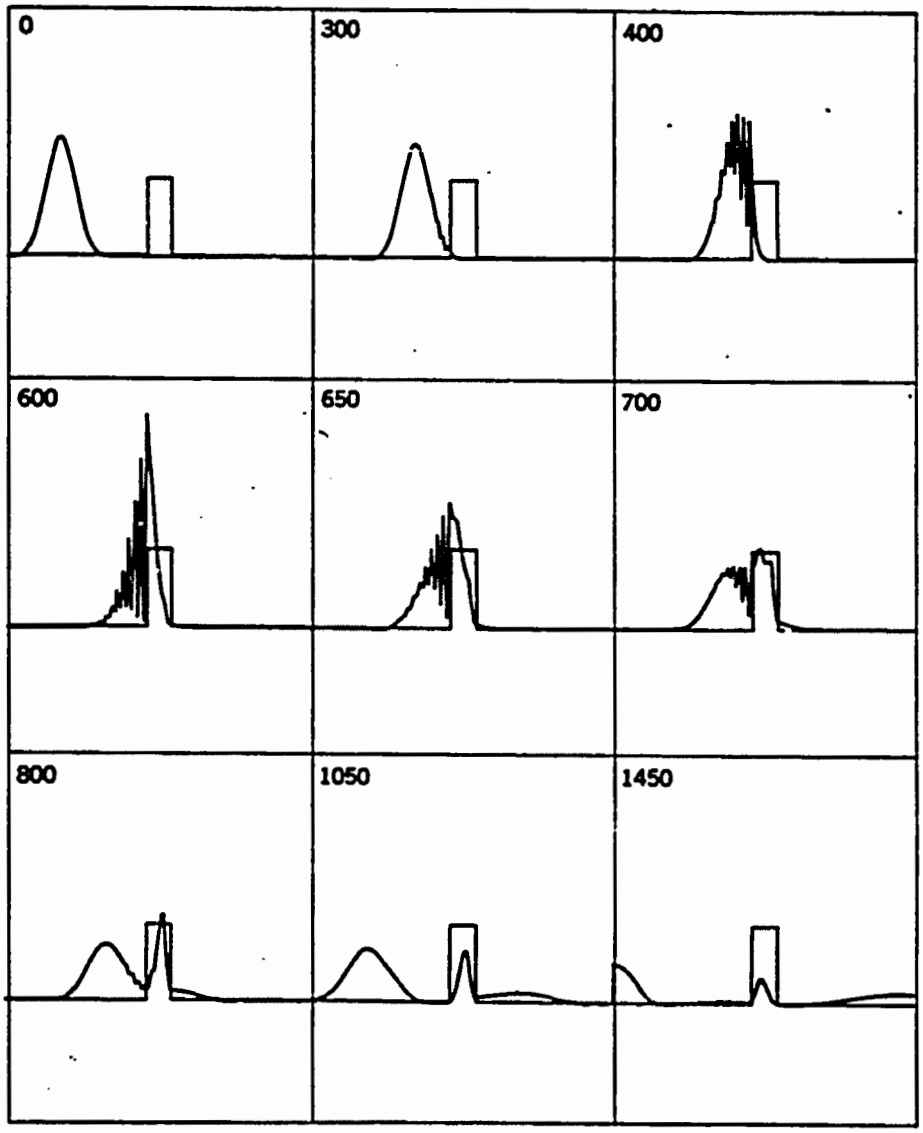
of these eight regimes, the physics of current transport throughout the MTOS device will be discussed.

Throughout the junction, the total current consists of a majority carrier (electron) component, and a minority carrier (hole) component. Each carrier current in turn, consists of various components representing different transport mechanisms, such as diffusion, drift and tunneling. These components vary widely in magnitude depending upon position in the junction, and the state of the system.

Oxide Current Transport (Tunneling, etc.)

For the oxide, the electron transport mechanism can be considered, in the simplest sense, to be one-dimensional quantum mechanical barrier tunneling. In this view, a stream of electrons in the metal (represented by a probability wave function) is incident upon the oxide energy barrier (a simple trapezoid). Solving Schrodinger's equation for this system yields a decaying wave function in the barrier, a reflected wave in the metal, and a propagating wave (of smaller amplitude) in the semiconductor. An electron wave packet incident upon the barrier is partially reflected and partially transmitted, as illustrated in Fig. 4.2-2.

From this simple model, several important features of current transport through the oxide may be described. The first is that the barrier penetration probability P_T depends exponentially upon the average barrier size E_B according to (for example, see Schiff, 1968, pg. 278)



4.2-2 Partial barrier transmission of an incident wave packet. (After Schiff, 1968. Reprinted with permission of the McGraw-Hill Book Company and the American Institute of Physics)

$$P_T = \exp -\left[\bar{E}_B/\hat{E}\right]^{1/2} \qquad 4.2\text{-}6$$

where $\bar{E}_B^{1/2} = \int_{X_L}^{X_R} \left[V(x) - E\right]^{1/2} dx$

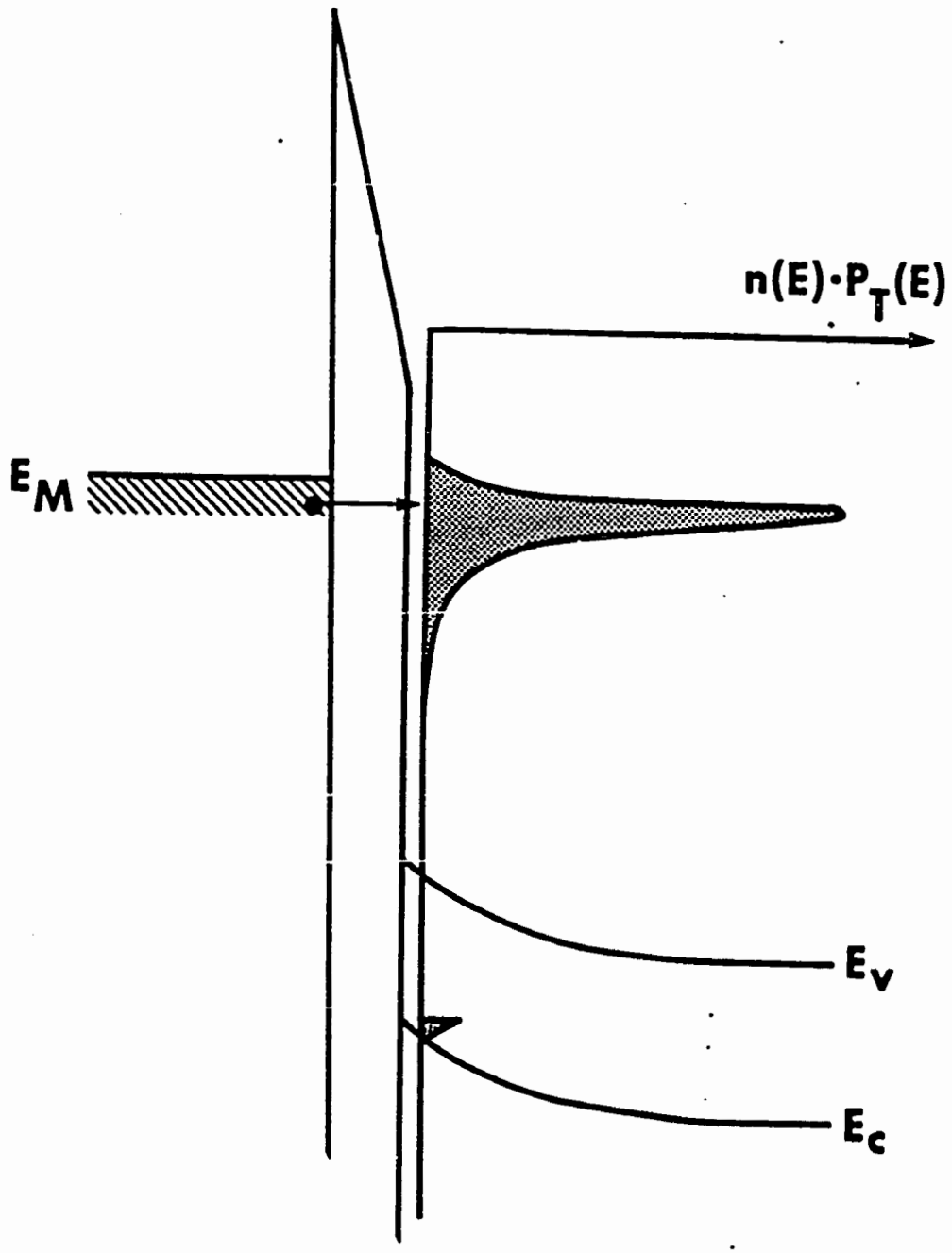
and $\hat{E} \triangleq h^2/8m$

The electron tunneling current J_{et} depends upon the number of electrons n(E) with energy E, their mean velocity $\bar{v}(E)$, and the barrier penetration probability $P_T(E)$ according to

$$J_{et} = \int_E qn(E)\bar{v}(E)P_T(E)dE \qquad 4.2\text{-}7$$

Since $P_T(E)$ depends exponentially upon the voltage drop across the oxide, the electron tunneling current is very sensitive to the total semiconductor charge, in particular Q_{INV}. This can be seen experimentally in Fig. 4.4-6 later in this chapter.

Another important feature of this simple model is that the current in a small energy range depends upon the $n(E)P_T(E)$ product. In Al-SiO_2 Si(n) tunnel junctions, this product peaks just below the metal Fermi level, as calculated by Lai (1979) and shown in Fig. 4.2-3. When this distribution of electrons emerges in the silicon, they are 'hot' because they have significantly more energy than thermally equilibrated silicon conduction band electrons. These hot electrons subsequently undergo net phonon (and possibly photon) emitting collisions with the silicon lattice. After many such collisions, they are no longer hot with respect to the silicon conduction electrons, and although these electrons still continue to collide with the lattice, they emit and absorb phonons at equal rates and hence are thermally equilibrated.



4.2-3 Distribution of tunneling electrons (After Lai 1979).

If the hot electrons have sufficient energy, a collision with the silicon lattice could result in the breaking of a silicon-silicon covalent bond, resulting in the generation of an electron-hole pair. This process is referred to as impact ionization and plays a crucial role in the bistable behavior of the MTOS junction, as will be elaborated upon shortly.

The simple one dimensional tunneling barrier model gives good qualitative insight into the MTOS junction behavior, but lacks the accuracy to yield quantitatively satisfying results. To refine the model, the full three dimensional electron dispersion relationship in the metal and in the semiconductor is introduced, but the simple trapezoidal potential barrier of the oxide is retained. In three dimensional tunneling, it is assumed that the tunneling process is specular, that is, only the wave vector which is in the direction perpendicular to the barrier plane is altered in the tunneling process. Hence, only the energy of the electron which is in the same direction counts for the barrier penetration probability. The wave vectors orthogonal to the above wave vector are conserved in the tunneling process. This wave vector conservation limits the final states available to a tunneling electron. In addition to wave vector conservation, the occupation probability of initial and final states (the latter is really an un-occupation probability) multiplies the penetration probability. To calculate the tunneling current including all these factors is an interesting yet long process, and has constituted the major portion of several theses [for example, Groudis 1971, Dressendorfer 1978].

Unfortunately such refinements still fail to yield quantitatively satisfying results. To provide greater accuracy, the model of the tunneling barrier can be improved. A two-band model for electron

dispersion in SiO_2 can be incorporated using a modified Franz form [Franz, 1956] to interpolate the imaginary wave vector in the forbidden energy range, using different effective masses at the conduction band and valence band edges. Image force barrier lowering may also be included. A non-uniform oxide thickness model (statistical variations on the order of an SiO_2 lattice cage size with a similar coherence length) may be considered. By slight and reasonable adjustment of oxide effective mass constants and the thickness variation parameter, Dressendorfer (1978) was able to achieve near perfect agreement between theory and experiment. Unfortunately, in such a finely tuned model, slight deviations in fitted values leads to enormous changes in tunneling currents, and as such, the predictive value of these models is somewhat moot.

Electrons in the metal can tunnel both into empty conduction band states and empty valence band states. In the latter case, it looks as though a valence band hole has tunneled through to the metal, therefore this process is sometimes referred to as hole tunneling. The penetration probability for an electron far enough below the metal Fermi level to be able to tunnel directly into the semiconductor valence band is many orders of magnitude less than for electron tunneling at the metal Fermi level into a hot state in the semiconductor conduction band. For thinner oxides (in the 20 Å range), calculations by Green and Shewchun (1974) indicate that the ratio of electron tunneling to hole tunneling currents should be in the 10^4 to 10^5 range. For thicker oxides, the ratio should be much higher.

Under these circumstances, 'hole' transport through oxides in the 35 Å thickness range may be dominated by mechanisms other than direct tunneling. One mechanism might be electron tunneling into interface trap states followed by subsequent hole recombination. The mechanism

favored by the author, but only in speculation, is that of a series of electron inelastic tunneling events through oxide neutral trap states, culminating in the electron exiting into the silicon valence band. Such a transport mechanism reduces the tunneling distance and hence the effective tunneling barrier dramatically. Although neutral traps in thick oxides have a relatively low density, it is possible that in a thin oxide where the entire oxide is composed of a strained bond 'transition' region [Grunthaner, et al., 1979], the number of neutral traps may be enhanced. Unfortunately such ideas must remain labeled speculation until experimental evidence discriminating among the transport mechanisms can be produced. Nevertheless, sufficient doubt about the hole transport mechanism remains in the author's mind to cause refrain from referring to hole transport as direct hole tunneling throughout this thesis.

Semiconductor Current Transport

The continuity equation for electrons and holes in a semiconductor is

$$\frac{\partial n}{\partial t} = G_{ext} - U + \nabla \cdot \vec{J}_n/q \qquad 4.2\text{-}8$$

and $$\frac{\partial p}{\partial t} = G_{ext} - U - \nabla \cdot \vec{J}_p/q \qquad 4.2\text{-}9$$

where G_{ext} is some non-thermal generation of carriers, U is the Shockley-Read-Hall (SRH) net recombination rate, and $\vec{J}_n$ and $\vec{J}_p$ are the electron and hole current densities. In the MTOS device, G_{ext} has two major sources. The first is the impact ionization process discussed in appendix C. The second is the optical electron-hole pair generation process in an illuminated device. Throughout this thesis illuminated MTOS devices are not considered since such generation was considered

extensively in the theses of Lai (1979) and Teng (1983). Thus, G_{ext} is consequential only under impact ionization conditions, and then only near the $Si-SiO_2$ interface.

The SRH net recombination rate U varies in sign and magnitude depending upon the state of the MTOS device and the position in the junction respectively. In deep depletion, $pn < n_i^2$ and U is negative. In 'thermal equilibrium' (i.e. $pn = n_i^2$), U is zero. In the super-inverted ON state, U is positive because $pn > n_i^2$ throughout the junction. In super-inversion, the pn product is pumped above thermal equilibrium due to the generation of electron-hole pairs by the impact ionization process.

The electron and hole currents $\vec{J}_n$ and $\vec{J}_p$ also depend on the state of the junction and position. The electron current throughout the entire junction is primarily drift toward the bulk. For holes, the current in deep depletion is characterized by drift in the space-charge region of thermally generated carriers toward the interface, and to a lesser extent, diffusion in the bulk toward the space-charge region. In super-inversion, the hole current is primarily diffusion from the super-inversion layer back toward the bulk.

Kirchoff Analysis of the Inversion Layer ($\dot{Q}_{INV}$)

For studying the transient switching behavior of the MTOS junction, the hole current is much more interesting than the electron current, although the electron current is what is measured externally. As was indicated above, the inversion layer of minority carriers Q_{INV} is a convenient state variable of the MTOS junction, and because the junction bias V_{TO} is fixed during the switching experiments performed in this investigation, Q_{INV} is the state variable which changes during switching. The time rate of change of Q_{INV}, which is labelled $\dot{Q}_{INV}$, is

controlled by the various components of hole current in the junction, which can be divided into charging (positive) currents and discharging (negative) currents. The charging currents are the thermal generation of holes J_{th}, optical generation of holes J_{op}, and the impact ionization generation of holes J_{ii}. The discharging currents are the oxide hole transport current J_{ht}, the recombination of holes J_{rec}, and the back diffusion of holes, J_{diff}. In addition to these currents present in an electrically isolated MTOS junction, a non-isolated junction may have a lateral injection/extraction current due to lateral diffusion or drift from adjacent devices which we call J_{lat}. Thus, we can write

$$\dot{Q}_{INV} = J_{th} + J_{op} + J_{ii} - J_{ht} - J_{rec} - J_{diff} \pm J_{lat} \qquad 4.2\text{-}10$$

The thermal generation and recombination of holes is actually the same current logically divided according to sign, and hence are not simultaneously present. The thermal recombination/generation current is defined according to

$$\left.\begin{matrix} -J_{th} \\ J_{rec} \end{matrix}\right\} \triangleq \int_{x=0}^{x_d} qU(x)dx \qquad 4.2\text{-}11$$

where x_d is the depletion layer depth and U is the SRH net recombination rate. The latter depends upon position, and has a delta function at the surface to account for interface trap generation/recombination. This term is normally disregarded, although it may contribute significantly to the integral. The problem is that for this oxide thickness range on (100) silicon, it is difficult to estimate the effective surface recombination velocity. Measurements by Teng (1983) on (111) silicon indicate that the surface recombination velocity is in a range similar to that of thicker oxides on (111) silicon. If a similar situation exists for (100) silicon, it may be justifiable to disregard surface

generation/recombination except under super-inversion conditions when both minority and majority carrier concentrations are significantly larger than n_i, and strong electric fields at the surface may enhance the recombination process.

As was discussed previously, the optically generated current was zero throughout the course of this investigation. However, as both Lai (1979) and Teng (1983) demonstrate, the optically generated current may control $\dot{Q}_{INV}$ under certain switching conditions. In deep depletion, the effect of illumination is to cause a current similar to that in a reverse-biased PN junction. However, when the pn product exceeds n_i^2, computing J_{op} is more complicated, and such a calculation is not considered here. Even in a thick MOS capacitor biased into inversion, the effect of illumination is to create a super-inverted layer of minority carriers in which the local optical generation rate exceeds the local recombination rate. To maintain steady state and zero total current, there is an equal minority and majority carrier flux away from the surface, in accordance with the continuity equations.

The impact ionization current is proportional to the electron tunneling current, where the proportionality constant ranges from 0.02 to 0.05 over the hot electron energy range probed by the bistable MTOS junction. The impact ionization current is nil below V_{OX} approximately equal to 2.2 volts for a 33 Å $Al/SiO_2/Si(n)$ junction, and becomes dominant in determining Q_{INV} for $V_{OX} > 2.4$ volts. The impact ionization process is described in greater detail in appendix C.

The diffusion current is defined at the depletion region edge, and is approximately equal to the total minority carrier current at this position. Solution of the continuity equation for zero electric field and constant recombination rate constant τ_p yields

$$J_{diff} = (qkT\mu_p/\bar{\tau}_p)^{1/2} [p(x_d) - p_{no}] \qquad 4.2\text{-}12$$

where $p(x_d)$ is the hole concentration at the depletion region edge. For deep depletion, J_{diff} is negative but negligible compared to J_{th} under most conditions. For example, using $N_D = 1 \times 10^{16}/cm^3$, $\bar{\tau}_p = 8.1$ μsec and $\mu_p = 480$ cm^2/V sec yields a diffusion current in deep depletion of 4.2 pA/cm^2 in contrast to J_{th} equal to 10 nA/cm^2. Under super-inversion conditions however, the diffusion current rises exponentially with decreasing surface potential and quickly exceeds J_{rec} as the dominant inversion layer discharging current component. Although it is a trivial observation to make, the minority carrier diffusion current in fact eventually recombines with majority carriers, so that it is only a matter of semantics to call one current recombination and the other diffusion. The author though, prefers to distinguish between the two currents. In super-inversion, with a typical surface hole concentration p_s of the order of 1×10^{21} per cm^3, and surface potential ψ_s of -0.56 volts, the back diffusion current is approximately 60 $\mu A/cm^2$ whereas the recombination current is calculated to be one hundred times smaller.

Regimes of MTOS State Space Operation

The continuum of MTOS state space (as parameterized by V_{TO} and Q_{INV}) can be conveniently divided into eight regimes of operation, which can be labelled one through eight in order of increasing Q_{INV} (increasing V_{OX}, decreasing ψ_s).

In the first regime, Q_{INV} is less than the inversion layer charge in the stable OFF state, and is characterized by deep depletion. The primary charging current is the thermal generation of holes in the space charge region. In this regime the hole oxide leakage rate is smaller than the thermal generation rate so that

$$\dot{Q}_{INV} = J_{th} - J_{ht}$$

In this regime, the oxide voltage drop is small so that the electron tunneling current is also small. When a bias voltage is initially applied to the MTOS junction, it operates in this regime, until the inversion layer builds up so that the device reaches the second regime.

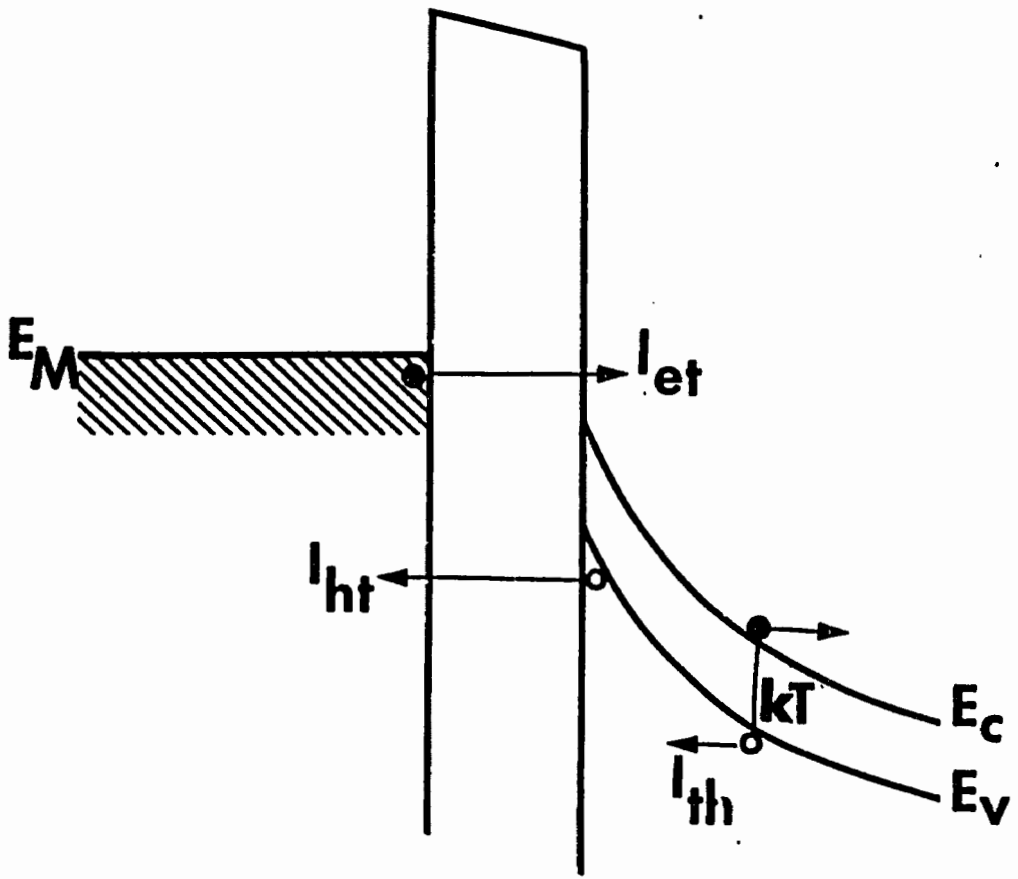
In the second regime, enough minority carriers have accrued at the interface so as to increase the hole leakage rate (and reduce the thermal generation rate) so that the discharging current balances the charging current and $\dot{Q}_{INV} = 0$. In this regime, the device is in the stable OFF state. The device is typically in a deep depleted low current mode since the oxide voltage drop is still small. If the oxide is of poorer quality (greater J_{ht}), the device currents balance at a higher value of J_{th} and lower value of J_{et}. If the number of generation centers is high (many interface or bulk traps), the device currents balance at a higher level of J_{ht} and a higher value of J_{et}. Thus, the old rule of thumb that low OFF currents reflect good device quality is in fact ambiguous. An energy band diagram depicting this state is shown in Fig. 4.2-4.

The third regime is categorized by net hole leakage through the oxide from the inversion layer, and spans from deep depletion to low level super-inversion. Throughout most of regime 3, the impact-ionization current is much smaller than the hole leakage current, except at the extremity of the regime where the two currents become comparable. In this regime

$$\dot{Q}_{INV} = - J_{ht} + J_{ii}$$

The thermal generation/recombination of minority carriers is small in comparison to $\dot{Q}_{INV}$, as well as any diffusion current.

In the fourth regime, the oxide voltage is large enough to cause the impact ionization charging current to balance the oxide hole transport



4.2-4 MTOS energy bands in the low current state.

discharging current. The net rate of charging is zero ($\dot{Q}_{INV} = 0$) and the device is in a metastable state. Experimentally, such a state is difficult to observe for durations longer than a few hundred milliseconds before some perturbation of the system knocks it out of the metastable state. This situation is roughly equivalent to balancing a pendulum pointing straight up. The inversion layer charge corresponding to this state is a critical one. For inversion layer levels less than this, an electrically isolated device's current will decay to the OFF state. For an inversion layer charge greater than this, the device current will rise to the ON state. Thus, this inversion layer charge is precisely the threshold charge required for charge packet threshold detector operation.

For inversion layer charge greater than that of the fourth regime, the device enters a fifth regime where the impact ionization charging current dominates and

$$\dot{Q}_{INV} = J_{ii} - J_{ht}$$

As Q_{INV} grows, the impact ionization current increases at a differentially faster rate than the oxide hole leakage current, increasing the positive feedback gain. The diffusion current remains smaller than J_{ht} in regime 5, by definition.

The sixth regime is defined for continued net charging but for back diffusion larger than hole leakage so that

$$\dot{Q}_{INV} = J_{ii} - J_{diff}$$

The back diffusion current doubles for every 18 mV change in surface potential and oxide voltage drop at room temperature, which is a differentially faster rate than the electron induced impact ionization tunneling current. This back diffusion current decreases the positive feed back gain, dampening the time rate of change of electron tunneling

current in an electrically isolated device.

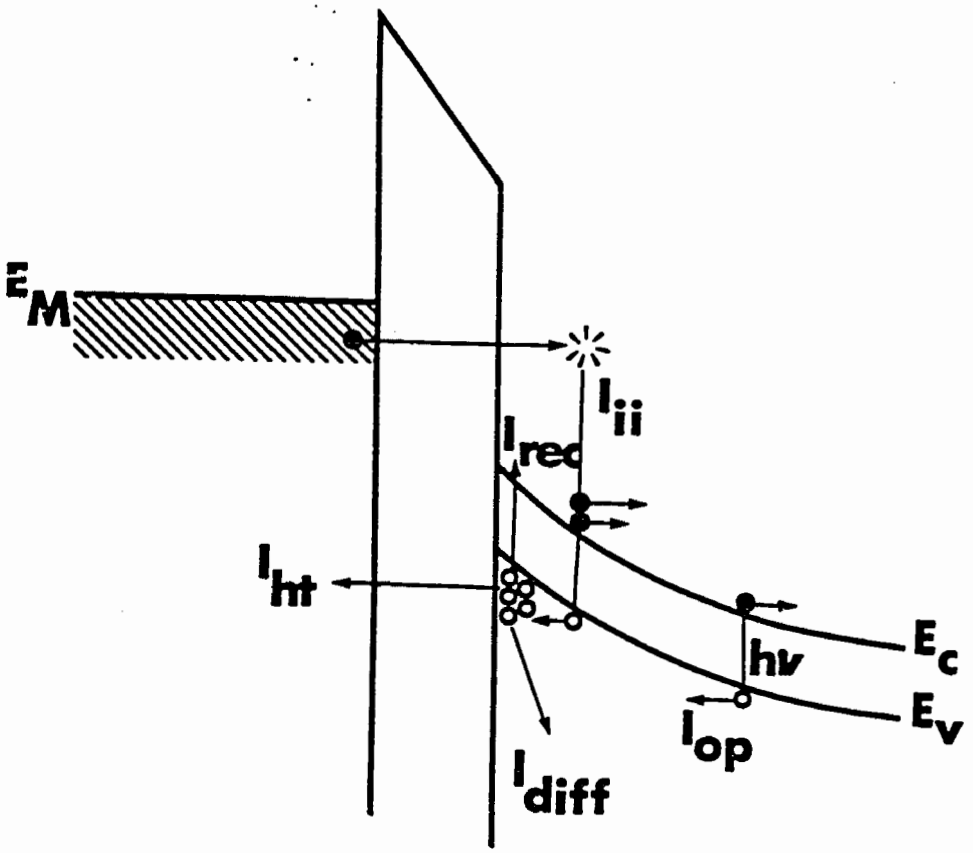
When the back diffusion discharging current has grown sufficiently to balance the impact ionization charging current, the device is in a stable super-inverted high current ON state, referred to as the seventh regime, and $\dot{Q}_{INV} = 0$. The stabilizing back diffusion current is such a strong function of surface potential, that the surface potential is approximately pinned independently of the applied bias voltage. Any change in V_{TO} falls across the oxide and boosts the electron tunneling current. Thus in the ON state, varying V_{TO} is like varying V_{OX} and the ON device current rises exponentially with V_{TO}. An energy band diagram schematically showing device operation in this regime is shown in Fig. 4.2-5.

The device can also exist in an ON-er than ON state which we refer to as the eighth regime. However, in this case $\dot{Q}_{INV}$ is very large and negative owing to the back diffusion current. An electrically isolated device in this state decays quickly to the seventh stable ON regime.

MTOS State Space Phase Diagram

The state of the MTOS device can be specified by two parameters, which can be used as the state variables of the system (such as V_{TO} and Q_{INV}). Using the two state variables, it is possible to draw a phase diagram of the MTOS state space as shown in Fig. 4.2-6, for a device with a 33 Å oxide. The horizontal axis is applied gate voltage and the vertical axis is inversion layer charge density normalized to a voltage by dividing by the oxide capacitance.

In this state space, curves of constant V_{OX} are nearly horizontal lines. To show this, recall Eq. 4.2-3 and note that for $N_D = 1 \times 10^{16}/cm^3$, and $V_{TO} = 5.0$ volts, the maximum value of Q_D is only 0.1 $\mu C/cm^2$ ($Q_{INV} = 0$). Thus,



4.2-5 MTOS energy bands in the high current state.

$$V_{OX} = \frac{Q_{INV}}{C_{ox}} + \Delta V \qquad 0 \lesssim \Delta V \lesssim 0.10 \text{ volts}$$

with ΔV less than one hundred millivolts for non-zero values of Q_{INV}.

Curves of constant surface potential ψ_s are almost exactly 45 degree straight lines. This can be seen by using Eq. 4.2-2 to write

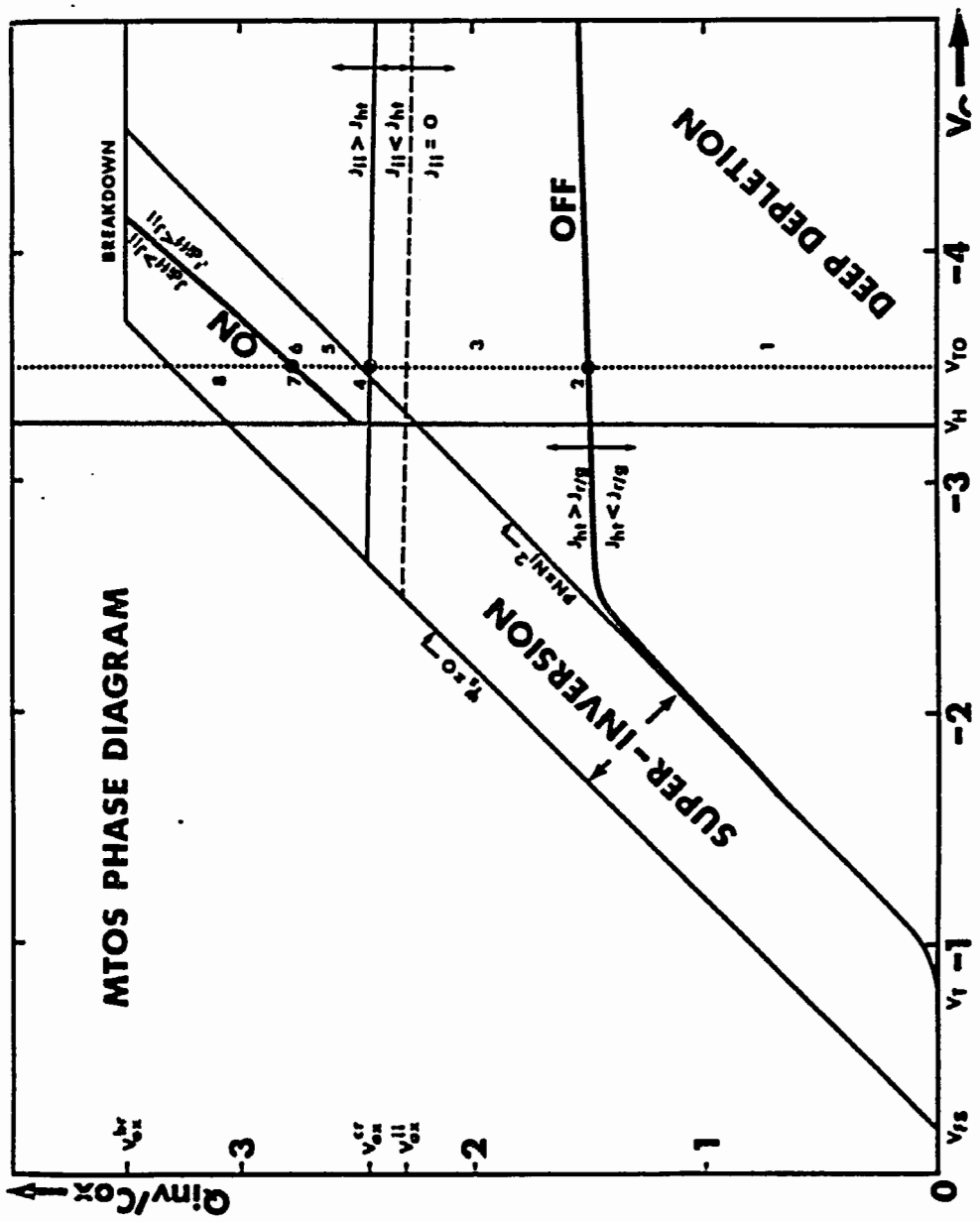
$$Q_{INV}/C_{OX} + \psi_s = V_{TO} - V_{FB} - \Delta V \qquad 4.2\text{-}12$$

The effect of varying the fixed value of ψ_s is to alter the horizontal axis intercept of the unity gain slope line.

The oxide breakdown voltage is determined by the product of the oxide thickness and the breakdown electric field strength of the oxide. This latter quantity is typically 10 MV/cm for tunnel oxides, and on the phase diagram it is represented by a nearly horizontal line. If the MTOS device is operated above this line, catastrophic oxide breakdown will result. It is believed that breakdown results in the formation of conducting channels in the oxide thereby shorting out V_{OX}. The failure is evidenced by loss of bistable device operation, nearly voltage independent current under illumination, and deep depletion capacitance under illumination. These experimental observations lead one to a conclusion that the oxide hole transport current mechanism has been significantly enhanced, preventing the formation of an inversion layer.

Zero surface potential is drawn on the phase diagram as a unity gain straight line with horizontal axis intercept V_{FB}. It is difficult to conceive of the device operating above this line since any inversion layer charge would be subject to strong electrostatic forces and the system would relax in a time comparable to the silicon dielectric relaxation time.

Another phase boundary which can be drawn corresponds to V_{OX}^{CR}, the



4.2-6 MTOS state space phase diagram.

fourth regime value of V_{OX}. Here the impact ionization current balances the hole leakage current and the device is in a metastable state. This phase boundary is drawn as a nearly horizontal line in Fig. 4.2-6. Also shown is a dashed horizontal line showing the value of V_{OX} when the impact ionization current becomes negligible.

The thermal equilibrium condition, $pn = n_i^2$, is drawn in the phase diagram as a nearly unity gain straight line except for small values of Q_{INV}. This line corresponds to the equilibrium inversion layer charge of an MOS capacitor if no oxide current transport took place. From appendix B, the equilibrium inversion layer charge can be calculated according to

$$Q_S^2 = \frac{2q\epsilon_s\epsilon_o}{\beta}\left[p_{no}(e^{-\beta\psi_s} + \beta\psi_s - 1) + n_{no}(e^{+\beta\psi_s} - \beta\psi_s - 1)\right] \qquad 4.2\text{-}13$$

In inversion, $Q_s \sim \exp(-\beta\psi_s/2)$ so that at room temperature, the surface potential changes by only 120 mV for a ten-fold increase in inversion layer charge. Indeed, for thick oxide MOS capacitors, ψ_s is normally considered to be pinned at $2\phi_n$. Thus, once the equilibrium inversion layer is established, it is reasonable to take the surface potential as nearly constant. For the distorted foot of the equilibrium phase curve, the horizontal axis intercept is V_T, the classic inversion layer threshold voltage.

The bistable MTOS states can be represented by two curves on the phase diagram. In the OFF state (regime 2) the inversion layer charge leakage rate (J_{ht}) must balance the thermal hole generation rate. For low values of Q_{INV}, the oxide drop is so small, that $J_{ht} = J_{th}$ implies

very small J_{th} and $pn \simeq n_i^2$. Note that small J_{th} does not imply a vanishingly small depletion region. This range of operation is the equilibrium diode mode of the MTOS junction. For larger values of applied bias, the oxide voltage drop causes the diode to switch from a tunneling-limited mode to a supply-limited mode, and V_{OX} becomes nearly pinned, and the OFF phase decouples from its nearly $pn \simeq n_i^2$ mode.

The ON phase is characterized by a nearly pinned surface potential as was discussed previously, due to the strong relationship between surface potential and back diffusion current. The ON curve is bounded by the breakdown phase and the holding voltage V_H. At the holding voltage, the charging impact ionization current is exactly balanced by oxide hole leakage and back diffusion. The fifth and sixth regimes have been shrunk to nearly coincide with the fourth and seventh regimes. For gate voltages below V_H, there is no stable ON state for the MTOS device because the back diffusion current plus hole leakage current always exceeds the impact ionization current.

Finally, on the MTOS state space phase diagram, a dotted line at $V_{TO} = -3.5$ volts is drawn, with the state space corresponding to each of the eight regimes duly labeled. Most of the ensuing experiments are performed at this bias.

This phase diagram represents the presently known operating phases of the MTOS device. The experimental device described in the next sections, injects a charge packet into the MTOS junction from a CCD input structure, and allows access to nearly every point on the phase diagram for measurement of the electron tunneling current, Q_{INV} and depletion layer depth for each MTOS state. As a challenge, the earnest reader might try to draw such a phase diagram for a thicker oxide wherein Fowler-Nordheim tunneling is the dominant electron transport

mechanism.

Switching Time

The switching time of the Al/SiO_2/Si(n) MTOS device is relatively and disappointingly slow. For an electrically isolated device, the switching time is determined by how long it takes to change the inversion layer charge from one level to another using the internal charging or discharging currents. For example, for the device to switch from regime 5 to the ON state, the time is simply determined by solving

$$Q_{INV}^{ON} - Q_{INV}\Big|_{t=0} = \int_{t=0}^{t^{ON}} \dot{Q}_{INV}\, dt \qquad 4.2\text{-}15$$

for the time t^{ON}. For most cases of interest, it is appropriate to write

$$\dot{Q}_{INV} \sim J_0\, e^{Q_{INV}/Q_0}$$

where J_0 and Q_0 are fitted parameters. Thus if the device switches from Q_{INV}^{o} at t = 0 to Q_{INV}^{ON}, the switching time is solved to be

$$t^{ON} = \frac{Q_o}{J_o}\left[\exp(-Q_{INV}^{o}/Q_o) - \exp(-Q_{INV}^{ON}/Q_o)\right] \qquad 4.2\text{-}17$$

For the devices measured in this work, (V_{TO} = -3.5 volts) J_0 is typically 1.00 pA/cm^2, Q_0 is 0.16 μC/cm^2 and Q_{INV}^{ON} equal to 2.7 μC/cm^2. Thus to switch from 2.4 μC/cm^2 to 2.7 μC/cm^2 might take 50 milliseconds. Note that to switch from 2.5 μC/cm^2 to the ON state takes nearly half as long, and that increasing Q_{INV}^{ON} (i.e. V_{TO}) does not significantly increase the switching time. The switching ON time is dominated by the

device operation near threshold, where the positive feedback gain is low.

The switching OFF time (from regime 3 to regime 2) is calculated in a similar manner, except that to switch OFF, $\dot{Q}_{INV}$ is negative. Eq. 4.2-16 can be used if J_0 is replaced with $-J_0$. In this case, the OFF switching time t^{OFF} is given by

$$t^{OFF} = Q_0 \left[\exp(-Q_{INV}^{OFF}/Q_0) - \exp(-Q_{INV}^{o}/Q_0)\right]J_0 \qquad 4.2\text{-}18$$

For devices measured in this work, Eq. 4.2-16 adequately describes $\dot{Q}_{INV}$ using J_0 approximately equal to 0.13 $\mu A/cm^2$ and Q_0 equal to 0.81 $\mu C/cm^2$. For the OFF state in a practical device, the current needs to be less than say, 1% of the ON current. For the experimental devices, this requires Q_{INV}^{OFF} less than 1.2 $\mu C/cm^2$ corresponding to an OFF switching time of approximately 1200 msec for Q_0 equal to 2.3 $\mu C/cm^2$.

The switching OFF time, although slow, is a measure of the oxide quality, since a 'perfect' oxide leads to the slowest possible switching time. Shorter switching OFF times are acheived only by a higher hole leakage rate. A thinner oxide would yield higher hole leakage rates but at the same time would require larger amounts of hole charge to leak away, since a thinner oxide would lead to a higher oxide capacitance. A thinner oxide would also suffer potential problems such as reduced uniformity, pinholes and durability.

Oxide Thickness Considerations

The viable oxide thickness range for bistable MTOS junctions has been discussed briefly by Lai (1979) and Teng (1983). For bistable operation, there must be two distinct stable states, though for practical device applications, the current should differ by at least a factor of two, if not larger. For the devices considered herein, the ON state is assumed to be maintained by impact ionization generation of

holes sustaining a super-inversion layer. The super-inversion layer in turn is responsible for the reduction of the oxide barrier to tunneling electrons. This super-inversion layer/impact-ionization process puts a lower limit on the oxide thickness due to the maximum field permissible in the oxide prior to catastrophic oxide breakdown. This can probably be alleviated slightly by choosing a gate metal with a lower work function which would result in enhanced impact ionization generation of holes through increased tunneling electron current and hot electron energy for a given voltage and oxide thickness. As mentioned above, a thinner oxide would also improve the OFF switching time. Still, problems in manufacturing reliable ultra thin tunnel oxides (< 25 Å) will probably limit the lower oxide thickness even if the gate metal is changed.

The upper oxide thickness limit is less certain. If the OFF state is to be one of deep depletion due to hole leakage, Teng (1983) estimates the upper oxide thickness limit to be approximately 45 Å. However, if the OFF state is one of near thermal equilibrium ($pn = n_i^2$), the Teng limit is increased. For example, consider an MOS capacitor where the equilibrium inversion layer size is just too small to cause significant electron tunneling, and the device rests in a stable, low current, near-equilibrium state. If the device is super-inverted, by illumination for example, the increased oxide electric field could lead to an enormous Fowler-Nordheim tunneling current and greater than 100% impact ionization probability due to the high hot electron energies (~ 3.5 eV). Such a large hole generation rate may be sufficient to maintain the super-inverted state, and compete favorably with back hole diffusion and recombination.

Although at the time of this writing it is mostly speculation, hot electrons in the oxide and at the $Si-SiO_2$ interface in the course of

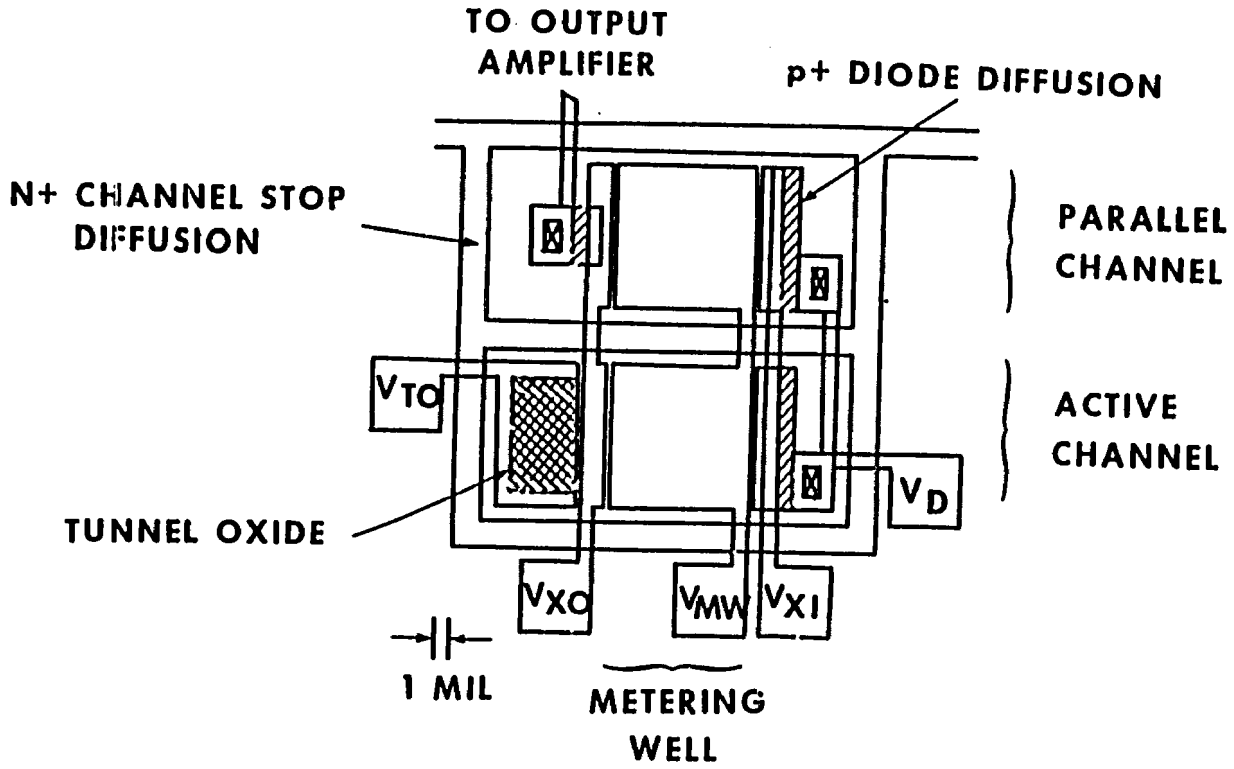
normal operation of a thicker oxide device may result in undesirable effects such as neutral and charged traps in the oxide and at the interface [Nicollian, et al. 1969]. The generation of a charge moment in the oxide would lead to a shift in the device operating characteristics, which in turn invalidates the potential utility of a thick oxide bistable device. Thus, from practical considerations, it may be undesirable to have hot electrons in the oxide, and hot electrons above a critical energy at the $Si-SiO_2$ interface.

In conclusion then, the oxide thickness limits for bistable operation are unclear, and that perhaps some investigation into the considerations described in this sub-section are warranted.

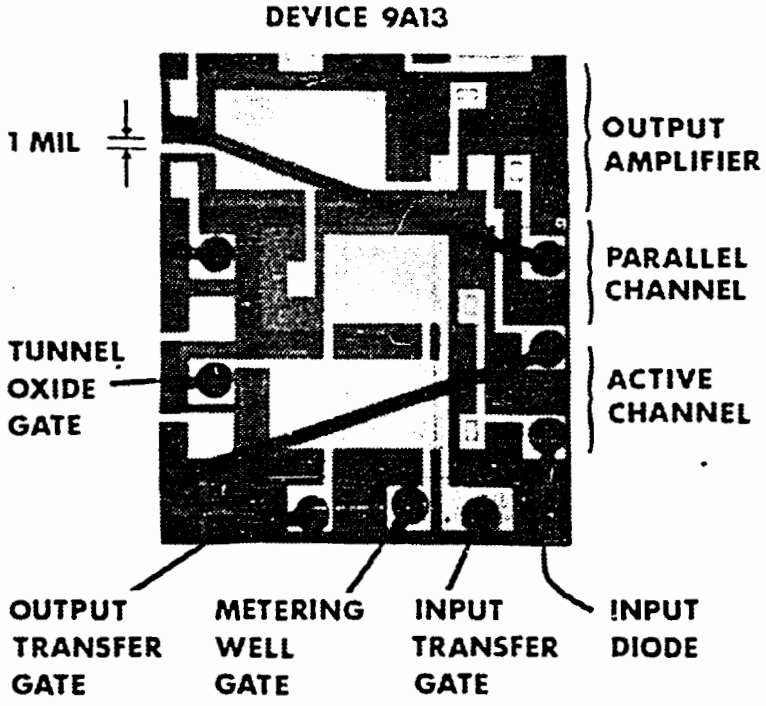
4.3 Experimental Device Layout and Fabrication

In order to test the proposed MTOS charge packet threshold detector, the layout shown in Fig. 4.3-1 was used. A photograph of the fabricated device is shown in Fig. 4.3-2. The structure consists of two parallel channels, each isolated by a surrounding channel stop diffusion. The two channels are wired in parallel to reduce the number of bonds required for the experiment, and to be sure that identical voltages are present in each channel. The channels are laid out in close proximity to assure that the physical characteristics of each channel are as closely matched as possible (such as oxide thickness and substrate doping concentrations). In retrospect, the entire structure may have been more versatile if the two channels were not hard-wired in parallel.

Each channel has an input structure consisting of an input p^+-n diode (V_D), and an output transfer electrode (V_{XO}). The input structure was designed to be operated in a surface potential equilibration mode [Tompsett, 1975] commonly referred to as 'fill and spill' as described in chapter 2. The charge packet formed in the active channel is then



4.3-1 Layout of MTOS charge packet threshold detector experiment. Interelectrode gap size is exaggerated.



4.3-2 Photograph of fabricated device structure for MTOS charge packet threshold detector experiment.

transferred to the MTOS device which is adjacent to the output transfer gate. A cross-section of the active channel is shown in Fig. 4.3-3. The charge packet formed in the parallel channel is transferred to a precharged MOSFET source-follower amplifier. The operation of the output amplifier is discussed in appendix D.

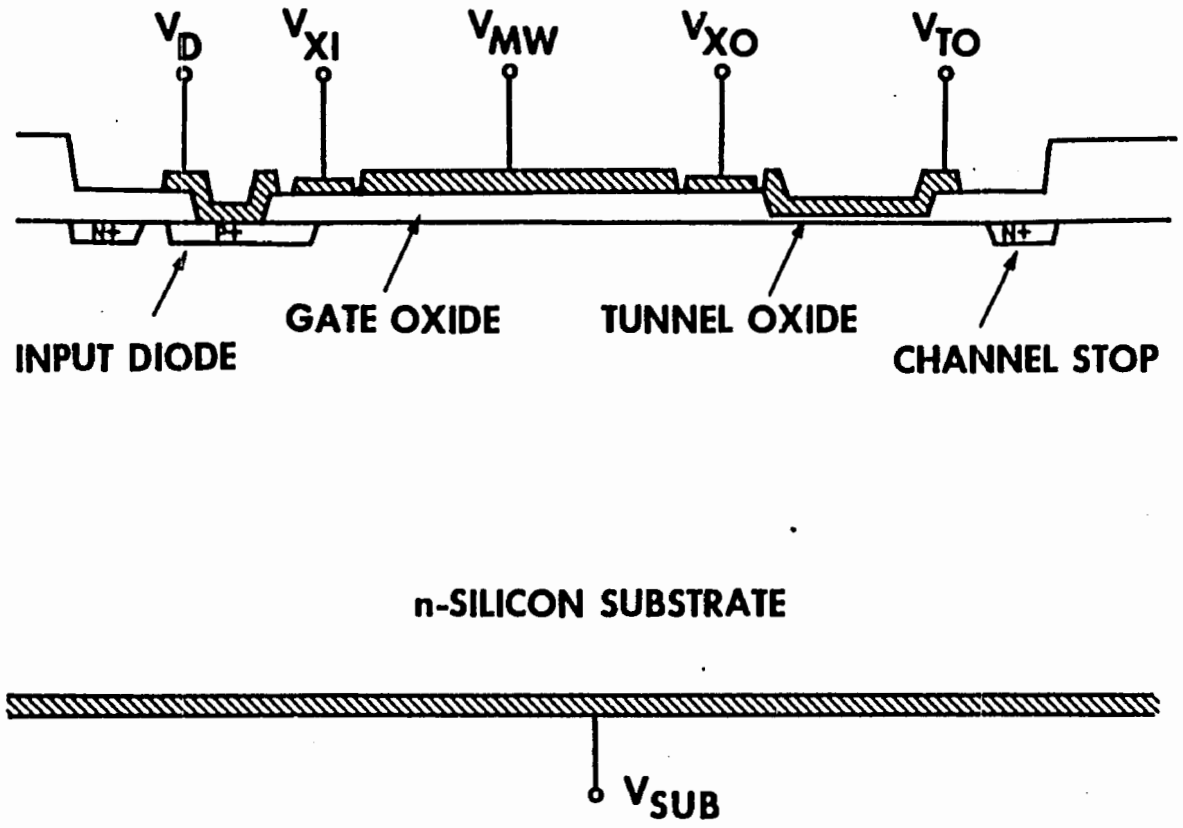
The purpose of the parallel channel is to create a charge packet identical to the one transferred to the MTOS device. In this manner, the magnitude of the latter packet can be inferred from the response of the output amplifier. Although designed to be used as a 'live' monitor of the charge packet injected into the MTOS junction, in practice, the large clock signal capacitive displacement currents and substrate leakage currents of the source-follower MOSFET obscure the MTOS junction current. Thus, sequential monitoring rather than simultaneous monitoring was performed. In addition, the parallel channel was used as a basic test of CCD input/output functionality.

There were several considerations in the geometrical design of the experiment. First, the area of the MTOS junction needs to be large enough to permit a reasonably noise-free measurement of the device tunneling current to be made, yet it must be small enough to be fabricated with reasonable yield (pinhole free, etc.). Also, the charge required to switch the device ON must be supplied from a modestly sized input structure. The charge per unit area of an inverted MOS capacitor is given approximately by

$$-Q_{INV} = C_{OX} (V_G - V_T) \qquad 4.3\text{-}1$$

where C_{OX} is the oxide capacitance, V_G the gate voltage, and V_T the inversion layer formation threshold voltage.

For an MTOS junction with an oxide 35 Å thick operated at 3.4 volts, the charge per unit area works out to be approximately 2.3 $\mu C/cm^2$. For



4.3-3 Cross-section of active channel of MTOS charge packet threshold detector experiment.

a thicker oxide of 500 Å, the charge per unit area is 0.07 $\mu C/cm^2$ per volt. To turn the MTOS device ON with a 15 volt bucket requires a junction area ratio of 2.3:(0.07 x 15) or 2.2:1. It was decided to use an area ratio of 2.5:1 by choosing the gate oxide area to be 100 $mils^2$ and the tunnel oxide area to be 40 $mils^2$.

There were some differences between design and practice. First, the tunnel oxide thickness was 33 Å. Second, Eq. 4.3-1 underestimates the size of the MTOS inversion layer charge in the ON state because the device is actually super-inverted. Third, the MTOS device was operated at 3.5 volts to stretch regime 5 as seen in the phase diagram of Fig. 4.2-6. The result of these three factors is that the ON state MTOS inversion layer charge density is closer to 2.8 $\mu C/cm^2$, nearly 25% higher than the design.

The gate oxide thickness was 555 Å thick, rather than 500 Å. In addition, Eq. 4.3-1 is inappropriate for the fill and spill operation, unless V_T is replaced with V_{XI}, the input transfer gate voltage. This latter voltage is typically 4.0 volts. Thus, to turn on the MTOS device, a metering well voltage of approximately 20 volts is required. This voltage is uncomfortably close to the 25 volt maximum swing of the MOS CCD clock drivers used in the test station, as described in appendix E.

Not only is the area of the MOS device important, but the W/L ratio of the individual oxide areas is important, too. (W is measured transverse to the direction of charge transfer). In general, for CCD devices, W/L is usually chosen to be large in order to improve the transfer time. This strategy was adopted in the layout of this experiment as much as possible, but it is cumbersome to layout a 100 $mils^2$ area device, and have the parallel channel described previously.

A compromise solution was settled upon and the W/L ratios of the input transfer gate, metering well gate, output transfer gate and tunnel oxide gate were chosen to be respectively 10:1, 10:10, 10:2, and 8:5. Although the transfer time for such enormous devices is slow compared to typical transfer times obtained in commercial optimized CCDs, the transfer time is still much more rapid than the MTOS switching time, and is therefore adequate.

The input diode was designed to span the width of the input transfer gate. In retrospect, with regard to substrate leakage currents, it is best to minimize the area of the input diode, since the diode area reduction has minimal impact on the speed of the fill and spill cycle. The speed of the latter is limited by the 100 $mils^2$ metering well geometry.

Although the fabrication process was discussed in some detail in chapter 3, this particular layout is sensitive to the alignment of the interelectrode gap to the tunnel oxide region edge, though the interelectrode gaps themselves are self-aligned. However, to minimize charge transfer inefficiency, the gap between the tunnel oxide electrode and the output transfer gate should be placed as close as possible to the tunnel oxide region, without having the tunnel oxide itself exposed 'thru' the gap. If the tunnel oxide is exposed, it is subject to possible contamination. More importantly, if the output transfer gate metal overlaps the tunnel oxide, application of relatively large transfer clock voltages will result in the breakdown of the tunnel oxide, destroying the experiment.

If the tunnel oxide electrode significantly overlaps the gate oxide, a barrier to charge transfer may result. The reason is that the small voltages applied to this electrode may be insufficient to form a channel

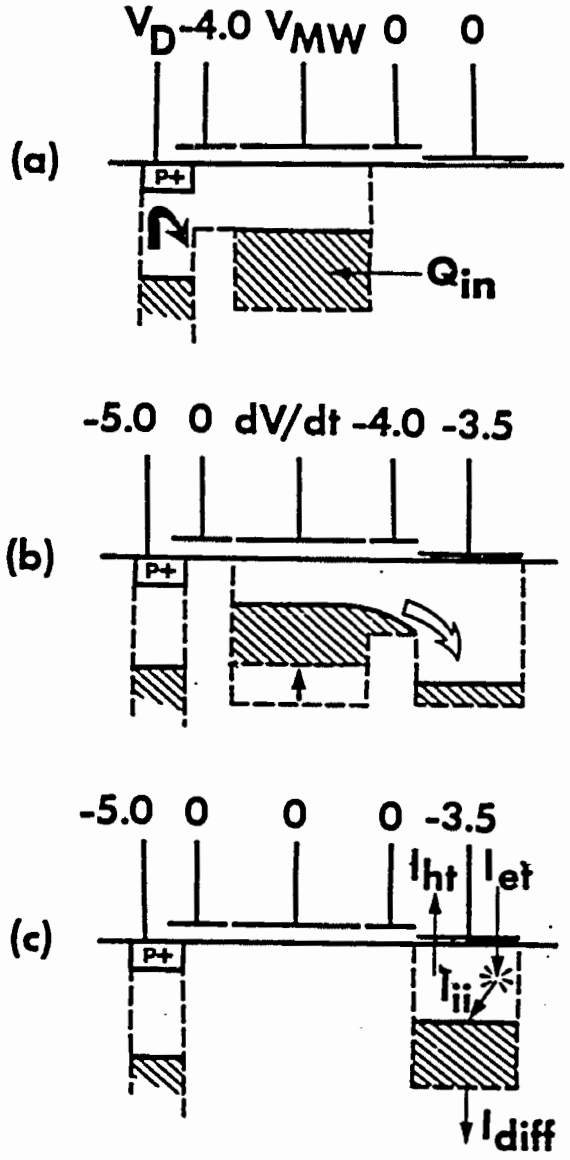
under the gate oxide for charge transfer. Experimentally, the gate oxide threshold voltage is -2.0 volts, so that applying a tunnel oxide bias of -3.5 volts, a conductive, albeit high impedance channel is formed. The consequence is that the high impedance channel acts as a bottleneck to lateral charge transfer, and the transfer rate must be reduced to accomodate the flow of charge. Experimentally, the transfer efficiency degrades when currents exceeding approximately 50 μA are pushed through this bottleneck. The theoretical saturation current for a MOSFET of this geometry using μ_p = 200 cm^2/V sec is 60 μA so that the experimentally observed bottleneck is reasonable.

4.4. Charge Packet Threshold Detector Experiment

Operation

The operation of the experiment can be divided into two stages. In the first stage, an input charge packet Q_{IN} is formed using the surface potential equilibration technique discussed in chapter 2. In the second stage, the charge packet is injected into a freshly formed MTOS potential well. The magnitude of the MTOS device current (after a physical computation time δt) is the output of the detector. If the input charge packet exceeds the threshold value Q_T, the device will switch ON and be in a high current state. If the charge packet is less than Q_T, the device will be in a low current state after time δt. A schematic cross-section of the active channel is shown in Fig. 4.4-1 depicting the operation of the experiment.

Initially all the applied voltages are zero, except for the diode which is reverse biased at -5.0 volts, and the semiconductor surface underneath the oxide is in a slightly accumulated state. A bias of -4.0 volts is applied to V_{XI}, followed by V_{MW} being applied to the metering well. The diode voltage is then raised to substrate potential (zero volts), resulting in a lateral forward injection of holes into the



4.4-1 Schematic cross-section of active channel illustrating charge packet injection sequence.
a. Charge packet is formed using fill and spill technique.
b. Metering well voltage ramped to zero causing ejection of charge packet.
c. MTOS junction switches into ON state.

metering well. As the diode voltage is reset to -5.0 volts, it extracts holes from the metering well and input transfer gate well until the equivalent MOSFET channel under the input transfer gate becomes depleted of all minority carriers and pinches off at both ends. As described by Tompsett (1975), the metering well retains a charge packet Q_{INV} whose magnitude is given by the expression

$$Q_{IN} = A_{MW}\, C_{OX}\, [V_{MW} - V_{XI}] \qquad 4.4\text{-}1$$

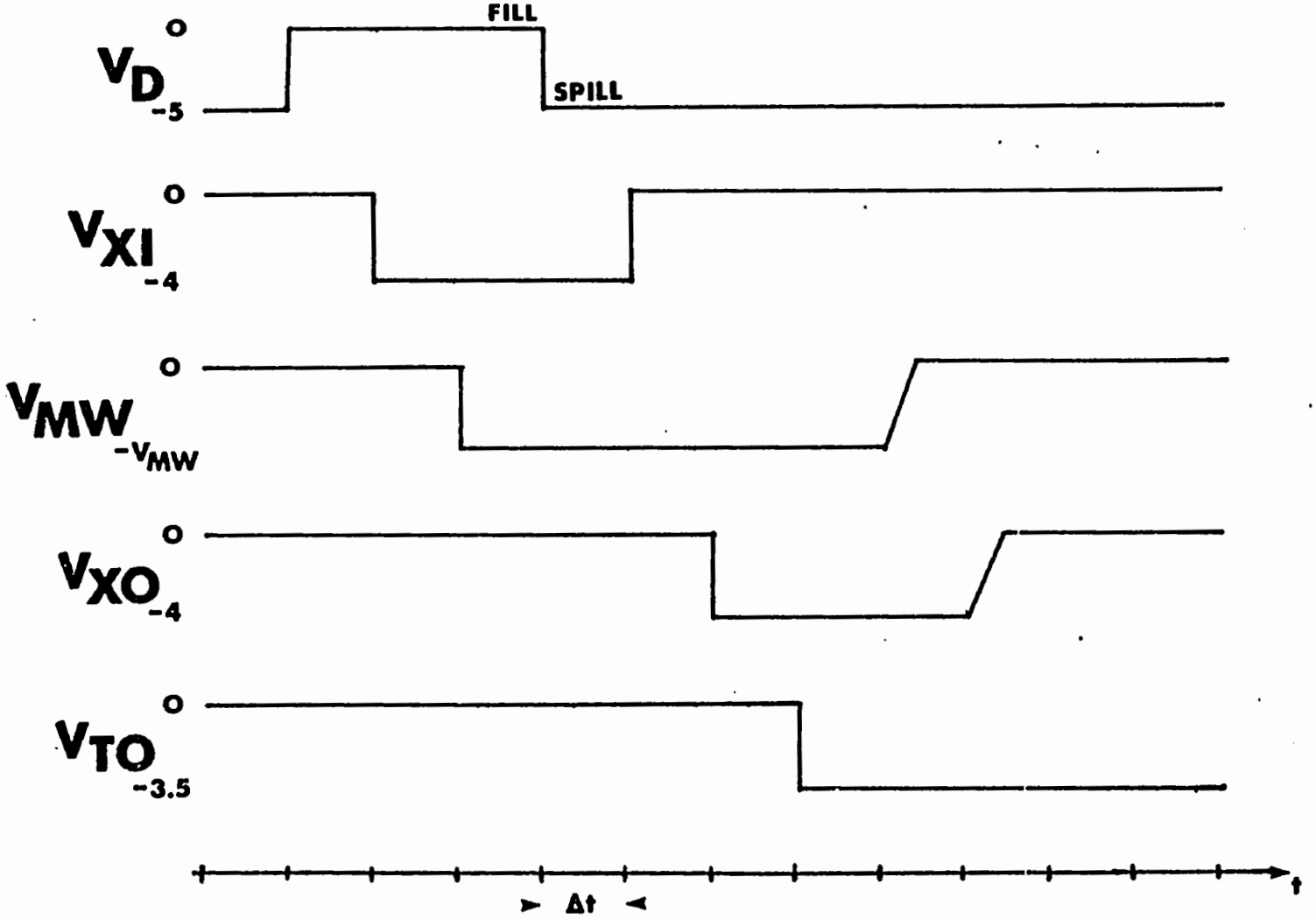
where A_{MW} is the area of the metering well and C_{OX} the specific gate oxide capacitance. [Note: Strictly speaking, this expression is written for an n-channel device, but for the sake of readability, Eq. 4.4-1. is written as shown rather than reversing V_{MW} and V_{XI}]. Once this spill cycle is complete, the input transfer gate voltage is set back to zero, isolating the metering well.

To inject the charge packet into the MTOS potential well, V_{XO} is set to -4.0 volts, followed by V_{TO} set to (typically) -3.5 volts. The metering well voltage is ramped to zero at approximately 0.25 volts/μsec causing the charge packet to be ejected from the metering well and injected into the MTOS potential well. After V_{MW} is ramped to zero, V_{XO} is clamped off to guarantee that the charge packet is confined to the MTOS potential well.

The timing described above is shown in Fig. 4.4-2. The charge packet Q_{IN} has set the initial state of the MTOS device to a point on the phase diagram of Fig. 4.2-6 given by V_{TO} equal to -3.5 volts (vertical dotted line) and an inversion layer voltage of

$$Q_{INV}/C_{OX} = A_{TO}\, Q_{IN}\, d_{OX}/\varepsilon_{OX}\, \varepsilon_0 \qquad 4.4\text{-}2$$

where C_{OX} and d_{OX} refer to the tunnel oxide, and A_{TO} is the area of the tunnel oxide region.



4.4-2 Timing diagram for charge packet injection sequence. Typical value of Δt is 350 μsec.

Experimental Results - Device Currents

A multiple exposure photograph of the substrate current as a function of time is shown in Fig. 4.4-3 for various values of V_{MW}, for device 9A10 operated with $V_{TO} = -3.45$ volts. The substrate current on the displayed scale can be considered to be the MTOS junction current because the reverse biased input diode current was below 1 nA.

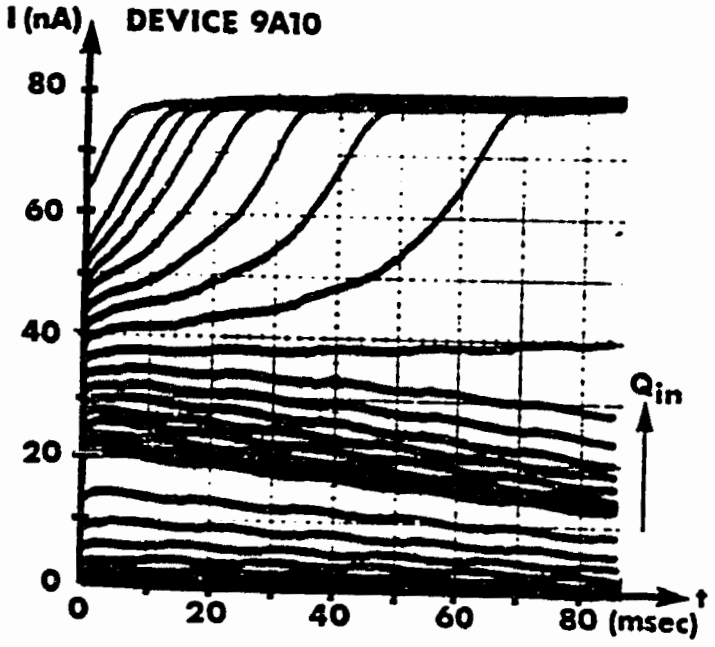
Using this photograph, a number of salient device features can be observed. First, the ON switching time is in the 10-100 millisecond range and the OFF switching time is in the 100-300 millisecond range. Second, the threshold value of Q_T corresponds to $V_{MW} = -19.5$ volts, or $Q_T = 634 \pm 2$ pC. Third, the transient response I(t) appears to be time invariant, that is

$$\dot{I} = f(I) \qquad 4.4\text{-}3$$

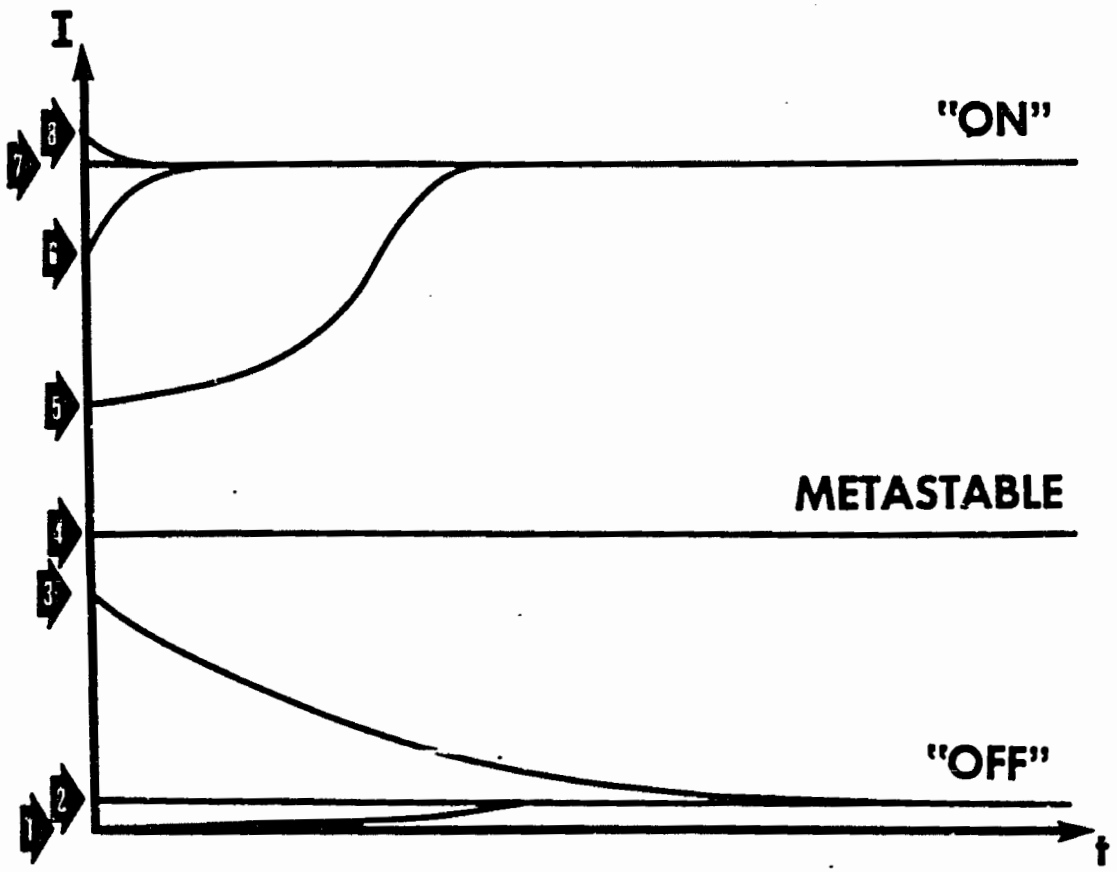
for a given value of V_{TO}. Physically this is reasonable since the switching time is much longer than the dielectric relaxation time (pico second range) so that the current could be used as a state variable in conjunction with the applied bias. Fourth, the initial transient device current just subsequent to charge packet injection increases exponentially with inversion layer charge.

In Fig. 4.4-4, another photograph is displayed which shows the transient device current for various values of V_{MW} for device 9A11 biased at $V_{TO} = -3.30$ volts. This photograph also illustrates device operation in regime 8 (ON-er than ON) for the largest values of V_{MW}.

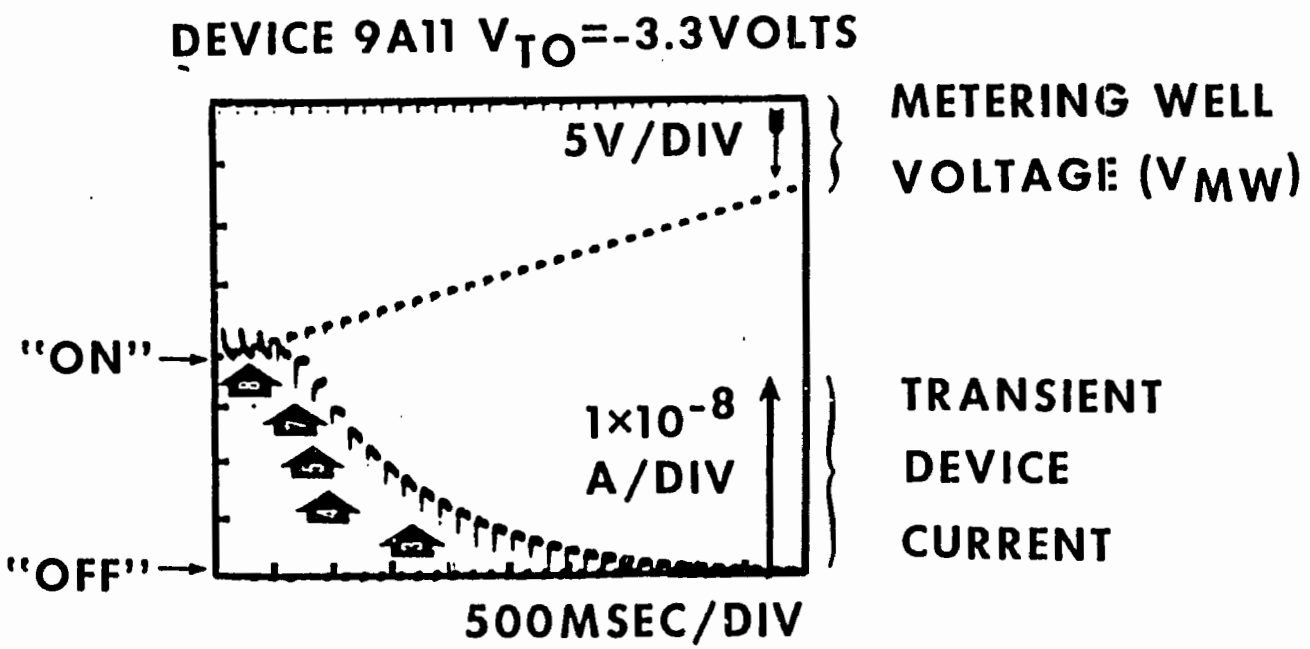
In order to make a more accurate measurement of device current versus inversion layer charge, a sample and hold circuit (as described in appendix E) was used to measure the transient device current just subsequent to charge packet injection. This device current is sampled



4.4-3a. Transient MTOS device current as a function of time for various sizes of injected charge packets. Metering well voltage varied from 12 to 22 volts in one volt increments except between 18 and 21 volts where the increment is 0.2 volts. Waveform ripple is residual 60 Hz noise. MTOS junction biased at -3.45 volts.



4.4-3b. Schematic drawing of transient current of Fig. 4.4.3a with initial MTOS state regimes labeled.



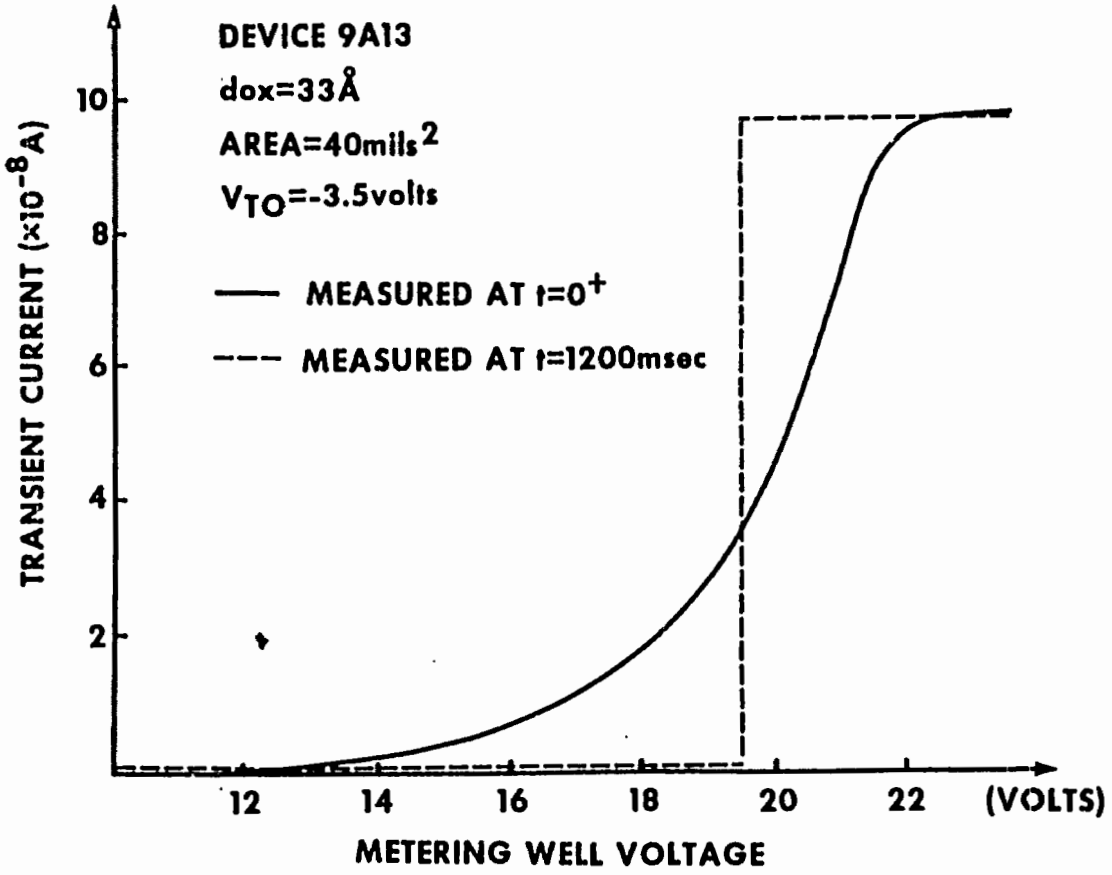
4.4-4 Transient MTOS device current biased at -3.30 volts for various values of injected charge packet magnitude. Note the regime 8 behavior for large injected charge packets.

before the MTOS inversion layer charge density departs significantly from the expression in Eq. 4.4-2 due to the impact ionization generation of holes, or the oxide leakage of holes. Unfortunately, the sampling is too slow to work well in regime 8 (partly due to the 'slow' response of the ammeter), thus the current as a function of increasing injected charge packet appears to flatten out for injected charge packets exceeding the ON state inversion layer charge. The result of this measurement is shown in Fig. 4.4-5. Also shown by the dashed line is the device current 1200 milliseconds after charge packet injection illustrating the thresholding properties of the device subsequent to the physical computation time.

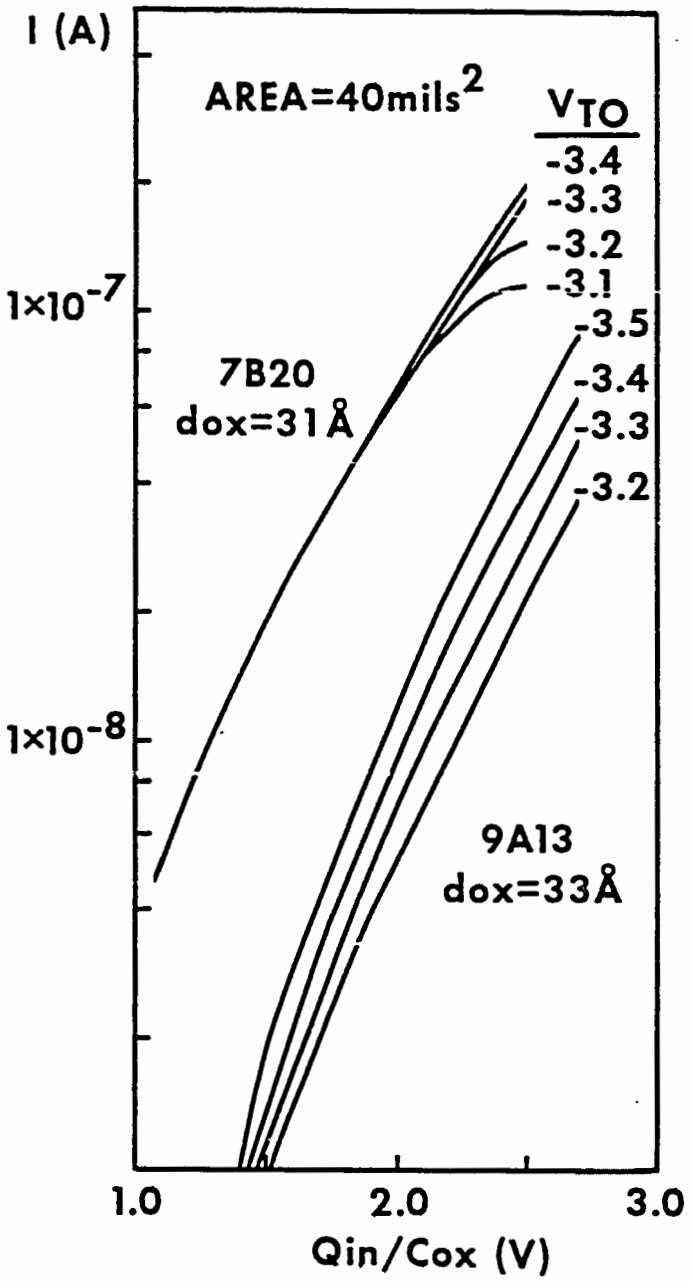
The same data is replotted in the semi-log plot of Fig. 4.4-6, with the data for different values of V_{TO} also shown. A similar measurement was performed on device 7B20 ($d_{OX} \simeq 31$ Å) and the results are also plotted in Fig. 4.4-6. Note that for the latter device, the current is almost independent of V_{TO} up to the saturation level, whereas for 9A13 there is a slight dependence of V_{TO}. This phenomenon was reinvestigated and it was discovered that the 9A13 V_{TO} dependency was anomalous and an artifact of the measured circuit, not present during the measurement of other devices such as 7B20. Because 9A13 did not survive long enough to be remeasured, other wafer 9A devices were retested, with the fortuitous result that the $V_{TO} = -3.5$ volts curve for 9A13 appears to be correct and reasonable.

For this latter curve, the data can be fitted using the expression

$$I(V_{TO}, Q_{INV}, t)\Big|_{\substack{t=0^+ \\ V_{TO}=-3.5 \text{ volts}}} = I_{et}^{o} \exp\left[Q_{INV}/Q_{et}\right] - I_{et}^{\infty} \qquad 4.4\text{-}4$$



4.4-5 Transient MTOS device current biased at -3.50 volts as measured at t = 0^+ and t = 1200 msec as a function of metering well voltage.



4.4-6 Semilog plot of transient MTOS device current as measured at $t = 0^+$ as a function of normalized injected charge packet magnitude for two different devices biased at various voltages.

with fitted values I_{et}^{o} = 26 pA, Q_{et} = 87 pC and I_{et}^{∞} = 800 pA, with the 'et' subscript referring to the electron tunneling current.

If a sample and hold operation is performed on the transient device current at $t = 0^{+}$ and at $t = \delta t$, the difference is a measure of $\delta I/\delta t$, the time rate of change of transient device current. A plot of $\delta I/\delta t$ measured using this technique is shown in Fig. 4.4-7 as a function of metering well voltage for device 9A13. In this experiment, δt was 4 milliseconds and the sample and hold difference signal was passively filtered using an R-C time constant of 2 seconds to reduce 60 Hz noise. The metering well voltage was ramped at 10 mV/sec from 18 volts to 23 volts.

To analyze the data, note that

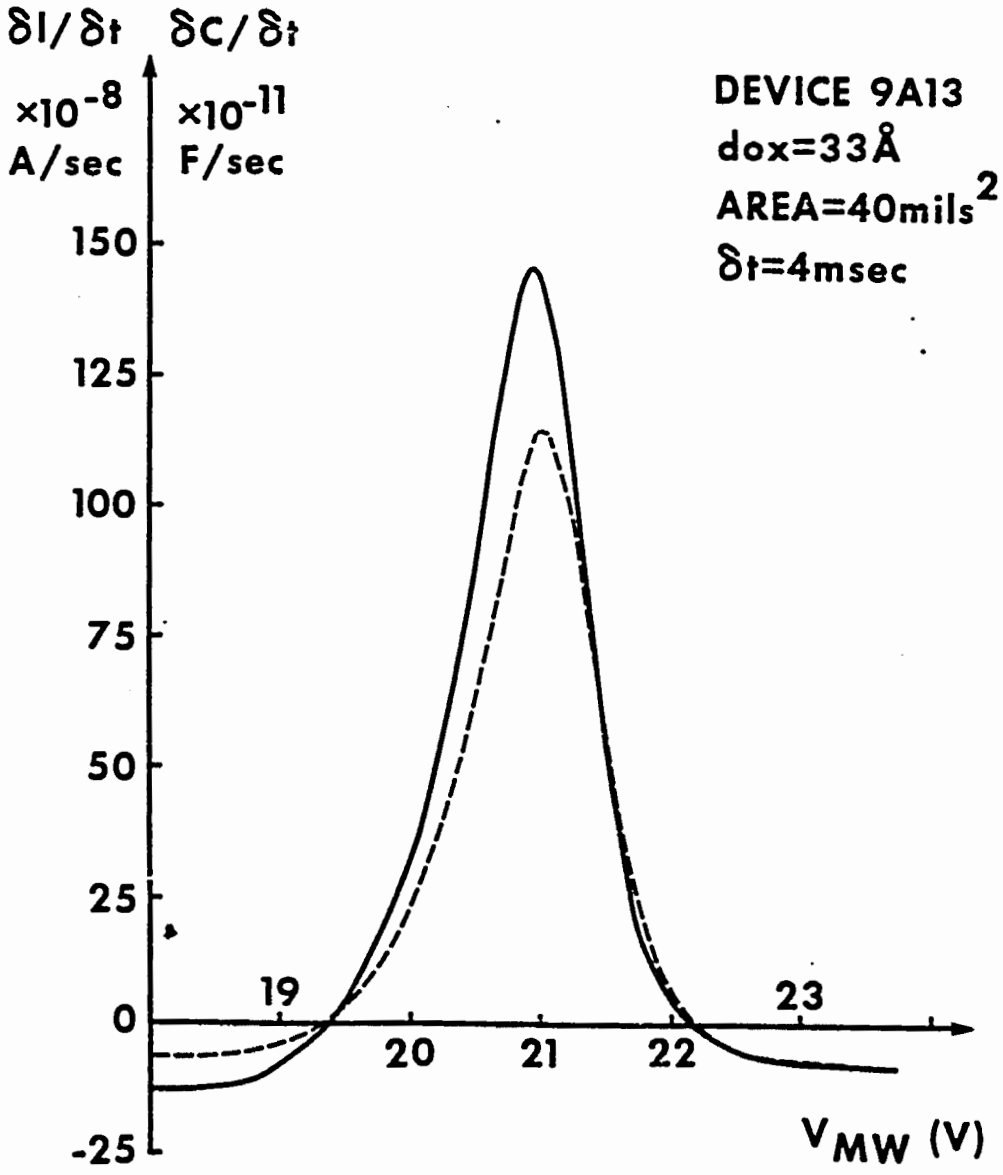
$$\frac{\partial I}{\partial t} = \frac{\partial I}{\partial Q_{INV}} \cdot \frac{\partial Q_{INV}}{\partial t} \qquad 4.4\text{-}5$$

Rearranging and using Eq. 4.4-4 yields

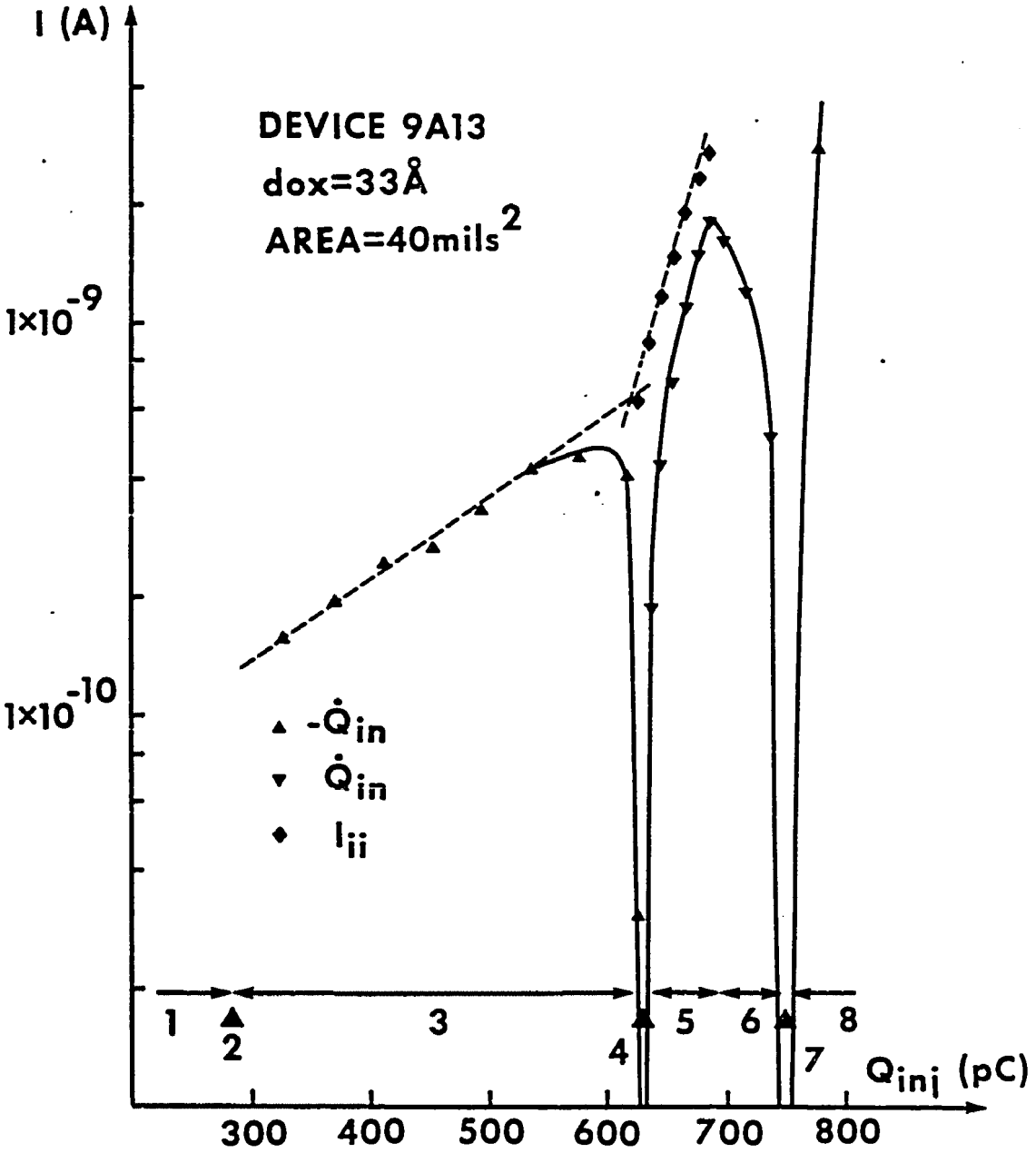
$$\frac{\partial Q_{INV}}{\partial t} = \frac{\delta I}{\delta t} \cdot \frac{Q_{et}}{I_{et}^{o}} \exp\left[-Q_{INV}/Q_{et}\right] \qquad 4.4\text{-}6$$

Thus, by measuring the time rate of change of the device current the time rate of change of the inversion layer can be inferred. A plot of $\dot{Q}_{INV}$ vs. Q_{INV} is shown in Fig. 4.4-8. For Q_{INV} = 550 pC (regime 3), the net rate of loss $\dot{Q}_{INV}$ can be interpreted as the oxide hole transport current. Using a least squares fit, it appears

$$I_{ht} = I_{ht}^{o} \exp\left[Q_{INV}/Q_{ht}\right] \qquad 4.4\text{-}7$$



4.4-7 Differential time rate of change of initial transient device current and initial transient device capacitance as a function of metering well voltage.
Device biased at -3.5 volts.



4.4-8 Time rate of change of inversion layer charge as a function of injected charge packet size. Also shown is the inferred impact ionization current.

with I_{ht}^{o} = 33 pA and Q_{ht} = 208 pC. The ratio between the electron tunneling current and hole leakage current ranges between 5 (for lower values of Q_{INV}) and 35 (for larger values of Q_{INV}). Such a low ratio places some doubt on the dominant hole current oxide transport mechanism being tunneling, since barrier penetration probability for 'hole' tunneling should be orders of magnitude lower as was discussed in section 4.2.

For Q_{INV} ≈ 650 pC (regime 5), the rate of increase $\dot{Q}_{INV}$ can be interpreted as the difference between the impact ionization current I_{ii} and the hole leakage current. Assuming the exponential dependence of hole leakage current on oxide voltage does not vary significantly from regime 3 to regime 5, the fitted relationship of Eq. 4.4-7 can be extrapolated to regime 5 and added to the measured value of Q_{INV}, and an estimate of the impact ionization current obtained. This current is plotted in Fig. 4.4-8. It too can be fitted to an exponential form as

$$I_{ii} = \dot{Q}_{INV} + I_{ht} = I_{ii} \exp [Q_{INV}/Q_{ii}] \qquad 4.4\text{-}8$$

with I_{ii}^{o} = 0.26 fA and Q_{ii} = 42 pC. Comparison of this impact ionization current with the electron tunneling current yields a total impact ionization probability $\bar{P}_{ii} = I_{ii}/I_{et}$ in the range 2% to 5%.

This average impact ionization probability can be compared to both the Shockley (1961) model and the Drummond and Moll (1971) model as discussed in appendix C. The energy of the hot tunneling electrons is approximately

$$E \approx E_M - E_C = q(\phi_s - \phi_m + V_{ox}) \qquad 4.4\text{-}9$$

The Shockley total impact ionization probability expression is

$$\bar{P}_{ii} = 1 - \exp\left[-(E - E_{ii})/mE_{ph}\right] \qquad 4.4\text{-}10$$

where E_{ii} is the minimum energy for which the impact ionization process can occur, E_{ph} is the silicon optical phonon energy ($k \sim 0$) $E_{ph} = 0.063$ eV and m is (in Shockley's model) the ratio of phonon and impact ionization scattering lengths. Since $\bar{P}_{ii}$ is on the order of $0.05 < < 1$, the exponential term in Eq. 4.4-10 can be expanded to yield

$$P_{ii} = [q(\phi_s - \phi_m + V_{ox}) - E_{ii}] \; / \; mE_{ph} \qquad 4.4\text{-}10$$

The oxide voltage drop V_{ox} is $(Q_{INV} + Q_D)/C_{ox}$, which for an MTOS device is approximately $Q_{INV}/C_{ox} + 0.05$ so that Eq. 4.4-11 can be rewritten as

$$\bar{P}_{ii} = \left[q(\phi_s - \phi_m + 0.05 + \frac{Q^o_{INV}}{C_{ox}} - E_{ii}\right]/mE_{ph} + q\Delta Q/mE_{ph}C_{ox} \qquad 4.4\text{-}12$$

where we have written $Q_{INV} = Q^o_{INV} + \Delta Q$.

Experimentally, $\bar{P}_{ii}$ can be written as

$$\bar{P}_{ii} = \frac{I_{ii}}{I_{et}} = \frac{I^o_{ii}}{I^o_{et}} \exp\left[Q_{INV}\left(\frac{1}{Q_{ii}} - \frac{1}{Q_{et}}\right)\right] \qquad 4.4\text{-}13$$

which for $\Delta Q < < (1/Q_{ii} - 1/Q_{et})^{-1}$ becomes

$$\bar{P}_{ii} = \frac{I^o_{ii}}{I^o_{et}} \exp\left[Q^o_{INV}\left(\frac{1}{Q_{ii}} - \frac{1}{Q_{et}}\right)\right] \cdot \left[1 + \Delta Q\left(\frac{1}{Q_{ii}} - \frac{1}{Q_{et}}\right)\right] \qquad 4.4\text{-}14$$

Comparing Eqs. 4.4-14 and 4.4-12 yields

$$mE_{ph} = q\left\{C_{ox}\frac{I^{o}_{ii}}{I^{o}_{et}}\exp\left[Q^{o}_{INV}\left(\frac{1}{Q_{ii}}-\frac{1}{Q_{et}}\right)\right]\left(\frac{1}{Q_{ii}}-\frac{1}{Q_{et}}\right)\right\}^{-1} \qquad 4.4\text{-}15a$$

and

$$E_{ii} = q\left\{(\phi_s - \phi_m + 0.05 - Q^{o}_{inv}/C_{ox}) - \left[C_{ox}\left(\frac{1}{Q_{ii}}-\frac{1}{Q_{et}}\right)\right]^{-1}\right\} \qquad 4.4\text{-}15b$$

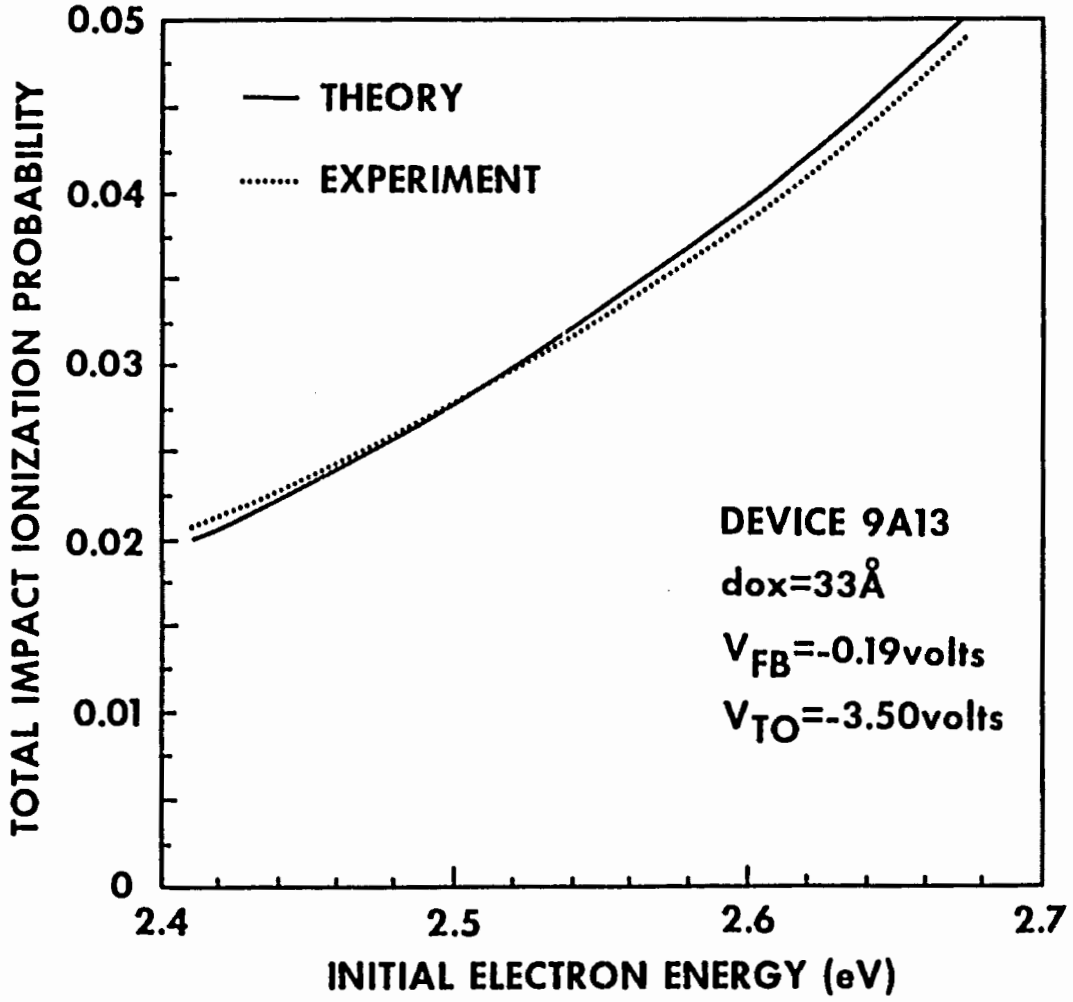
Putting in the fitted values for I^{o}_{ii}, I^{o}_{et}, Q_{ii}, Q_{et} and choosing Q_{INV} = 630 pC we obtain E_{ii} = 2.04 eV and m = 206. The value of E_{ii} is similar to that obtained by Lai (1979) using a surrogate optical generation and capacitance technique, but the value of m is several times larger.

The experimental data can also be compared to the improved impact ionization model of Drummond and Moll, as shown in Fig. 4.4-9. The model is plotted as a solid line and $\bar{P}_{ii}$ as described by Eq. 4.4-13 is plotted as a dotted line. The agreement is remarkably good and supports the assumption made in appendix C that

$$\bar{P}_{ii}(E) = P_{ii}(E_m - E_c)$$

The data can be made to fit more accurately by slight adjustment of the model parameter A (=648) and the Kane band structure exponent (=4.2), but Fig. 4.4-9 displays the unadulterated data.

It is interesting to note that the Shockley scattering length ratio m can be interpreted as the ratio of scattering rates r_{ph}/r_{ii}. Evaluating the latter ratio using the Drummond and Moll model at an energy corresponding to Q^{o}_{INV} = 630 pC (E = 2.45 eV) yields r_{ph}/r_{ii} = 196, which is in close agreement with m = 206 as obtained above. Although it is tempting to interpret (as Shockley does in his phenomenological model) the ratio m as a ratio of real scattering distances, this is misleading.



4.4-9 Total impact ionization probability as a function of hot electron energy as measured experimentally and calculated theoretically (Drummond and Moll, 1971).

For example, putting m = 200 and l_{ph} = 70 Å yields l_{ii} = 1.4 μm. This does not indicate that impact ionization is occurring one micron into the bulk, rather it seems likely that the physical scattering distance into the silicon for impact ionization is the same as for phonon scattering, but the probability that such a lattice collision results in an ionization event is approximately 1/200.

The fitted expressions for I_{ii} and I_{ht} can be used to calculate a semi-empirical metastable operating point corresponding to the threshold level. From section 4.2, the threshold point occurs when $I_{ii} = I_{ht}$. Thus using Eqs. 4.4-7 and 4.4-8,

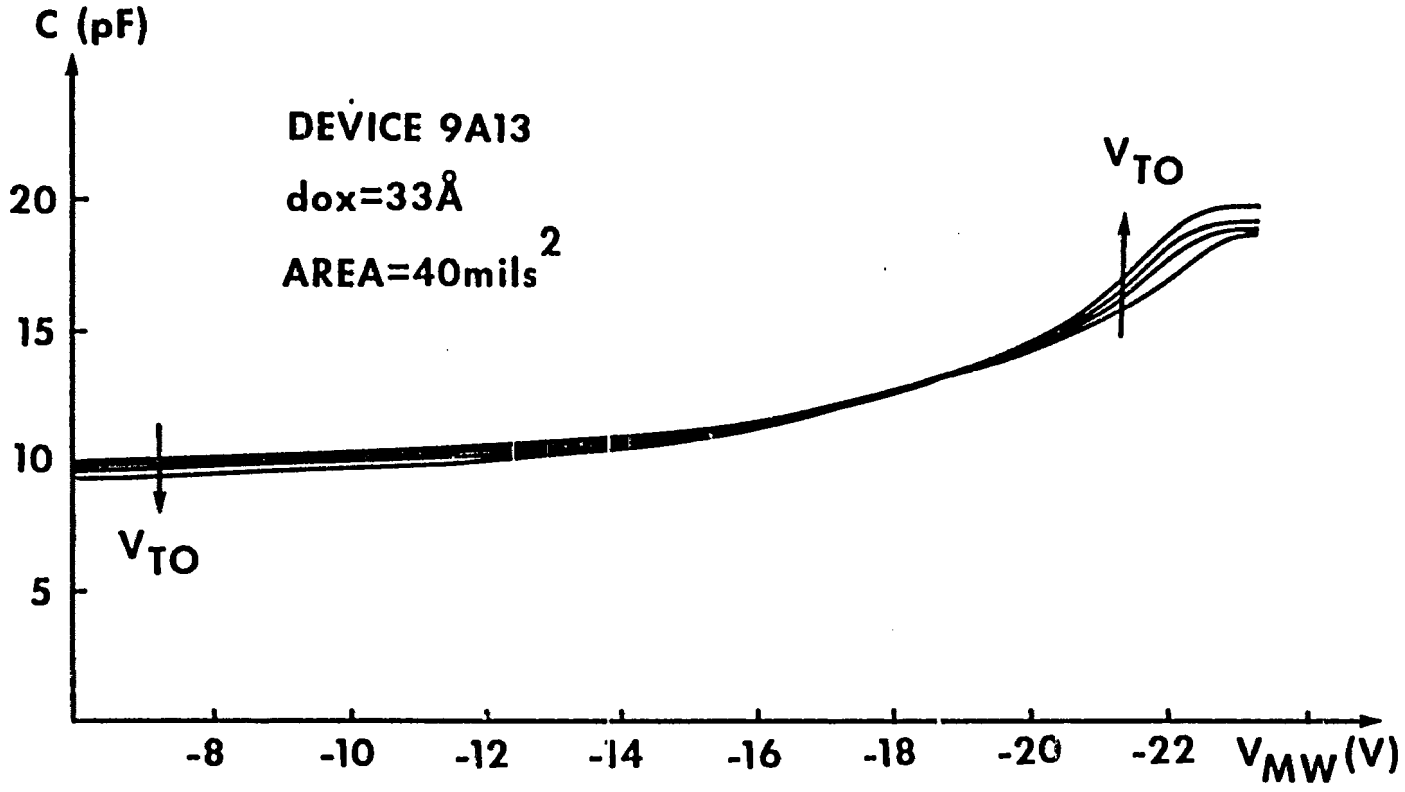
$$Q_T = [1/Q_{ht} - 1/Q_{ii}]^{-1} \ln (I^o_{ii}/I^o_{ht}) \qquad 4.4\text{-}16$$

Putting in the fitted values yields Q_T = 620 pC, or V_{MW} = 19.1 volts. Experimentally, V_{MW} at the threshold is found to be 19.4 volts from Fig. 4.4-5 and to be 19.3 volts from the zero crossing of $\delta I/\delta t$ in Fig. 4.4-7. The discrepancy in the thresholds is due to the experimental uncertainties in determining Q_{INV} and then fitting the result to an exponential. Note that $\delta I/\delta t$ in Fig. 4.4-7 also passes through zero at V_{MW} = 22.2 volts, which implies that the stable ON operating point corresponds to Q_{INV} = 745 pC for V_{TO} = -3.5 volts and d_{ox} = 33 Å.

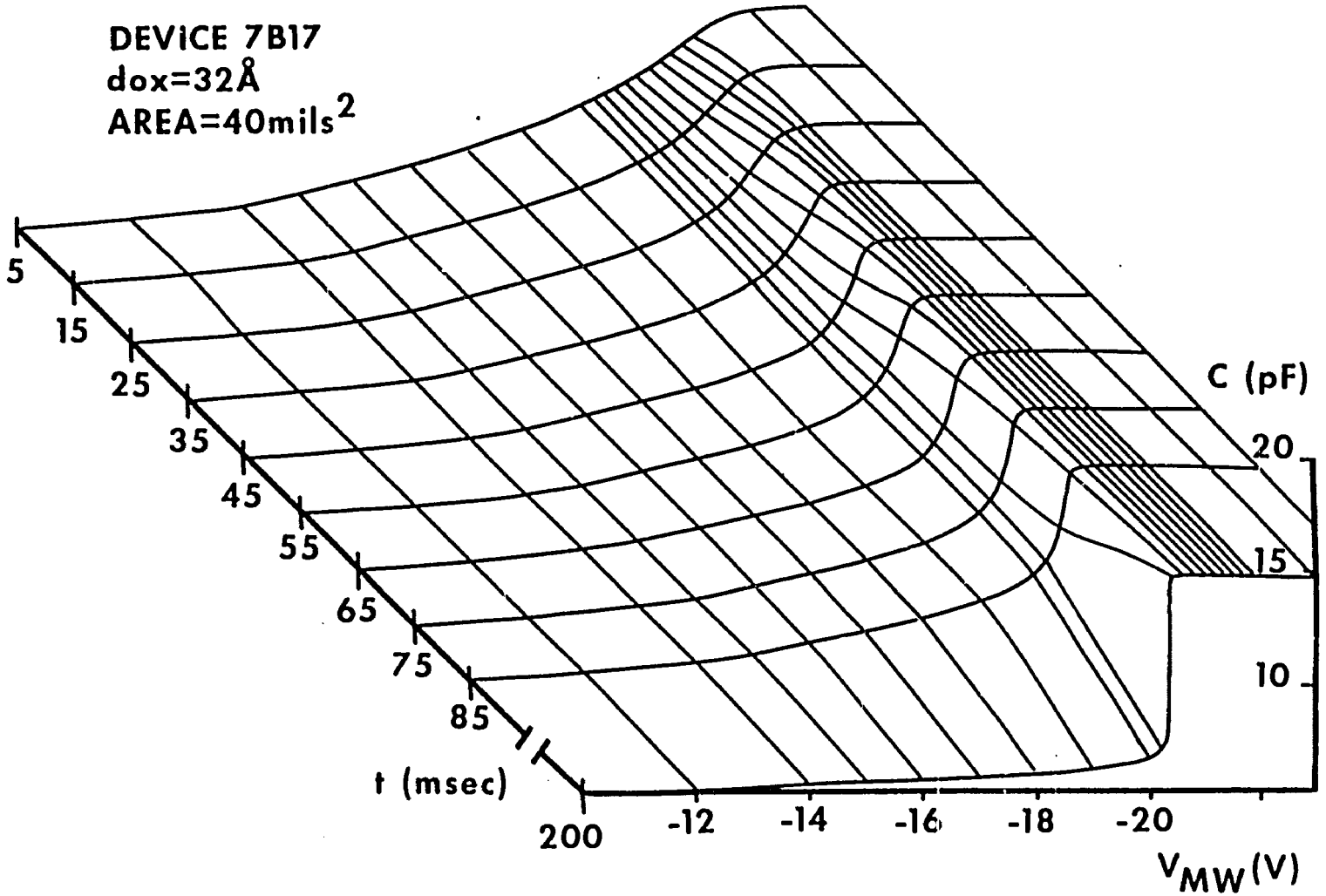
Experimental Results – Device Capacitance

Thus far attention has been confined to the transient MTOS device current as a function of injected charge packet. The transient capacitance at $t = 0^+$ as a function of V_{MW} for device 9A13 biased at various values of V_{TO} is shown in Fig. 4.4-10. A transient capacitance surface is shown in Fig. 4.4-11 for device 7B17 biased at V_{TO} = -3.30 volts, by plotting the measured transient capacitance as both a function of injected charge packet and time.

The capacitance is also time invariant, and indicative of the state



4.4-10 Transient MTOS device capacitance as measured at $t = 0^+$ as a function of metering well voltage, for various values of V_{TO} (-3.2, -3.3, -3.4, -3.5 volts).



4.4-11 Transient MTOS device capacitance as a function of time and metering well voltage.

of the MTOS junction. As such, it also can be used in principle to calculate Q_{INV} by writing

$$\frac{\partial Q_{inv}}{\partial t} = \frac{\partial Q_{inv}}{\partial C} \cdot \frac{\partial C}{\partial t} \qquad 4.4\text{-}17$$

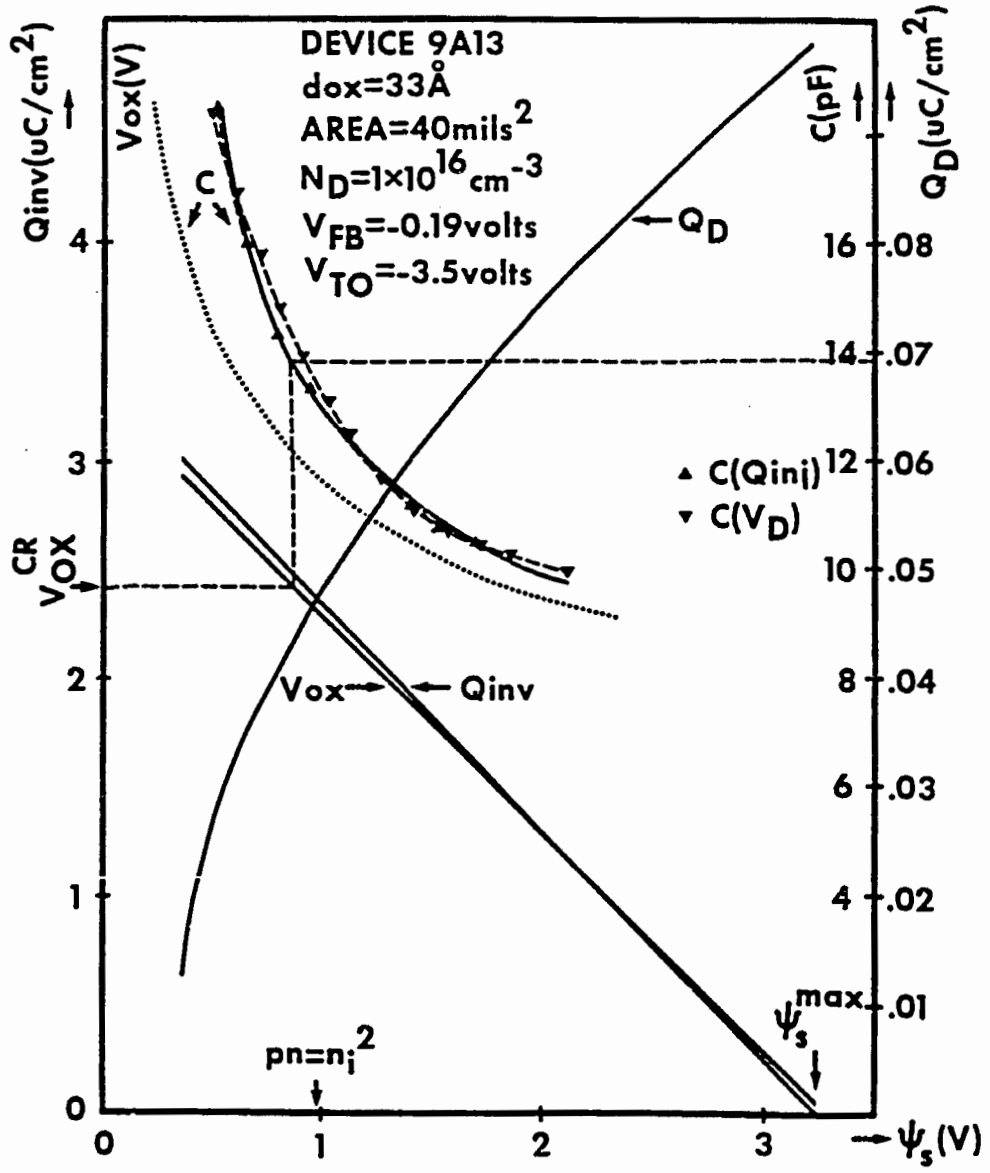
and proceeding in a manner similar to that outlined in the previous subsection. However, the capacitance is neither a strong function of time nor Q_{INV} and is more prone to errors when calculating the terms of Eq. 4.4-17 than calculating the analogous current terms.

The capacitance can also be used to calculate the semiconductor surface potential ψ_s. In appendix B, the zero-current approximation was used to develop a computer model relating carrier concentration profiles to depletion layer depth and surface potential under non-equilibrium conditions. In this way, it is possible to calculate device capacitance as a function of surface potential or inversion layer charge. The latter relationship can be used to compare the model's performance to the experimental results of Fig. 4.4-10.

The total device capacitance $\underline{C}$ is given by

$$\underline{C} = \underline{C}_{ST} + A_{TO}\,(1/C_{ox} + x_d/\varepsilon_o\varepsilon_{ox})^{-1} \qquad 4.4\text{-}18$$

where $\underline{C}_{ST}$ is the stray device capacitance. The stray capacitance can be estimated by examining the device layout. The device consists of a thick oxide bonding pad of area 26 mils^2, a gate oxide n^+ overlap of 2 mils^2, and a gate oxide overlap area of 21 mils^2. Using the thick oxide capacitance of 21 pF/540 mils^2, an accumulated gate oxide capacitance of 83 pF/207 mils^2, and an inverted gate oxide capacitance of 30 pF/207 mils^2, the stray capacitance is estimated to be 4.8 pF. Exercising the computer model for the device capacitance as a function of surface potential (i.e. depletion layer depth x_d as a function of surface



4.4-12 MTOS device parameters as a function of surface potential. Solid curves are results of a computer model employing the zero current approximation. Dotted curve shows capacitance calculated using depletion approximation. Dashed curves show experimental capacitance measured using charge packet injection technique and steady state hole injection technique. Straight dashed lines highlight critical values of metastable state. $C(V_D)$ fitted using $\psi_s = -0.53$ volts. $C(Q_{inj})$ fitted using $C_{ST} = 4.4$ pF.

potential) yields the solid curve of Fig. 4.4-12. In contrast, the capacitance as a function of surface potential calculated using the depletion approximation is shown by the dotted curve. The difference can be ascribed to the potential drop across the inversion layer charge as measured by the horizontal shift in the two curves. The depletion approximation ignores this voltage-drop and the depletion layer depth is given by the constant doping parabolic band bending as

$$x_d = (2\varepsilon_s\varepsilon_o\,\psi_s/qN_D)^{1/2} \qquad 4.4\text{-}19$$

The voltage drop across the inversion layer can be seen to range from 0.25 to 0.35 volts.

The same model can also be used to compute the inversion layer charge as a function of surface potential. This relationship, which is quite linear, is also shown in Fig. 4.4-12 (see Fig. 2.3-3). This curve (line) can be used as an ordinate for plotting the experimentally observed relationship between injected charge packet magnitude and MTOS junction capacitance, as was shown in Fig. 4.4-10. The result is also shown in Fig. 4.4-12 as a dashed curve, which is almost totally obscured by the theoretical curve. To obtain such near perfect agreement, the theoretical stray capacitance was adjusted from 4.8 pF down to 4.4 pF. Nevertheless, the agreement remains satisfying and lends confidence to the model. Also shown in Fig. 4.4-12 is the depletion layer charge Q_D and oxide voltage drop as a function of surface potential.

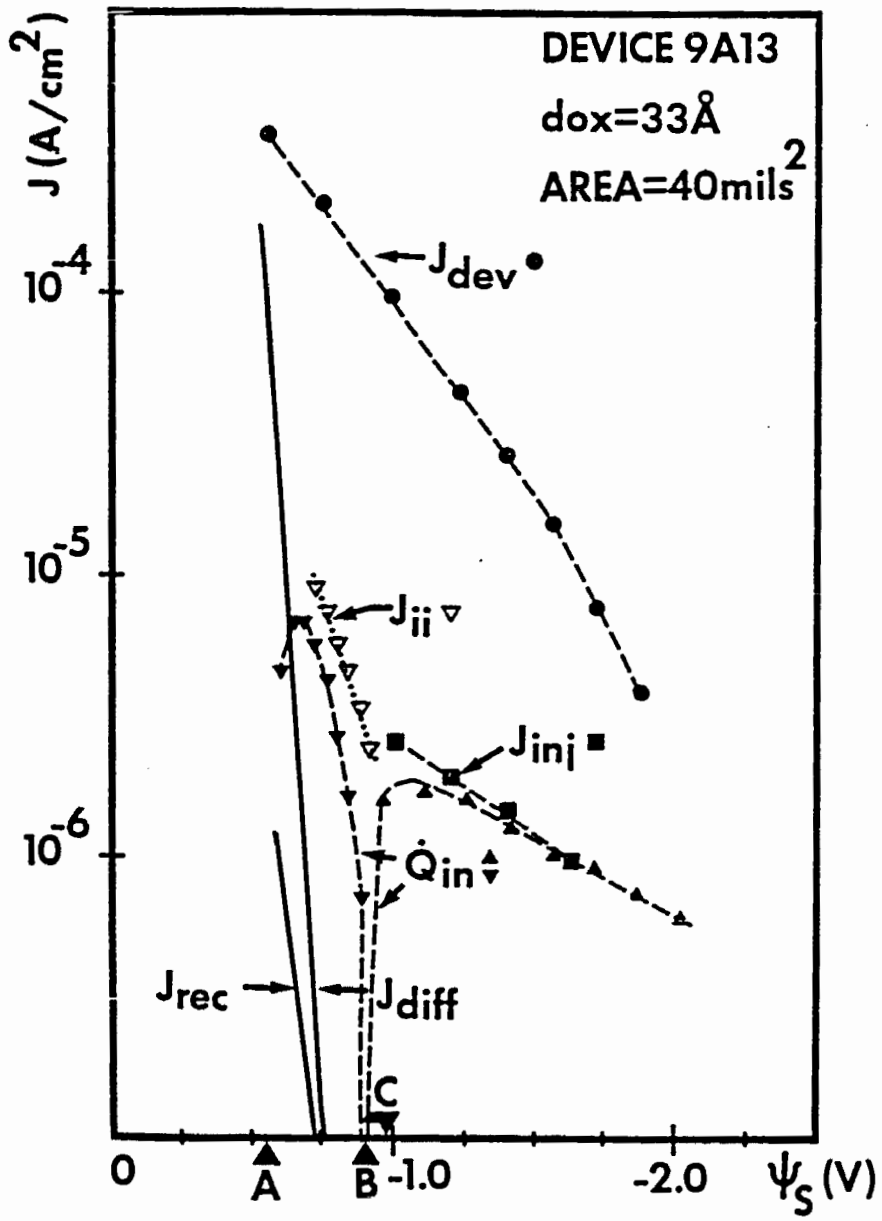
Modelling of ON State Stabilization Currents

Having established confidence in the computer model, the currents measured in the previous sub-section can be plotted as a function of surface potential as shown in Fig. 4.4-13. The computer model can be used to to compute the recombination current (as defined by Eq. 4.2-10) by integrating the Shockley-Read-Hall net recombination rate throughout

the space charge region. This current is plotted in Fig. 4.4-13 using the experimentally measured recombination/generation rate constant $\bar{\tau}_p$ = 8 μsec, and can be seen to be much smaller than most other currents of interest.

In addition to the recombination current, the computer model can be employed to calculate the diffusion current calculated in Eq. 4.2-11 by determining the minority carrier concentration at the depletion region edge. As can be seen in Fig. 4.4-13, the diffusion current is a strong function of ψ_s and is the dominant loss current component of $\dot{Q}_{INV}$ in the ON state. Although it can be argued that the relative sizes of the two loss currents depends on the choice of $\bar{\tau}$, in order for the recombination current (which varies as $[\tau^{-1}]$) to be comparable to the diffusion current (which varies as $[\tau^{-1/2}]$), the average lifetime would have to be smaller by nearly three orders of magnitude. Such an error is doubtful on two counts; first it would mean an enormous difference in τ within 10 mils measured laterally on the surface, and second it would result in a shifted ON state to a significantly lower than observed current level. In fact, experimentally the ON state agrees well with the predicted ON state found by the intersection of the impact ionization current and the back diffusion current.

The relationship between the magnitudes of the recombination current and diffusion current is exactly the same as in a PN junction. In that case, recombination in the space-charge region is virtually ignored, and the primary current is the diffusion current calculated at the depletion region edge. The only difference is that in an MTOS junction, the P side is actually an inversion layer (looks P^+) with the possibility of surface recombination/generation. As was discussed in section 4.2, the contribution of interface traps to recombination has been disregarded.



4.4-13 MTOS device current densities as a function of surface potential for device biased at -3.50 volts. J_{dev} is experimental total device current (electron tunneling current) $\dot{Q}_{INV}$ are points transcribed from Fig. 4.4.8 and measure time rate of change of inversion layer charge. J_{inj} is lateral hole injection current measured from steady state measurement. J_{ii} is impact ionization current. J_{diff} and J_{rec} are the back diffusion and recombination currents respectively as obtained from computer model.

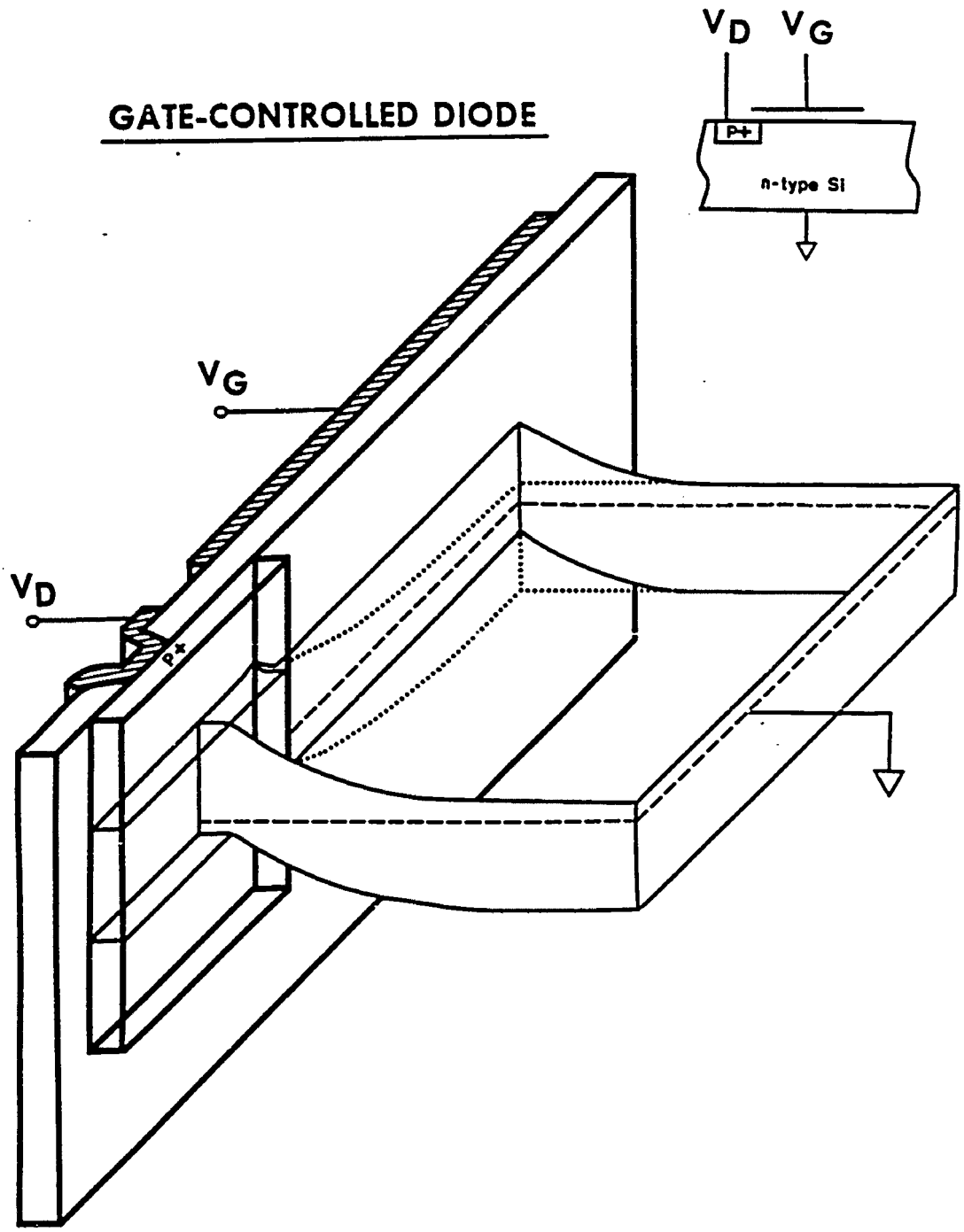
The conclusion that a strong back hole diffusion current stabilizes the ON state is further supported by experimental evidence in section 4.6.

4.5 Steady State Hole Injection using a Diode Controlled Inversion Layer

The electrode structure for the charge packet threshold detector experiment can be employed for steady state hole injection experiments. Biasing the three input structure electrodes at -4.0 volts yields an effective large area gate-controlled diode [Forster and Veloric, 1959]. In this mode, the voltage applied to the diode V_D, controls the surface potential under the three electrodes. If the tunnel oxide electrode is biased at $V_{TO} = -3.5$ volts, its surface potential is similarly controlled. In this way, a three-terminal MTOS device is created in which the MTOS total junction bias V_{TO} and surface potential ψ_s can be independently controlled.

Gate-Controlled Diode Theory

The gate-controlled diode theory is discussed by Grove and Fitzgerald [1966]. In this section, our attention is restricted to the situation where the surface under the controlling gate is inverted. Although Grove and Fitzgerald approached this problem from the point of view of the effect of the gate on the diode I-V characteristics, the effect of the diode voltage on the 'controlling' gate's inversion layer is emphasized here. From this point of view, the structure behaves as a diode-controlled MOS capacitor. A schematic cross-section for this structure showing the energy band bending for a diode-controlled biased and unbiased MOS capacitor is shown in Fig. 4.5-1. In this device, the gate is actually three separate electrodes biased in parallel at a voltage greater than the threshold voltage (-2.0 volts). Since the three electrodes are separated by submicron gaps, the device behaves as though the electrode structure were continuous. The applied diode



4.5-1 Energy bands of ideal gate-controlled diode structure. Dotted bands are for zero gate bias and solid bands are for inversion gate biases. In both cases the diode is grounded. Dashed line shows Fermi level. [After Grove and Fitzgerald].

voltage is in the range $-1.0 \leq V_D \leq +0.2$ volts.

For $V_D = 0$ volts, the pn product throughout the semiconductor is equal to n_i^2, and the surface potential under the gated n-region is the equilibrium inverted surface potential ψ_s^o. In the lateral direction between the p^+ region and the n region, the band bending is just such that minority carrier drift balances minority carrier diffusion. The p^+ region is slightly accumulated due to the negative gate bias, and the n region is inverted. The complexity of this structurally simple system should not be underestimated.

When $V_D < 0$ is suddenly applied to the diode, the initial effect is to raise the valence band in the p^+ region relative to the valence band in the inverted n region. This causes a net drift current of holes from the electrostatically unfavorable inversion layer into the p^+ region, resulting in a raising of the inversion layer valence band until drift/diffusion balance is reestablished. The surface potential in the inversion layer ψ_s is approximately equal to

$$\psi_s = V_D + \psi_s^o \qquad 4.5\text{-}1$$

The semiconductor under the gate is in a state of slight deep depletion ($pn < n_i^2$) and the diode must extract the thermal hole generation current from the MOS capacitor in order to maintain the surface potential according to Eq. 4.5-1. If one measures the reverse biased PN junction current, it is the sum of an isolated PN junction reverse bias current and the thermal generation current of a slightly deep depleted MOS capacitor. This effect can be used to probe the MOS dark current, but the MOS gate area must be large enough to permit the observation of dark current in the 10 nA/cm^2 range [Rabbani and Lamb, 1983].

For small positive values of V_D, Eq. 4.5-1 continues to describe the surface potential. In this situation, the diode is injecting holes

toward the MOS surface in order to support a super-inversion layer ($pn > n_i^2$). The injected holes subsequently recombine in the space-charge region or diffuse away into the bulk. Under either forward or reverse diode bias, the effect of the MOS inversion layer coupling is to increase the apparent area of the PN junction, because the inversion layer of holes acts exactly as a virtual p^+ region. Such an effect is used in inversion layer solar cells [for example, Okuyama and Ma, 1984].

Diode-Controlled MOS Capacitor Experiment

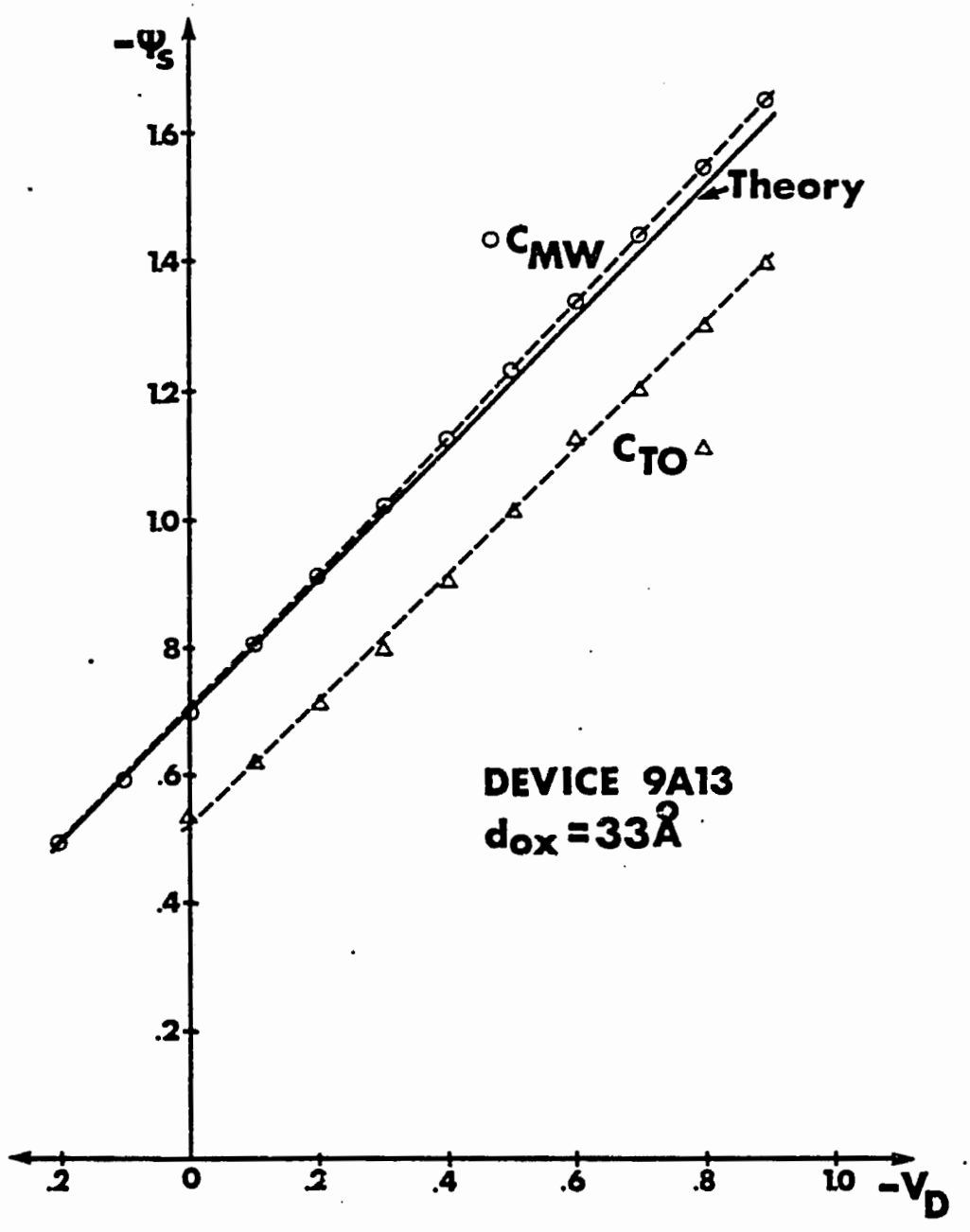
To check this simple theory exemplified by Eq. 4.5-1, the capacitance of the metering well electrode was monitored as a function of V_D in order to measure the surface potential of the capacitor. The bias applied in parallel to the three electrodes (V_{XI}, V_{MW}, V_{XO}) was -4.0 volts. The surface potential was determined from the capacitance through the use of Eqs. 4.4-18 and 4.4-19 (which neglects the voltage drop across the inversion layer; permissible for gate oxides 555 Å thick). This surface potential is plotted as a function of diode voltage in Fig. 4.5-2. Experimentally it appears that

$$\psi_s = 1.06\ V_D - 0.70 \text{ (volts)} \qquad 4.5\text{-}2a$$

The value $\psi_s^o = -0.70$ volts is almost the classic equilibrium inversion layer surface potential ($2\phi_n$) so that the agreement between the Grove and Fitzgerald theory and the confirmation experiment is excellent.

The same experiment can be performed using the MTOS capacitor rather than the metering well electrode. The three input structure electrodes remain biased at -4.0 volts and the MTOS junction is biased at -3.5 volts as its capacitance is monitored as a function of V_D. The surface potential is calculated using the results of the computer model, and is also plotted in Fig. 4.5-2. Experimentally,

$$\psi_s = 0.97\ V_D - 0.53 \text{ volts} \qquad 4.5\text{-}2b$$



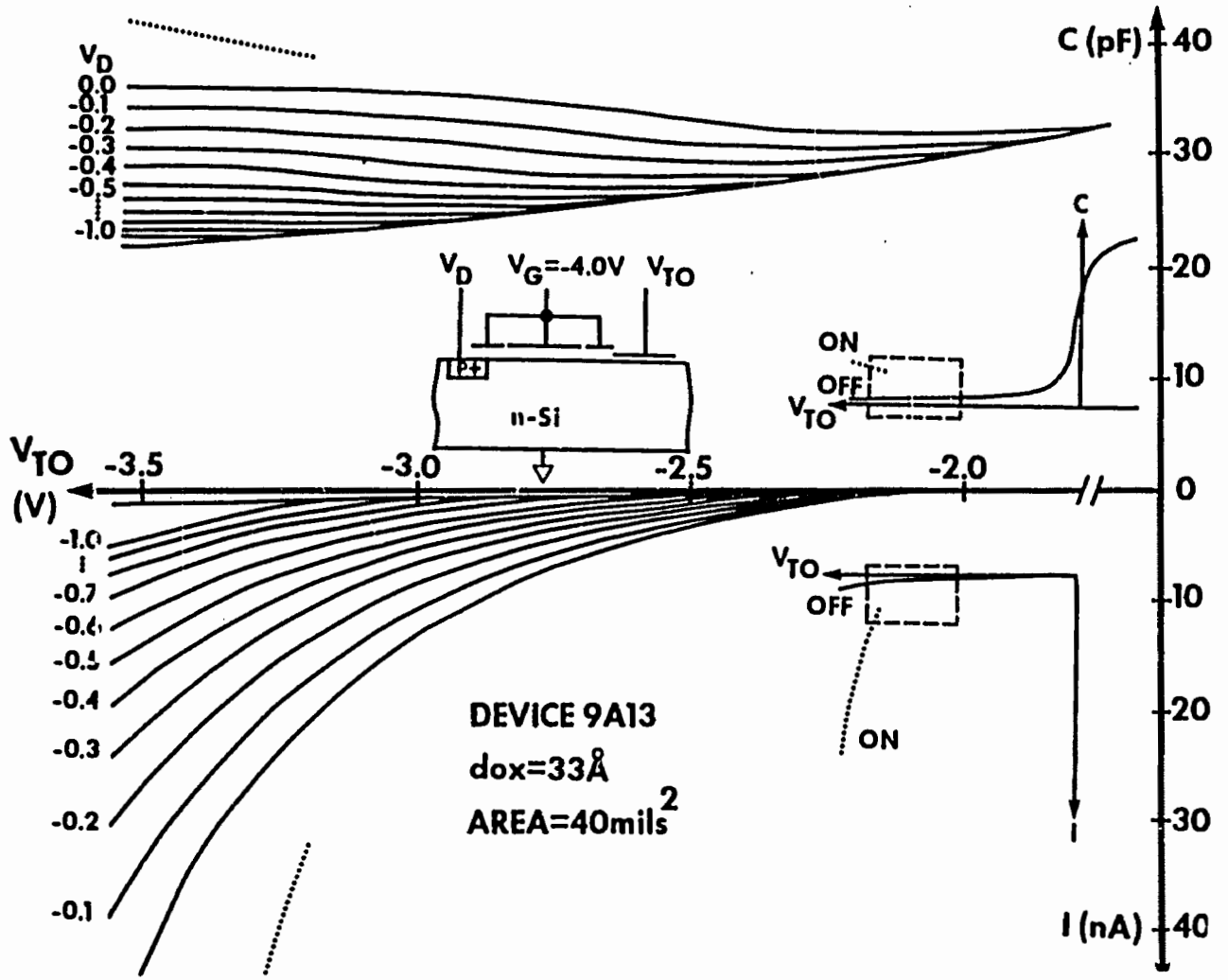
4.5-2 MOS surface potential as a function of diode voltage as determined from diode-controlled MOS capacitor experiments for both the metering well (C_{MW}) and tunnel oxide device (C_{TO}).

This value of ψ_s^o is considerably smaller than the value $\psi_s^o = -0.98$ volts predicted by setting $pn = n_i^2$ and employing Eqs. 4.2-2 and 4.2-13. The reason for this discrepancy is not presently clear. The major differences between the tunnel oxide device and the gate oxide device are that a current is flowing through the tunnel oxide with possible impact ionization generation of holes occuring for ψ_s smaller than -1.5 volts, and that the surface hole concentration is much higher for the tunnel oxide device. The former difference could account for part of the discrepancy, but the latter difference would predict the opposite effect. Another possibility is that lateral minority carriers are responding to the A.C. probe signal resulting in increased capacitance [T-P. Ma, private communication]. The MTOS junction may be more sensible to such lateral effects due to the high value of C_{ox}, so that such a hypothesis is reasonable but presently unproven.

The outstanding observation of this simple diode-controlled MOS capacitor experiment is that the diode voltage controls the MOS capacitor and MTOS capacitor surface potential in a linear and nearly unity gain fashion.

Diode-Controlled MTOS Junction

In Fig. 4.5-3, the capacitance-voltage and current-voltage characteristics of the MTOS junction are shown for various values of diode voltage V_D. In all cases $V_{XI} = V_{MW} = V_{XO} = -4.0$ volts except for the dotted lines which indicate the characteristics for $V_{XO} = 0$. The effect of fixing ψ_s once the MTOS device is inverted is to flatten the C-V characteristic. This is similar to the case of thick oxide MOS capacitors where in inversion, ψ_s becomes fixed at some level ($\sim 2\,\phi_n$). In this experiment though, the 'fixed' level of ψ_s can be varied by altering the diode voltage. By pinning the surface potential, additional voltage applied to the thin oxide electrode falls across the



4.5-3 MTOS device current and capacitance as a function of MTOS junction bias for various values of controlling diode voltage. Insets show biasing arrangement and enlarged portions of C-V and I-V characteristics. Dotted lines show isolated MTOS device ON characteristics.

oxide. This is seen in the I-V characteristics by the exponential increase in electron tunneling current for increasing V_{TO}.

As was discussed above, fixing $V_D = 0$ volts should result in $pn = n_i^2$ throughout the semiconductor. The C-V curve corresponding to $V_D = 0$ can be interpreted as the MTOS C-V for a thin (33 Å) 'non-tunnel' oxide. By biasing $V_D = 0$, the MTOS state is determined on the MTOS phase diagram by the $pn = n_i^2$ trajectory. Varying V_D results in operating the MTOS device along a diagonally shifted but nearly parallel curve in state space.

Instead of fixing the diode voltage and sweeping the tunnel oxide junction bias, the latter can be fixed and the diode voltage swept. This measurement is shown in Fig. 4.5-4. The substrate current increases exponentially with V_D as expected since $\Delta V_D = \Delta V_{OX}$. Experimentally we observe that

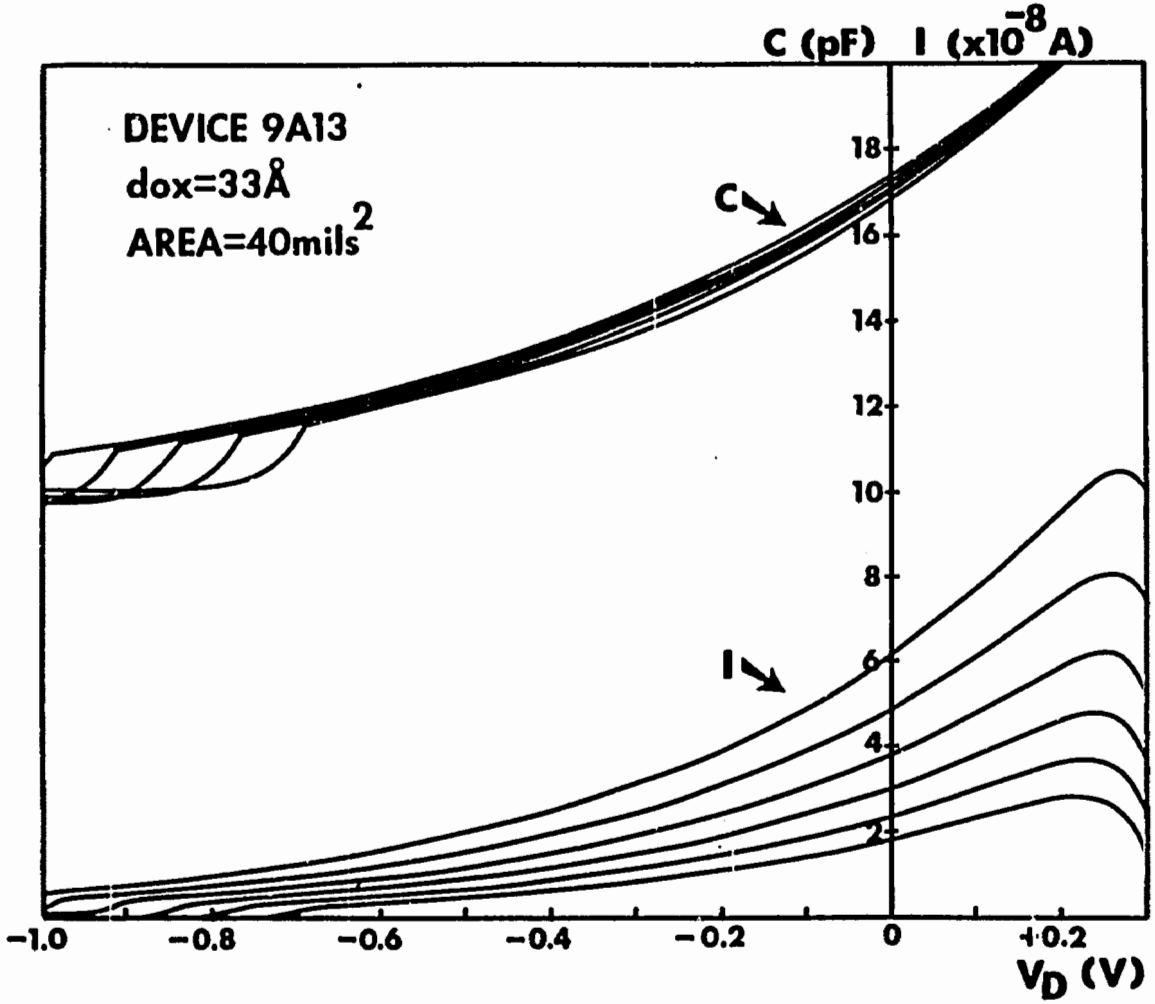
$$I(V_D, V_{TO}) = I(V_D + \Delta V, V_{TO} + \Delta V) \qquad 4.5\text{-}3$$

This behavior is indicative that the device current (i.e. electron tunneling current) is a function only of V_{OX}, since

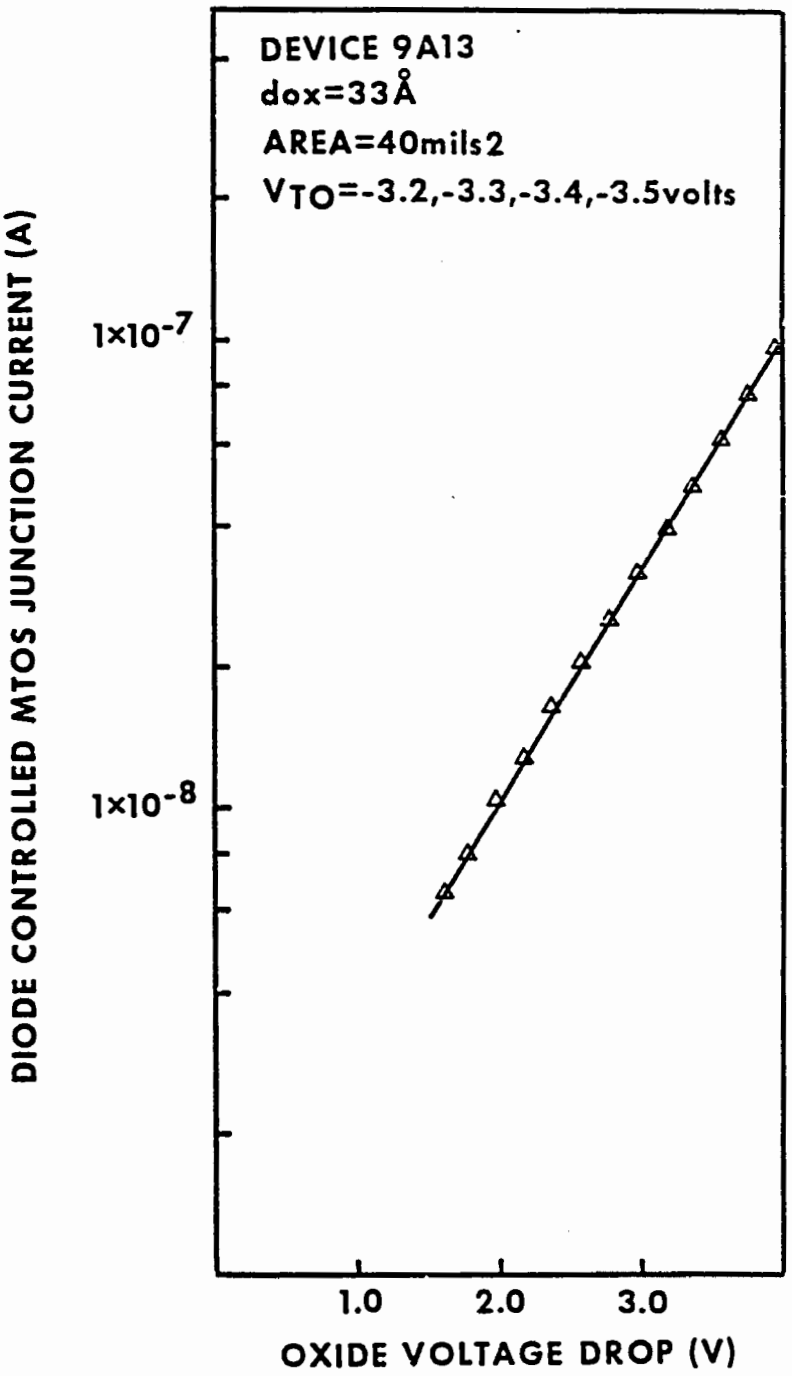
$$V_{OX} = V_{TO} - V_D - (\psi_s + V_{FB}) \qquad 4.5\text{-}4$$

and equal increases in V_{TO} and V_D cancel. This same behavior was alluded to in the discussion of Fig. 4.4-6.

In the experimental set-up for Fig. 4.5-4, the device current is measured at the substrate, and can be interpreted as an MTOS junction current for small levels of diode current. However, for positive values of V_D, the diode current grows rapidly leading to the downturn in total substrate current as seen in Fig. 4.5-4. For values of V_D before the diode current becomes large, the device current is plotted as a function of V_{OX} using $\psi_s^o = -0.53$ volts in the semi-log plot of Fig. 4.5-5. This



4.5-4 MTOS device capacitance and substrate current as a function of controlling diode voltage for various values of MTOS junction bias. In order of increasing capacitance and current, junction bias ranges from -3.0 volts to -3.5 volts in 0.1 volt increments.



4.5-5 Semi-log plot of MTOS junction current as a function of tunnel oxide voltage as measured by diode-controlled MTOS junction technique. Data transcribed from Fig. 4.5.4.

figure can be compared to Fig. 4.4-.6. The slope in the latter is slightly steeper, but may be due to the uncertainties in the indirect calculation of oxide voltage in either or both of the two experiments.

The MTOS capacitance as a function of V_D for various values of MTOS junction bias is also shown in Fig. 4.5-4. This data was used in preparing Fig. 4.5-2, and was replotted in Fig. 4.4-12 as a function of surface potential. In order to obtain good agreement between the experimental capacitance curve of Fig. 4.4-12 obtained from the charge packet injection experiments, and this latter data, ψ_s^o was chosen to be -0.53 volts. This fitting process is the origin of the value of ψ_s^o quoted previously in this section.

In both the capacitance and current characteristics of Fig. 4.5-4, there is an odd looking distortion for $V_D \simeq -0.8$ volts (position depends upon V_{TO}). The distortion is one of lower capacitance (i.e. deeper depletion) and lower current (less oxide voltage drop, i.e. deeper depletion), than one might expect by extrapolating the smooth curves. This distortion is most likely due to the loss of the coupling inversion layer under the three input structure electrodes. The loss is due to the relatively large reverse bias diode voltage, making it energetically unfavorable for any minority carriers to remain in the MOS capacitor inversion layer. The negative diode voltage pulls out all the holes under the input structure electrodes. The MTOS junction then relaxes to its deep depleted OFF state and is no longer influenced by the diode. This behavior is discussed in more detail in section 4.6.

N-Type Negative Resistance caused by Hot Electron Impact Ionization

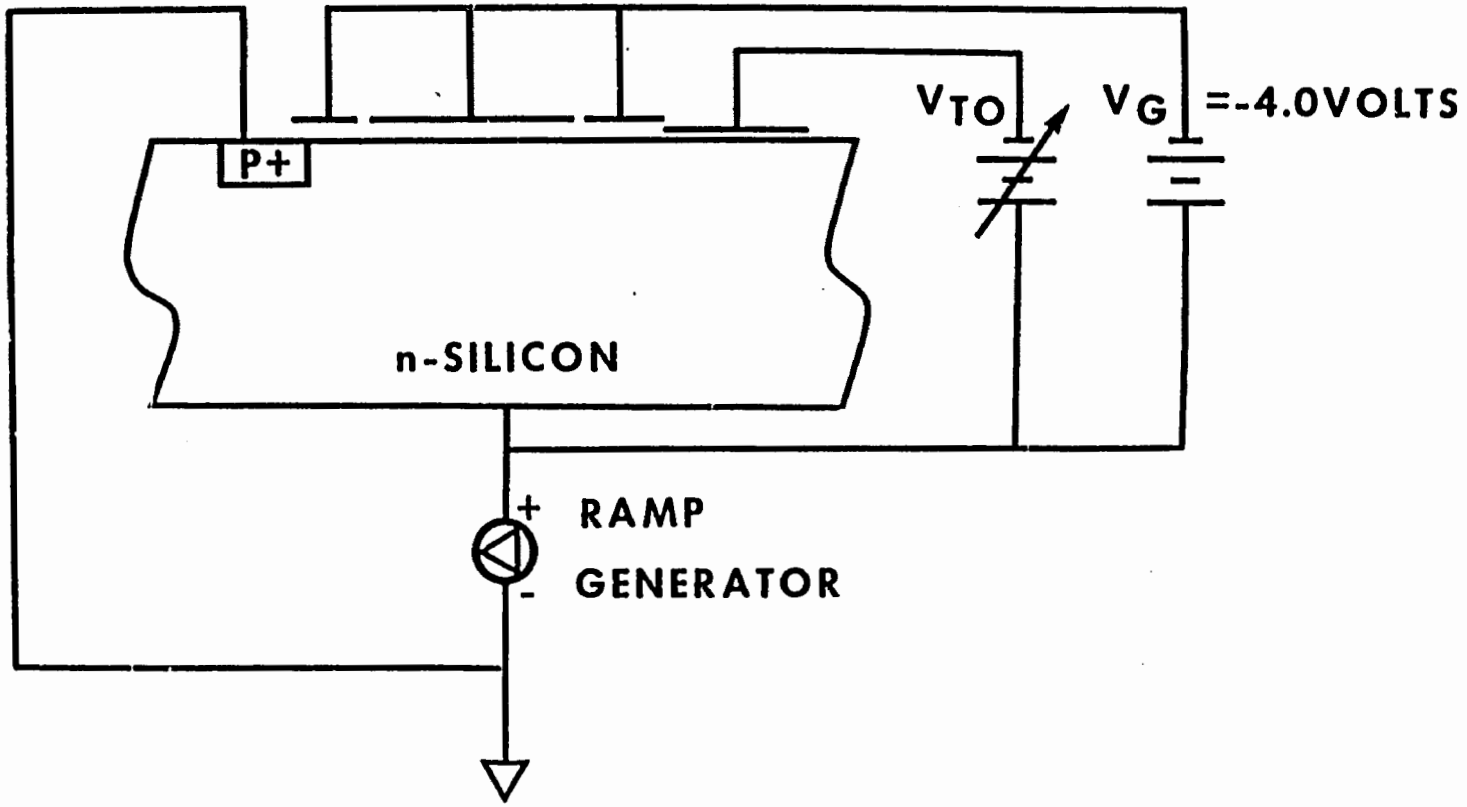
When current through the PN junction is measured in the diode-controlled MTOS junction experimental set-up, an interesting N-type negative resistance is observed. This measurement is in principle

straight forward, but in actuality Murphy's Laws of Instrumentation seem to be enhanced in the experimental configuration, which has resulted in the 'death' of many devices, including 9A13. Fortunately, it was possible to make one measurement of 9A13 before it passed away using the experimental set-up shown in Fig. 4.5-6. Because most ammeters with picoampere to nanoampere sensitivity operate best in a single-ended mode, the diode was grounded through the ammeter, and the substrate bias was ramped. To maintain constant bias across the input structure and tunnel oxide electrodes, a floating supply was used between the substrate and the electrodes.

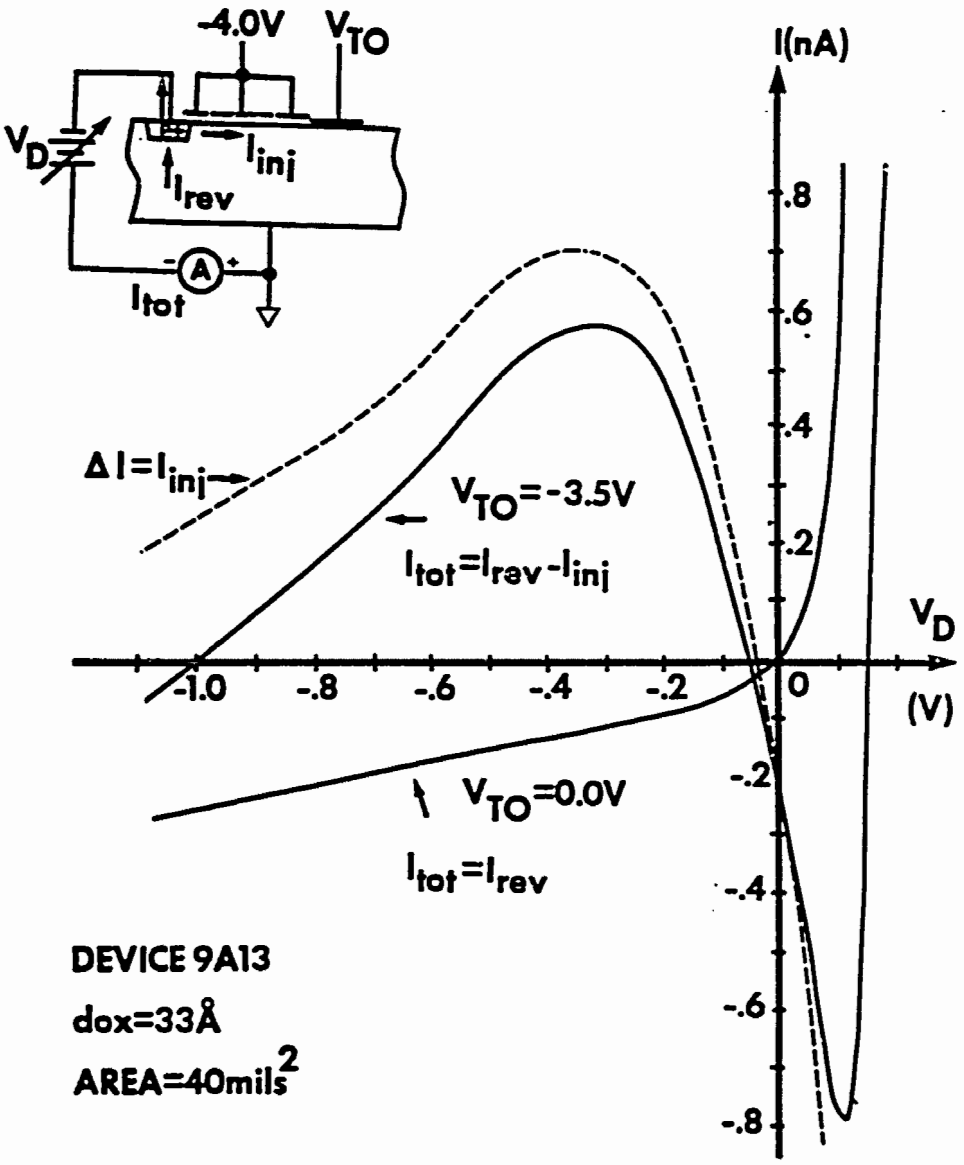
The PN junction current was measured with the input structure electrodes biased at -4.0 volts and MTOS junction shorted to the substrate. Then, the same measurement was repeated but with the MTOS junction biased at -3.5 volts. These measurements are shown in Fig. 4.5-7. The difference in the two measured currents is due to the lateral flow of holes between the MTOS surface and the PN junction.

The lateral injection current as measured for V_D between -1.0 and -0.4 volts can be interpreted as the tunnel oxide hole leakage current. In this surface potential regime, $\dot{Q}_{INV} = I_{lat} - I_{ht}$, and in the steady-state measurement of Fig. 4.5-7, $\dot{Q}_{INV} = 0$. The hole leakage current measured by this method is shown in Fig. 4.4-13. It is seen that the agreement between this latter measurement and the hole leakage current obtained using the charge packet transient technique is quite reasonable. Such consistency tends to validate both approaches and lends confidence to the results.

In the bias regime $V_D \geq -0.4$ volts, the oxide voltage has reached a sufficient level to cause hot electron impact ionization generation of holes. This hole generation partially resupplies holes lost to oxide



4.5-6 Experimental set-up for measuring lateral injection current.



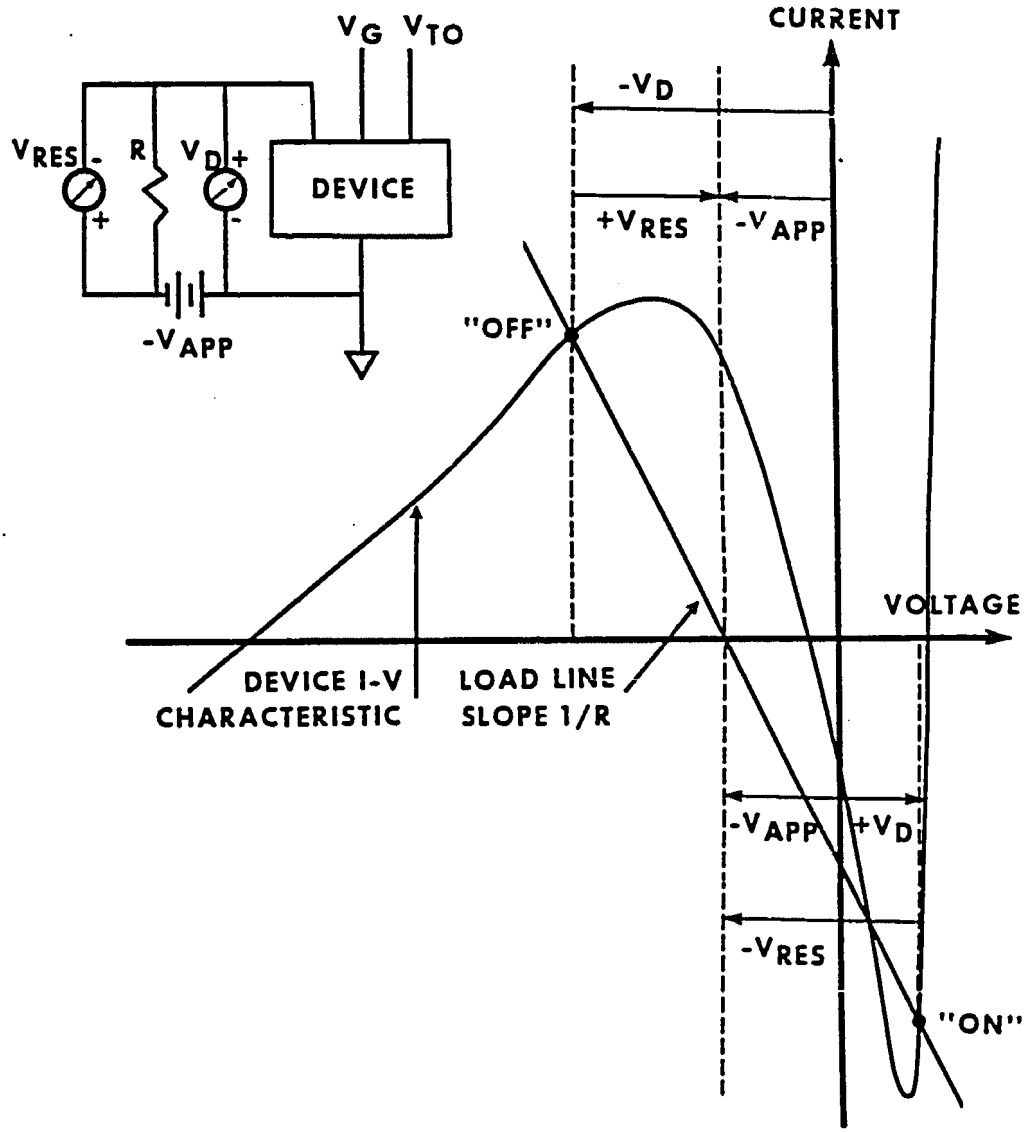
4.5-7 Negative resistance controlling diode current as a function of diode voltage for MTOS junction grounded to substrate and for MTOS junction biased at -3.50 volts. Dashed line shows difference in two currents and interpreted as lateral hole current.

leakage and reduces the lateral hole current requirement from the diode, and the diode current begins to drop. In this regime,

$$\dot{Q}_{INV} = 0 = I_{lat} + I_{ii} - I_{ht}$$

As the impact ionization current continues to grow, the net diode current decreases in a N-type negative resistance fashion, and even changes sign. When J_{ii} is equal to J_{ht}, the lateral current is zero and the diode current is the same as when the MTOS junction is unbiased. The surface potential for this condition is less than that obtained from the metastable operating point of the charge packet injection experiment. The difference is not well understood, but may be due in part to the finite impedance of the ammeter used in Fig. 4.5-6, which may distort the I-V characteristic, so that the charge packet experiment is felt by the author to be more reliable. This same lower-than-expected surface potential phenomenon was encountered in the diode-controlled MOS capacitor experiment described in the previous subsection.

The total diode current, which also decreases in a negative resistance fashion from $V_D \geq -0.4$ volts to $V_D \simeq +0.1$ volts, increases again beyond the latter limit and changes sign a second time due to the additive effect of the PN junction current, and completes the N-type characteristic. Such a characteristic can be used for a bistable switching circuit if a load of sufficient value is placed in series with the PN junction. This is shown in the load time analysis diagram of Fig. 4.5-8 (wherein the load is chosen to be a passive resistor, as opposed to an FET active load). By choosing the proper load, it is even possible to apply zero volts to the circuit yet have two stable current states, one with positive current and the other with negative current. The circuit is switched by causing the MTOS junction to become super-inverted beyond the threshold level (regime 4) such as by illumination



4.5-8 Negative resistance load line analysis diagram showing bistable operating points of inset circuit.

or any of the other techniques described in this work.

Although such a circuit may have many interesting applications, the switching speed of this circuit depends upon the strength of the negative resistance, which depends upon the hole generation current due to impact ionization. Raising the junction bias does this but is limited by the breakdown voltage. By increasing the area of the device, the total hole generation rate is increased but the capacitance is also increased. It is preferrable to increase the impact ionization probability and simultaneously reduce the capacitance. Although preliminary, a collaborative effort between the author and another investigator has shown that the same N-type negative resistance curve may be obtained using a thicker ($\sim$90 Å) oxide biased into the Fowler-Nordheim tunneling regime. In this case, the specific oxide capacitance is reduced three-fold and the impact ionization probability is increased approximately thirty-fold to nearly 100% or larger. Thus the negative resistance switching capability is enhanced by approximately two orders of magnitude.

The author regrets that due to time constraints, the ramifications of this newly discovered N-type negative resistance remain largely unexplored. Perhaps future investigations will reveal an optimum way to employ this phenomena in a practical circuit.

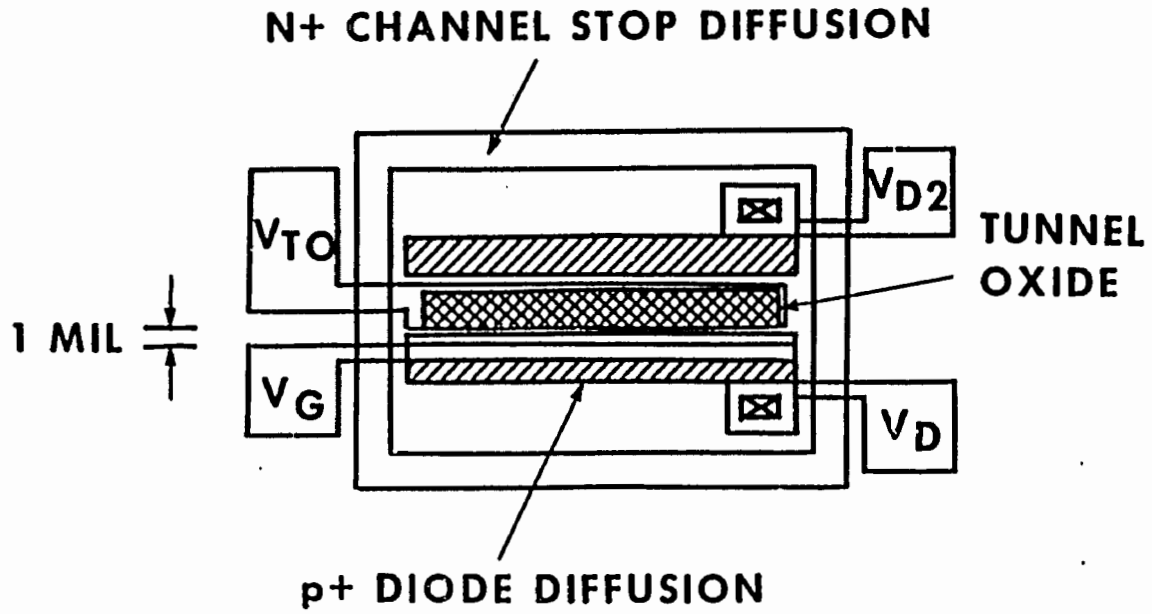
4.6. Effect of Control Gate Voltage on PN Junction - MTOS Junction Coupling

In the previous section, the PN junction diode voltage control of the MTOS junction surface potential was examined. In the cases so far discussed, the control gate was either biased such that a strong inversion layer was present or was clamped off so that the diode was decoupled from the MTOS junction. In this section the intermediate range of control gate biases will be examined.

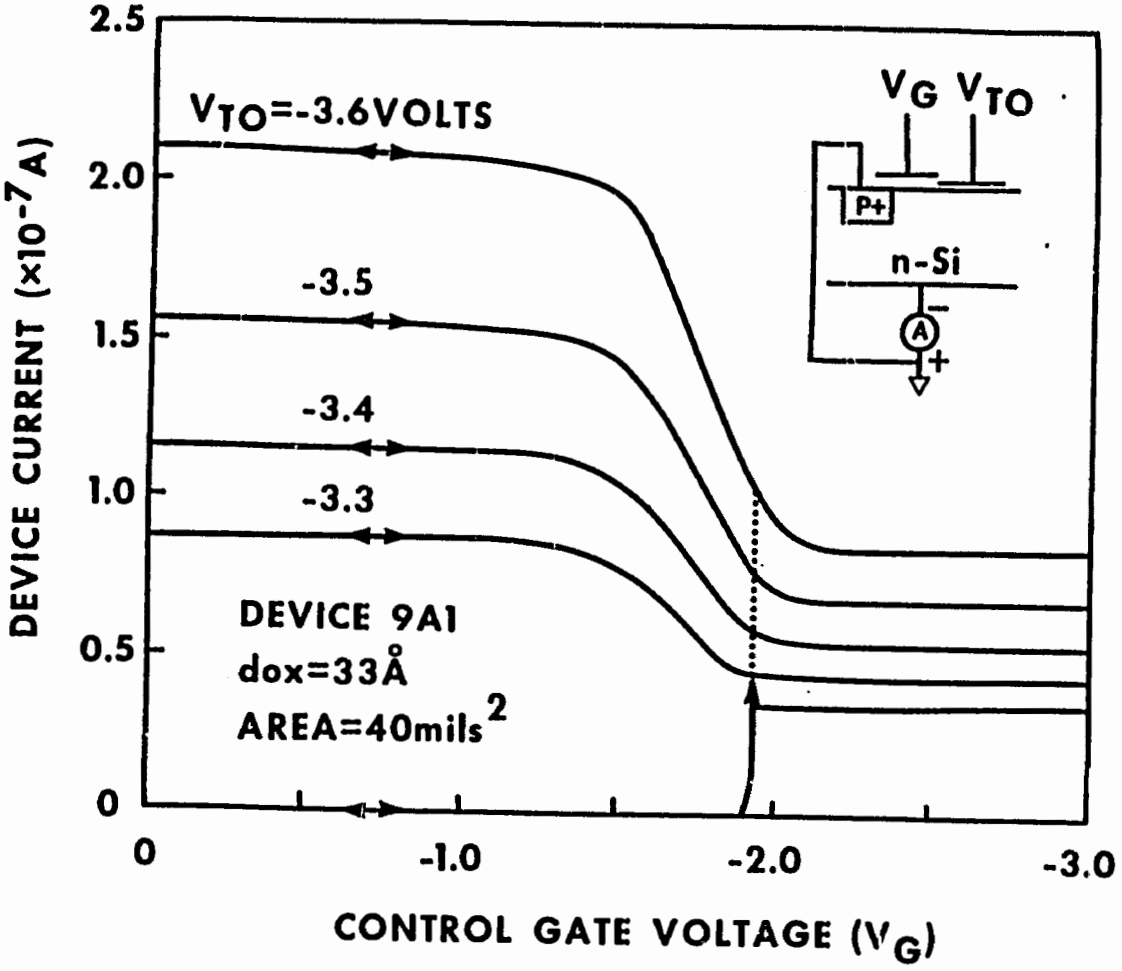
For the investigation, a somewhat simplified experimental structure which was fabricated on the 8205 test chip and whose layout is shown in Fig. 4.6-1 is used. The structure is elongated in the direction transverse to current flow, compared to the structure discussed previously in this chapter, and there is only one control gate as compared to three. The PN junction is located only one mil away from the MTOS junction so as to enhance the coupling of the two devices by lateral diffusion when the control gate is OFF.

There were four basic sets of current-voltage measurements that were performed on this device structure. Since there are three independent terminals (substrate grounded), one terminal's bias is fixed, and current measured as the second is swept and the third bias used as a parameter. There are three interesting places to measure current which are the substrate, the diode, and the MTOS gate. However, the substrate current is almost equal to the MTOS junction current, and the diode current is only easily measured when the diode voltage is fixed and equal to the substrate voltage. Thus, four sets of measurements have been made, and these measurements are sufficient for understanding this complex semiconductor device structure.

In the first measurement performed, the diode is grounded to substrate, and the substrate current is monitored as a function of control gate voltage, for various values of MTOS junction bias. The results of this measurement are shown in Fig. 4.6-2. The gate voltage is ramped to -3.0 volts with the MTOS junction initially in the OFF state. At approximately -1.9 volts, the lateral subthreshold control gate channel conductance is sufficient to allow lateral hole injection from the diode to take place. The MTOS junction current rises rapidly and then switches into a high current state. At this point, the coupling between the diode and MTOS junction has changed. The MTOS



4.6-1 Layout of MTOS junction - PN junction coupling experiment with intermediate control gate. Interelectrode gap exaggerated.



4.6-2 Substrate current as a function of control gate voltage for various values of MTOS junction bias.

junction inversion layer is limited by the subthreshold diffusion of holes from the MTOS super-inversion layer through the control gate channel to the diode.

Increasing the control gate voltage results in the formation of an inversion layer which strongly couples the MTOS junction surface potential to the diode voltage in the previously discussed manner of a diode-controlled MOS capacitor. The MTOS current is independent of control gate voltage as long as the coupling inversion layer is present, and the current is determined by the tunnel oxide voltage drop as described by Eq. 4.5-2.

Reduction of the control gate voltage below -2.0 volts with the MTOS junction in a high current state (V_{TO} must be above the holding voltage V_H), retards the ability of the diode to control the MTOS surface potential through the lateral current. The MTOS inversion layer grows until the back diffusion loss term makes up for the retarded control. As will be seen in Fig. 4.6-6, the lateral diffusion current to the diode actually grows as the control gate bias is reduced from -2.0 volts to -1.5 volts. This counter-intuitive effect is due to a differential negative 'resistance' relationship. As the lateral channel is clamped off, the MTOS inversion layer grows resulting in increased electron tunneling current and impact ionization currents. The increase in super-inversion results in an increase in lateral diffusion current. Thus, clamping off the lateral channel results in increased lateral diffusion current. The percentage of the lateral diffusion current exiting the MTOS super-inversion layer collected by the diode is also a function of control gate voltage, as the number of majority carriers in the control channel is regulated from depletion toward flat-band.

For control gate voltages less than ~ -1.5 volts, the MTOS junction

current is nearly independent of control gate voltage, even though the internal inversion layer currents continue to rearrange themselves.

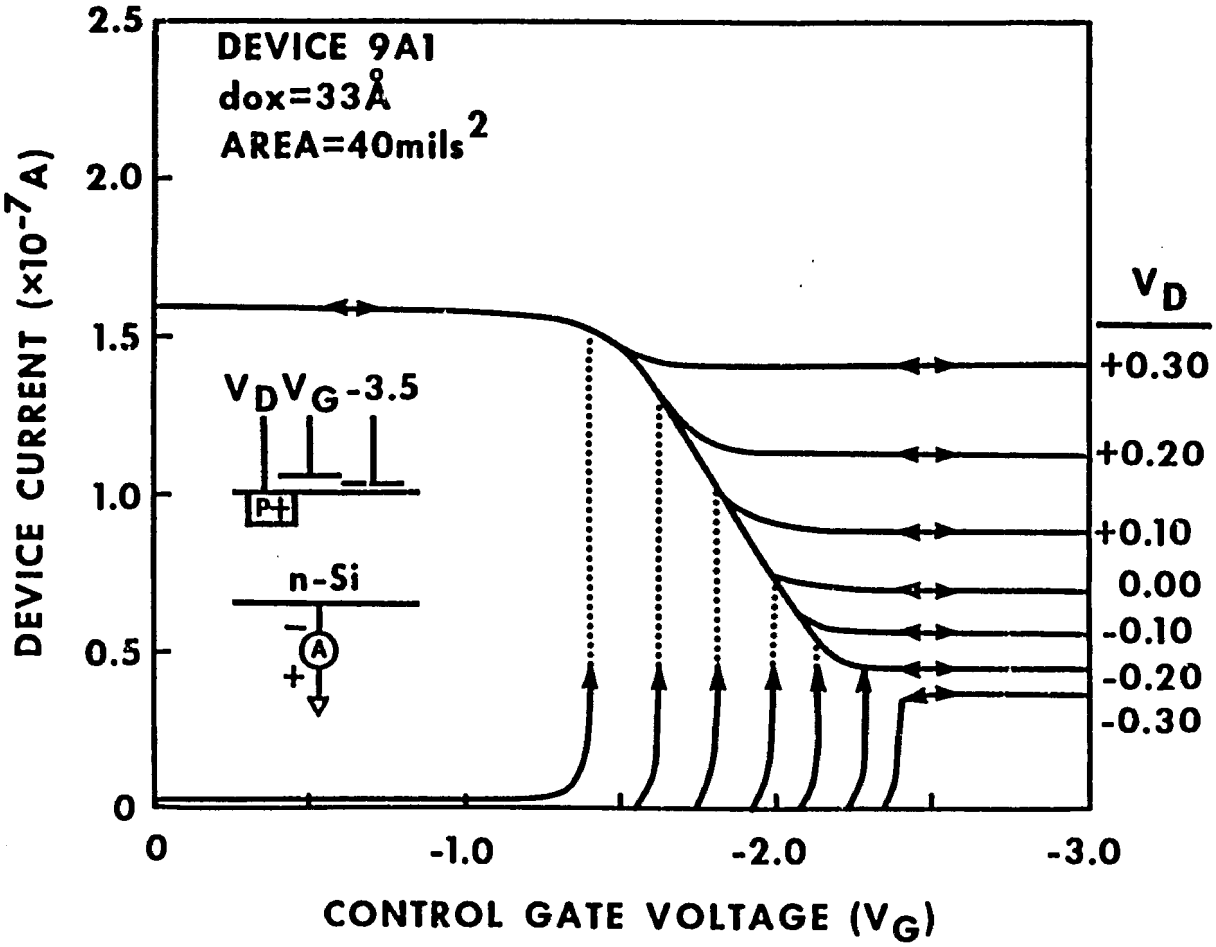
In Fig. 4.6-3, a similar measurement is performed except that the MTOS junction bias is held fixed at -3.5 volts and the diode voltage is discretely varied. The effect of the diode voltage is seen to be two-fold. The first effect is to vary the control gate threshold voltage required to initiate switching of the MTOS junction from a low current to a high current state. This effect is similar to the apparent threshold lowering effect observed in MOSFETs if a non-zero 'source' bias is applied relative to substrate [Muller and Kamins, 1977]. This 'body' effect is that

$$V_T' = V_T - (2\phi_o\phi_n - \phi_o V_S)^{1/2} + (2\phi_o\phi_n)^{1/2} \qquad 4.6\text{-}1$$

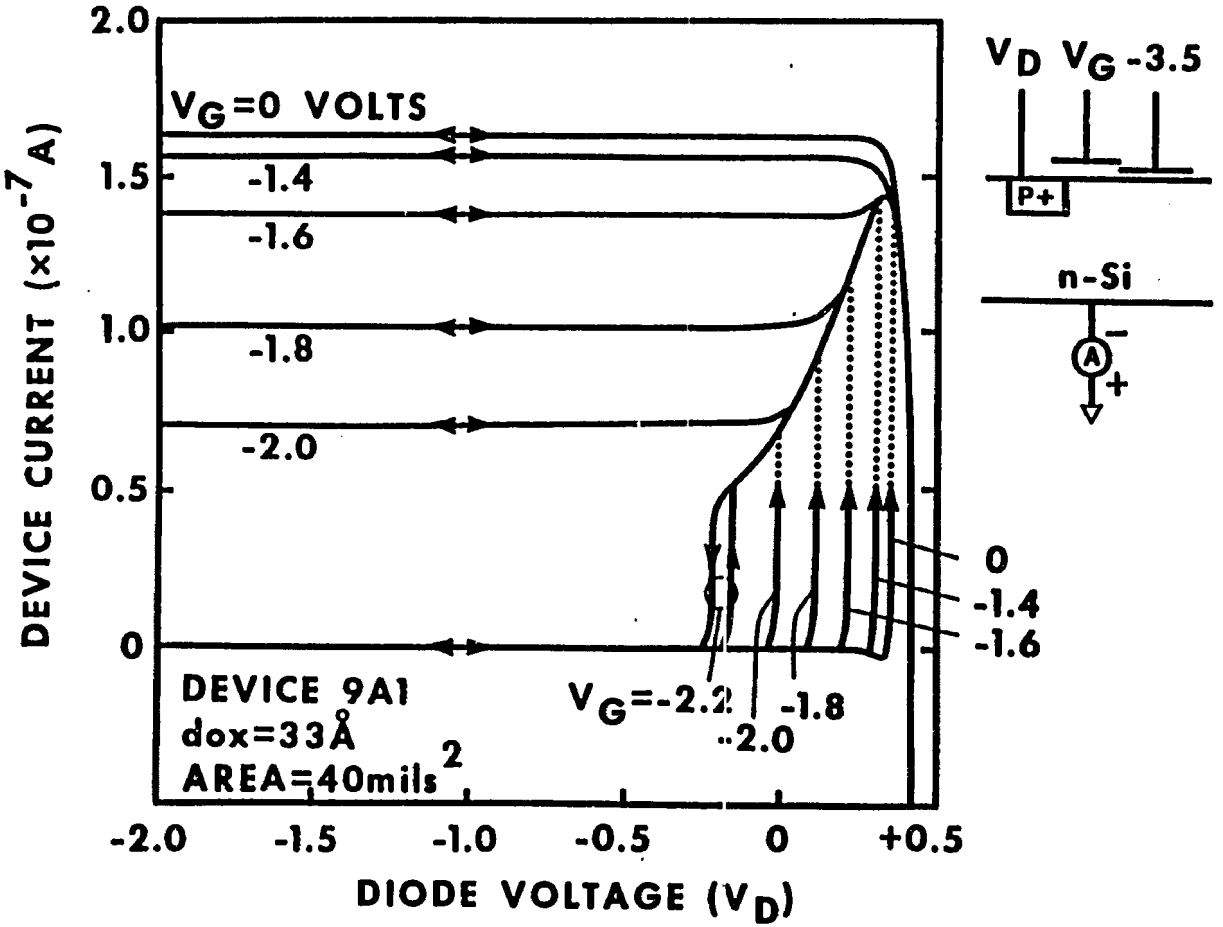
Once the inversion layer has been formed, the diode controls the MTOS surface potential. By comparing Fig. 4.6-2 and Fig. 4.6-3, the relationship of Eq. 4.5-3 is again verified. (Compare the currents at $V_G = -3.0$ volts).

This system can be considered to be a gate-controlled threshold detector. If the control gate voltage exceeds a certain threshold value, the MTOS junction switches to a stable high current state. The threshold level, as illustrated in Fig. 4.6-3 can be altered by adjusting the diode voltage over approximately a half volt range.

Another variation on the theme is shown in the measurements of Fig. 4.6-4. Here the MTOS junction is fixed at a bias of -3.5 volts and substrate current is monitored as the diode voltage is swept for various values of control gate voltage. Consider first the case for a control



4.6-3 Substrate current as a function of control gate voltage for various values of PN junction bias.



4.6-4 Substrate current as a function of PN junction bias for various values of control gate voltage.

gate voltage V_G = -2.0 volts. The MTOS junction is initially in a low current state as V_D is ramped from -2.0 volts toward zero. It is energetically unfavorable for there to be an inversion layer under the control gate in this case until $V_D > 0$ volts so that for $V_D \leq 0$, the diode is decoupled from the MTOS junction. As V_D is ramped above zero, an inversion layer forms and the MTOS junction surface potential becomes coupled to the diode voltage and the MTOS junction switches to a high (diode-controlled) current state. The device current rises exponentially with increasing V_D (i.e. increasing V_{OX}) until at V_D +0.3 volts, the forward biased diode current obscures the MTOS electron tunneling current and the net substrate current begins a sudden nose dive exhibiting differentially negative resistance.

Sweeping in the reverse direction from +0.4 volts to -2.0 volts, the opposite happens except that when the inversion layer dissipates close to V_D equal to zero volts, the MTOS junction remains in a high current state. The MTOS state is pinned by lateral diffusion toward the diode, the strength of which depends upon the control gate channel's state of depletion (i.e. V_G).

If the control gate voltage is large enough (for example V_G = -2.2 volts) when the diode voltage is swept toward -2.2 volts, the MTOS inversion layer is less than the critical level required for the impact ionization positive feedback when the coupling inversion layer dissipates so that the MTOS junction relaxes back to a low current state.

For small values of control gate voltage (less than -1.4 volts) the coupling of the MTOS junction and PN junction while the MTOS device is in a low current state is by hole diffusion between the PN junction and the MTOS deep depletion potential well. In this situation, the two

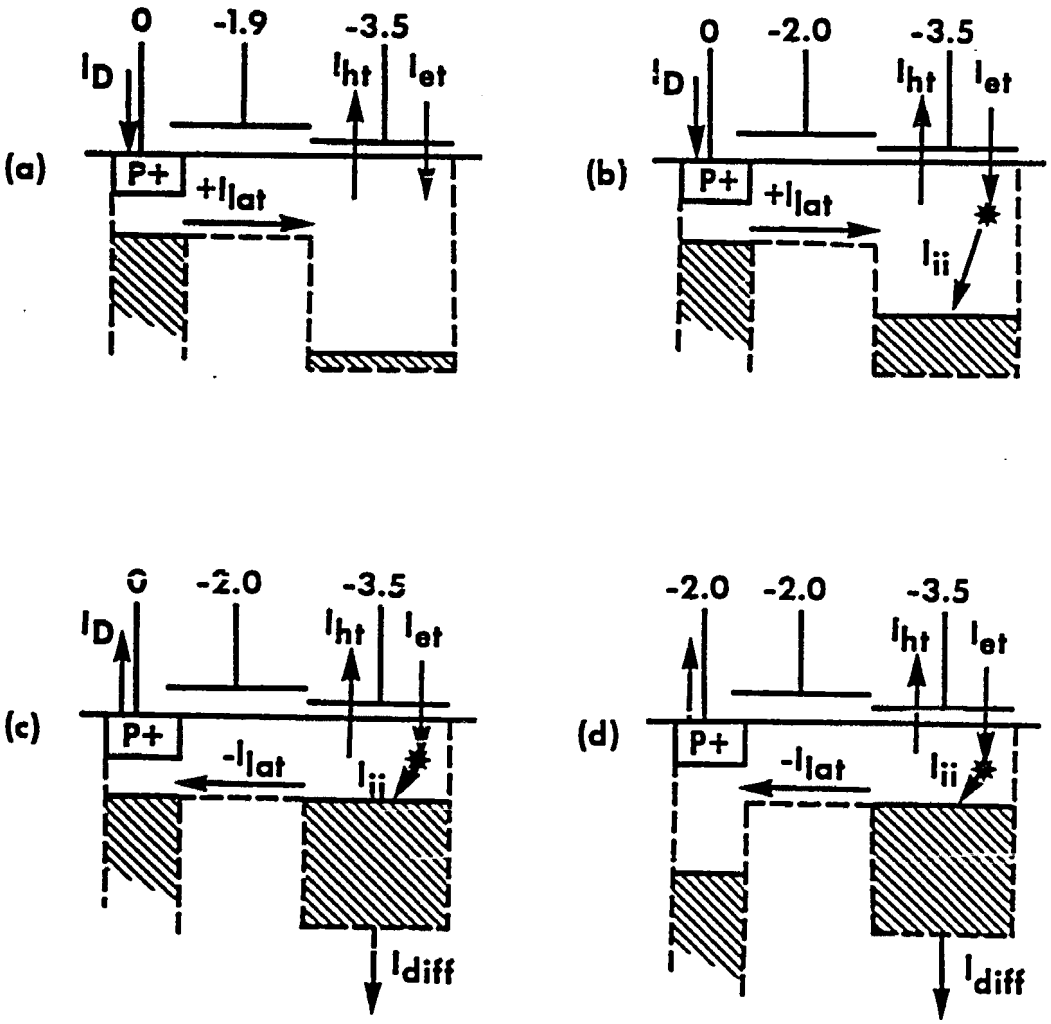
devices behave as the devices studied by Teng (1983). The efficiency of such coupling is poor so that a significant total PN junction current must flow before switching occurs.

It is difficult to diagrammatically show the physical properties of the device, but in Fig. 4.6-5, a potential well-like diagram is shown which attempts to illustrate the different coupling states of the system. Before proceeding, it is important to note that diffusion can only be marginally illustrated by such a diagram, and the control gate well is not really a well at all. Super-inversion doesn't really fit into this pictorial scheme as it was originally intended. Nevertheless, in Fig. 4.6-5a, the system is shown for the MTOS device in a low current state with some sub-threshold conduction of holes from the diode taking place. This lateral flow is balanced by oxide hole leakage so that $\dot{Q}_{INV} = 0$.

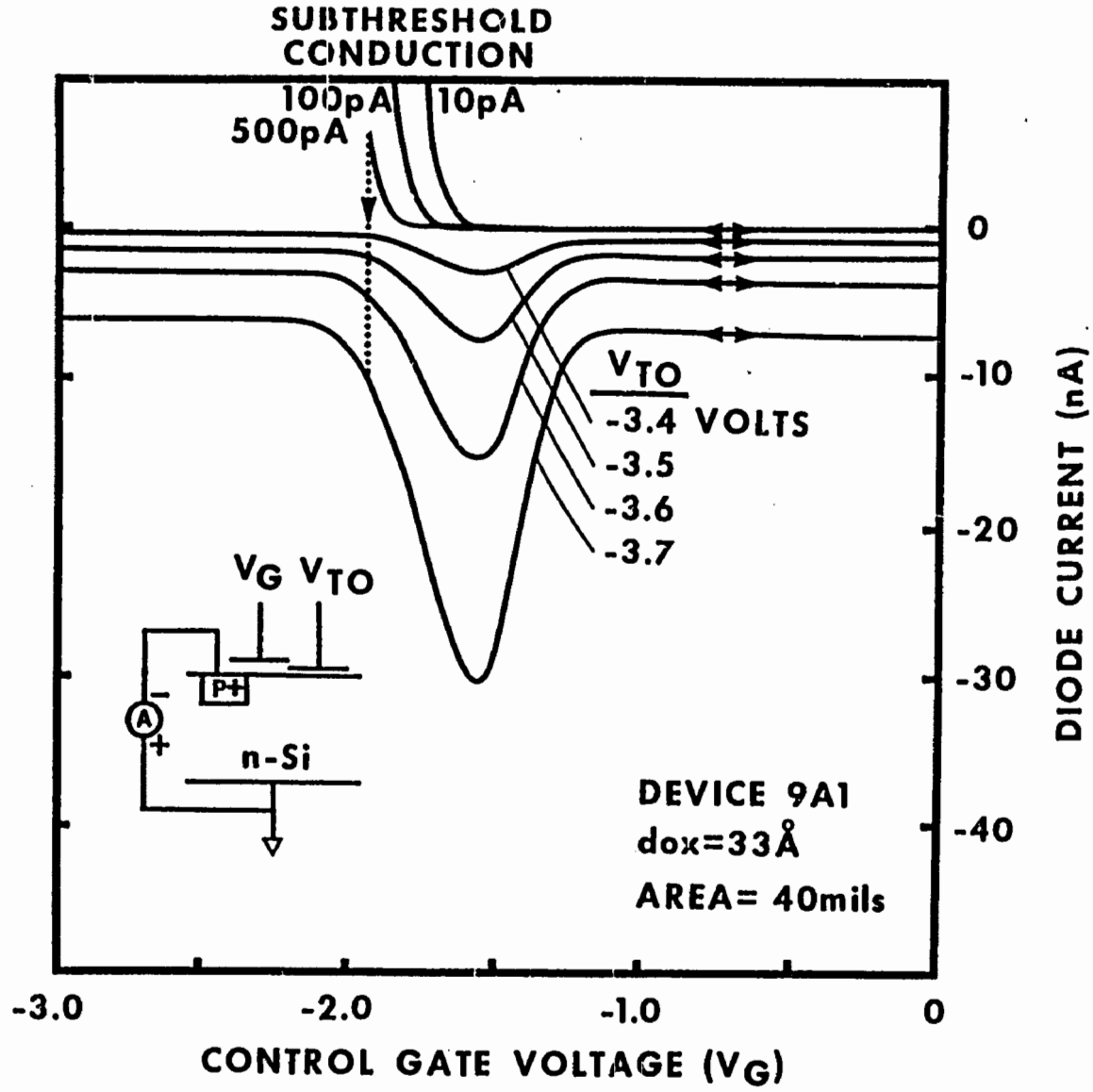
In Fig. 4.6-5b, the lateral current has increased sufficiently (by increasing V_D or V_G) to initiate switching of the MTOS junction. The growth of the MTOS inversion layer is stabilized in Fig. 4.6-5c by lateral hole diffusion toward the diode and by back hole diffusion. Reducing V_G can be imagined to increase the MTOS inversion layer and hence to vary the various components of $\dot{Q}_{INV}$ which will still sum to zero.

In Fig. 4.6-5d, a negative diode voltage is seen to have little effect on the MTOS inversion layer, since the control gate continues to limit lateral diffusion from the super-inversion layer.

So far measurements involving the substrate current have been discussed. Measurement of the diode current is shown in Fig. 4.6-6. Consider $V_{TO} = -3.5$ volts and refer to the similar condition of 4.6-2. As the control gate voltage is increased from zero toward -3.0 volts



4.6-5 Potential well-like diagrams for subthreshold control gate voltage switching of coupled MTOS junction and PN junction. See text.



4.6-6 PN junction current as a function of control gate voltage for various values of MTOS junction bias.

(with the MTOS junction initially in a low current state), it can be seen that the diode current increases exponentially. This lateral hole injection current is identical to subthreshold MOSFET conduction current. However, at $V_G = -2.0$ volts, the hole injection via subthreshold conduction is sufficient to cause the MTOS junction inversion layer to grow autonomously through the impact ionization hole generation mechanism (Fig. 4.6-5b). The growth is limited by the lateral subthreshold diffusion of holes toward the diode and back diffusion to the bulk (Fig. 4.6-5c). The diode current is observed to change sign and magnitude.

After switching, the MTOS-PN junction system is in that counter-intuitive regime of operation. Increasing V_G improves the control of the diode over the MTOS junction surface potential, and the MTOS junction current and lateral hole diffusion current decrease as the super-inversion region is brought under the reins of the diode voltage until firm control is established. The lateral current flow is primarily drift once the control gate layer is established.

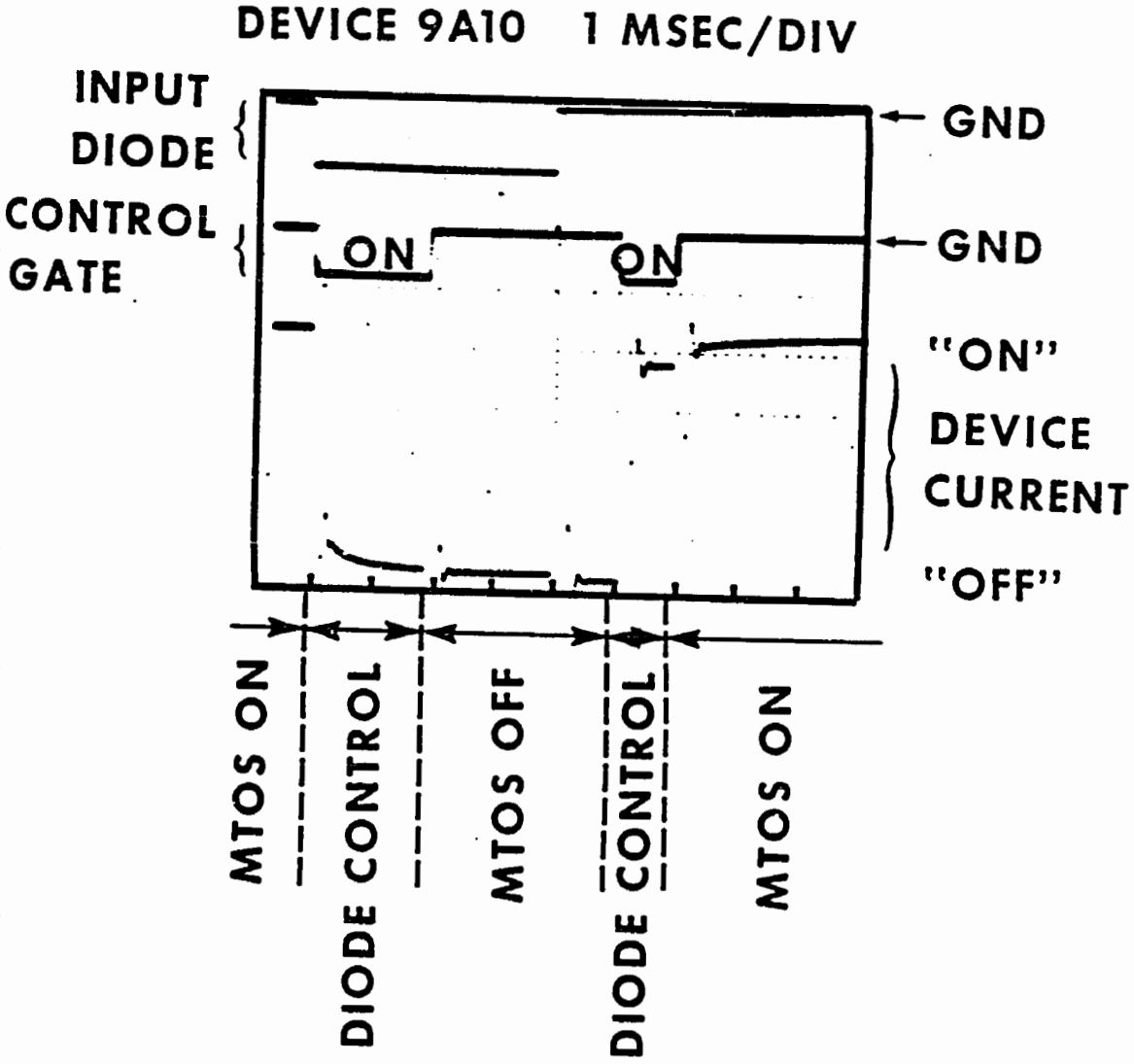
Sweeping the control gate voltage back toward zero decouples the diode voltage from the MTOS junction surface potential, but the latter is still limited by the control gate voltage (Fig. 4.6-5d, but imagine $V_D = 0$ and $V_G = 1.7$ volts) like an overflowing bathtub. Reducing V_G increases the maximum level of the MTOS inversion layer, and results in an increased lateral diffusion current as collected by the diode. When the control gate voltage is reduced below approximately -1.5 volts, the majority carriers in the coupling channel limit the amount of lateral diffusion current which actually reaches the diode and is collected. This reduction in collection efficiency results in a reduction of diode current, until the collection efficiency reduction saturates near $V_G = -1.2$ volts.

The presence of a continued diode current for a clamped off channel provides hard evidence for the concept of lateral and back diffusion of minority carriers from a super-inversion layer. Such a back and lateral diffusion current may in fact be a practical drawback of this device, or a useful property depending upon the application. Excess minority carriers can cause latch-up in CMOS circuits for example, an undesirable feature. However, diffused carriers can fill a CCD potential well, allowing for soft coupled detection of the MTOS ON state, or provide a base current to a PNP transistor yielding simple integrated signal amplification.

4.7. Pulsed Control Gate Switching of the MTOS Junction

In Fig. 4.6-4, it can be readily discerned that with the MTOS junction initially in the low current state, and the diode grounded to substrate, applying a control gate voltage greater than -2.0 volts will cause the device to switch to a high-current state. Removing the control gate bias will further increase the device current. In this manner, it is actually a simple manner to switch the MTOS junction ON electrically. To electrically switch the junction OFF is also straight forward and can be done in two ways. The classical way is to briefly remove the MTOS junction bias, which will result in the recombination and diffusion of the inversion layer. Another way is to begin with the control gate off and bias the diode at perhaps -1.0 volt. Pulsing the control gate on will cause the diode to extract the MTOS inversion layer and switch the junction OFF. Removing the control gate bias will result in the isolated MTOS junction in its stable OFF state. A oscilloscope photograph illustrating the electrical switching of the MTOS junction is shown in Fig. 4.7-1. Note that the junction can also be switched OFF by transferring out the charge in the MTOS potential well in a CCD manner.

Perhaps more interesting is to preset the initial state of the MTOS



4.7-1 Pulsed control gate switching of MTOS junction using biased PN junction.

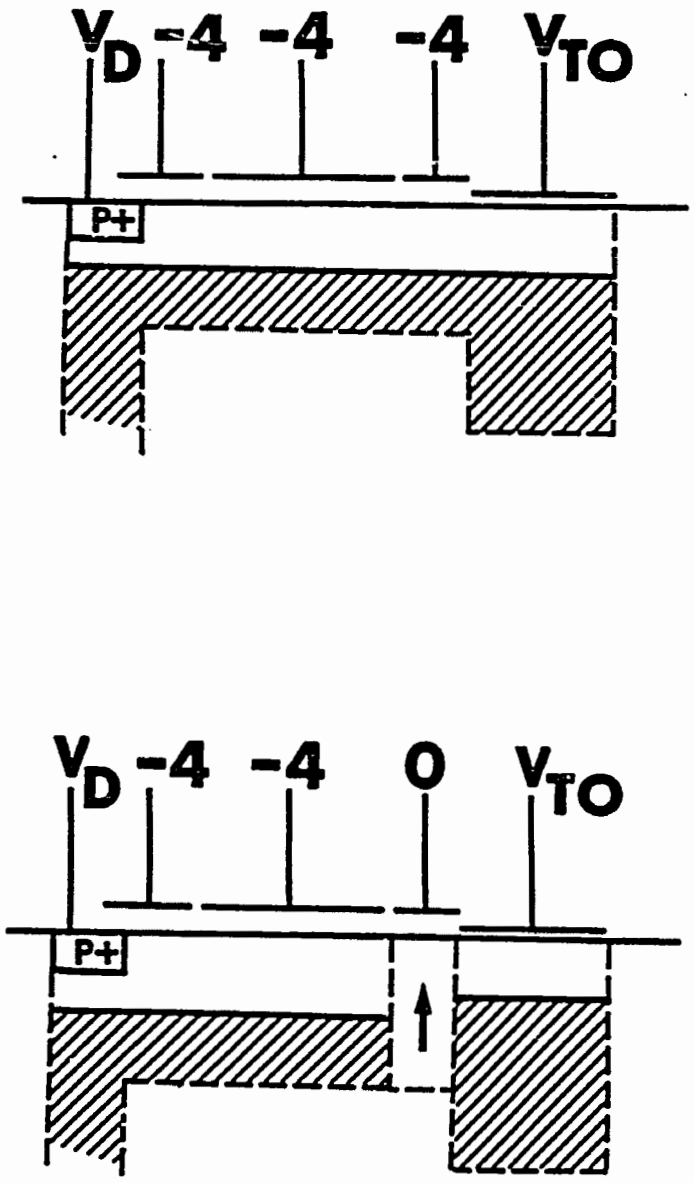
junction using the diode voltage. In this operation, the control gate is initially on with the diode voltage setting the MTOS junction surface potential to some steady-state intermediate level. The control gate is suddenly clamped and the MTOS junction either relaxes to the OFF state or switches to the ON state, depending upon the preset value of V_{OX}. This situation is entirely analogous to the charge packet initiated switching experiment in which the size of the injected charge packet presets the initial state of the junction. As such, it is not surprising that the transient device current resembles that of the charge packet initiated switching experiment as illustrated in the multiple exposure photograph of Fig. 4.7-3 for various values of V_D.

One slight difference about this preset surface potential transient switching experiment concerns the discontinuity observed at t = 12 msec in the transient current corresponding to the instant when V_{XO} is suddenly clamped off. The reason for the discontiniuty is that when the inversion layer under V_{XO} is clamped off, the holes which comprised the layer are ejected both back toward the diode and forward into the MTOS junction. Approximately half the inversion layer goes in one direction, and half in the other. The extra charge ejected into the MTOS junction can be calculated by writing

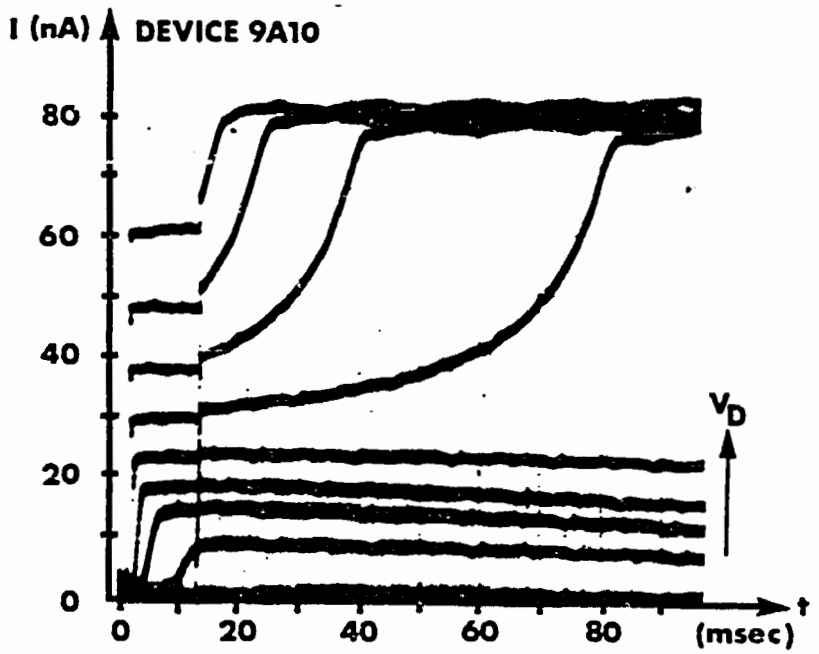
$$\Delta Q = \frac{1}{2} A_{XO} C_{OX} (V_{XO} - V_T') \qquad 4.7\text{-}1$$

where A_{XO} is the relevant area of the output transfer gate and C_{OX} is the specific gate oxide capacitance. Using $A_{XO} C_{OX}$ equal to 8 pF and $V_{XO} - V_T'$ equal to 2 volts we obtain ΔQ equal to 8 pC. The change in current is well approximated by expanding Eq. 4.4-4 to obtain

$$\frac{\Delta I}{I} = \frac{\Delta Q}{Q_{et}} \qquad 4.7\text{-}2$$



4.7-2 Potential well diagram illustrating preset MTOS junction surface potential experiment.



4.7-3 MTOS junction transient current as a function of time for various values of preset surface potential. MTOS junction bias applied at t = 0, and output transfer gate clamped off at t = 12 msec. Diode voltage increased from −0.6 volts to +0.2 volts in 0.1 volt increments. MTOS junction biased at −3.45 volts. Waveform ripple due to residual 60 Hz noise.

which using Q_{et} = 87 pC yields ΔI/I approximately equal to 10%. This is almost exactly the value observed.

Note also that in this experiment, the more negative V_D becomes, the longer it takes to precharge the MTOS potential and preset the surface potential. This is because the equivalent MOSFET conductance of the control gate channel varies as the square of $V_G-V_D-V_T'$ (see Eq. 4.6-1).

In principle, nearly all the calculations performed in section 4.4 could be applied to the transient current response of Fig. 4.7-2. However, due to the uncertainties in the preset surface potential as alluded to throughout this chapter, it was felt that the errors associated with analyzing data collected in this manner would be greater than those using the charge packet method. In addition, since no new information would be garnered, demonstration of the qualitative equivalence of the two methods is satisfactory.

4.8 Summary

The new results and ideas presented in this chapter are now summarized.

First, the successful operation of the MTOS junction as a charge packet threshold detector was demonstrated. Its characteristics are summarized in Table 4.8-1. The quintessential feature of the device is its relatively slow physical computation time (10-100 milliseconds) which is too slow for most applications. There is (undemonstrated) promise of increasing the switching speed by altering the oxide thickness and gate material work function.

In the fabricated Al/SiO_2/Si(n) device, the electron tunneling current, impact ionization current and oxide hole leakage currents were measured using two techniques. One technique was to preset the MTOS

junction to an unstable state by injecting a packet of hole charge of a known and calibrated amount, and observing the transient device current as the junction relaxed to its OFF state or switched to its ON state. The second technique was to use a diode-controlled MOS capacitor structure and measure the diode current as a function of MTOS junction surface potential in a steady-state fashion. The two techniques yielded basically identical results.

The average total impact ionization probability obtained from the ratio of hole generation current to electron tunneling current was measured as a function of hot electron energy. This experimental result was compared to a theoretical scattering rate calculation, and the results were found to agree remarkably well.

When the MTOS junction was coupled to a PN junction through an intermediate control gate, a new N-type negative resistance was discovered, whose strength is controllable with the applied MTOS bias. This N-type negative resistance has since been observed for thicker oxides biased into the Fowler-Nordheim tunneling regime, which by themselves are not bistable. This N-type negative resistance may be useful in bistable switching circuits.

The use of the control gate with MTOS devices has allowed the electrical switching of the junction with significantly greater efficiency than has been done previously. The control gate allows the thresholding operation to be transferred from a minority carrier current in the MTOS depletion region to a control gate voltage. The switching threshold using the control gate can be altered by varying the PN junction voltage.

A model of the MTOS junction carrier concentration profile was developed which properly estimates the potential drop across the super-

inversion layer as demonstrated by comparing capacitance calculations to experimental results. This model also predicts that the ON state stabilization mechanism is mainly due to back diffusion of holes from the super-inversion layer. This diffusion current was experimentally observed in the fabricated devices. Such a stabilization mechanism is in contrast to a recombination mechanism (in the space-charge region) proposed by earlier workers.

In the course of this work, experimental techniques have been established which enable the control of the MTOS junction surface in a manner which allows access to the entire state-space of the junction, and gives the capability to map device currents, potentials and capacitances throughout all regimes of operation.

Some of the areas which need further investigation include resolving the oxide hole transport mechanism issue, a fuller understanding of how the diode coupled to the MTOS junction controls its surface potential, and applications of the newly discovered N-type negative resistance effect.

The author is grateful to his predecessors who have established firm foundations for this work. It is hoped that the techniques and methods introduced in this chapter may be useful for the continued study of thin oxides and other dielectric films. The metal/tunnel-oxide/semiconductor system remains one of the most complex yet structurally simple systems encountered by the author. Coupling such a system to a PN junction or CCD may extend the useful role thin oxides can play in practical circuits, but at the same time seriously compounds the device complexity.

Table 4.8-1 Summary of Charge Packet Threshold Detector Characteristics

(9A13) $Al/SiO_2/Si(n)$	Actual	Area Normalized
Area	$2.58 \times 10^{-4} cm^2$	1
	40.0 $mils^2$	1
Oxide Thickness	33.0 ± 0.5 Å	33.0 ± 0.5 Å
Flat-Band Voltage	-0.19 volts	-0.19 volts
Oxide Capacitance	266 pF	1.03 $\mu F/cm^2$
OFF State (V_{TO} = -3.5 volts)		
Current (Electron Tunneling)	2.0×10^{-9} A	7.7×10^{-6} A/cm^2
Inversion Layer Charge	355 pC	1.4 $\mu C/cm^2$
Surface Potential	-1.4 volts	-1.4 volts
Oxide Voltage	-1.9 volts	-1.9 volts
Equilibrium State ($pn = n_i^2$)		
Current	2.4×10^{-8} A	9.5×10^{-5} A/cm^2
Inversion Layer Charge	594 pC	2.3 $\mu C/cm^2$
Surface Potential	-0.98 volts	-0.98 volts
Oxide Voltage	-2.33 volts	-2.33 volts
Metastable (Critical) State		
Current (Electron Tunneling)	3.3×10^{-8} A	1.3×10^{-4} A/cm^2
Inversion Layer Charge	622 pC	2.4 $\mu C/cm^2$
Surface Potential	-0.88 volts	-0.88 volts
Oxide Voltage	-2.43 volts	-2.43 volts
ON State (V_{TO} = -3.5 volts)		
Current (Electron Tunneling)	9.8×10^{-8} A	3.8×10^{-4} A/cm^2
Inversion Layer Charge	720 pC	2.8 $\mu C/cm^2$
Surface Potential	-0.56 volts	-0.56 volts
Oxide Voltage	-2.75 volts	-2.75 volts

Table 4.8-1 Summary of Charge Packet Threshold Detector Characteristics (cont'd.)

	Actual	Area Normalized
Switching ON Time	10 - 100 msec	10 - 100 msec
Switching OFF Time	100 - 300 msec	100 - 300 msec
Impact Ionization Probability	2 - 5%	2 - 5%
Electron/Hole Oxide Current Ratio	5 - 100 (extrap)	5 - 100 (extrap)
	5 - 50 (meas.)	5 - 50 (meas.)
I_{et}^{o}, Q_{et}^{o}	26 pA, 87 pC	100 nA/cm^2, 0.34 $\mu C/cm^2$
I_{t}^{o}, Q_{ht}^{o}	33 pA, 208 pC	128 nA/cm^2, 0.81 $\mu C/cm^2$
I_{ii}^{o}, Q_{ii}^{o}	26 fA, 42 pC	1 pA/cm^2, 0.16 $\mu C/cm^2$

CHAPTER 5

CHARGE-PACKET DIFFERENCER CIRCUIT

5.1 Introduction

The charge-coupled differencer circuit discussed in this chapter is a novel implementation of the functional block discussed in chapter 2. This circuit creates an output charge packet equal in magnitude to the difference in magnitudes of two input charge packets. The advantage of the implementation of the differencer presented in this chapter is that the physics of the device predict inherently linear transfer characteristics with exactly unity gain. In principle, the circuit can be fabricated using a minimal amount of real-estate, which makes it particularly attractive for charge-coupled computer array applications. A third important feature of this invention is that the origin of the input charge and the differencer electrodes can be separated by structures in the silicon yet connected by a wire passing over those structures, an advantage not normally enjoyed by charge-coupled devices.

In cooperation with the Hughes Aircraft Company's Corporate Patents Office, a novelty search was initiated revealing several circuits which perform a charge packet differencing operation. The prior art which was found to be the closest to the circuit disclosed in this chapter was invented by Fagan (1978). The primary differences between the two circuits are that first Fagan generates two voltages which are proportional (possibly with two different proportionality constants) to two reference charge packets. Fagan then uses two sequential fill and spill cycles utilizing one of these voltages on each cycle to generate the output charge packet. Such a technique produces an inherently non-linear relationship between input charge packet and requires more real

estate and more complicated timing than the circuit presented in this work.

There are two drawbacks to the new charge-packet differencer circuit. The first is that the input charge packets are destructively sensed. This nuisance can be circumvented in the charge-coupled computer by cloning the input charge packets prior to the sense operation through the utilization of the differencer's replicator mode. The second drawback is that the inherent unity gain is affected by stray capacitance. Since the stray capacitance is often a function of voltage, the change in gain appears as a distortion in the transfer characteristic. The stray capacitance can be minimized by proper layout and advanced fabrication technologies.

In the Yale 8205 implementation of the charge packet differencer, much larger geometries than what would be used in a true charge-coupled computer array were employed, due to the limitations imposed by fabricating the devices at Yale. The consequences of these oversized geometries were that charge transfer inefficiency, dark current and PN junction leakage current were all in excess of what they might be were a better facility available. These consequences are compounded due to interplay among the non-optimized characteristics. For example, to improve charge transfer efficiency, clocking rates were kept relatively slow. Such increased time intervals amplified the detrimental aspects of both dark current and leakage currents.

Even with the limitations imposed by the processing facility, the charge packet differencer implemented on the Yale 8205 chip demonstrated the validity of the performance predictions made based on initial analysis of the proposed structure. The device performs charge packet differencing in a linear and accurate way and has a large dynamic range.

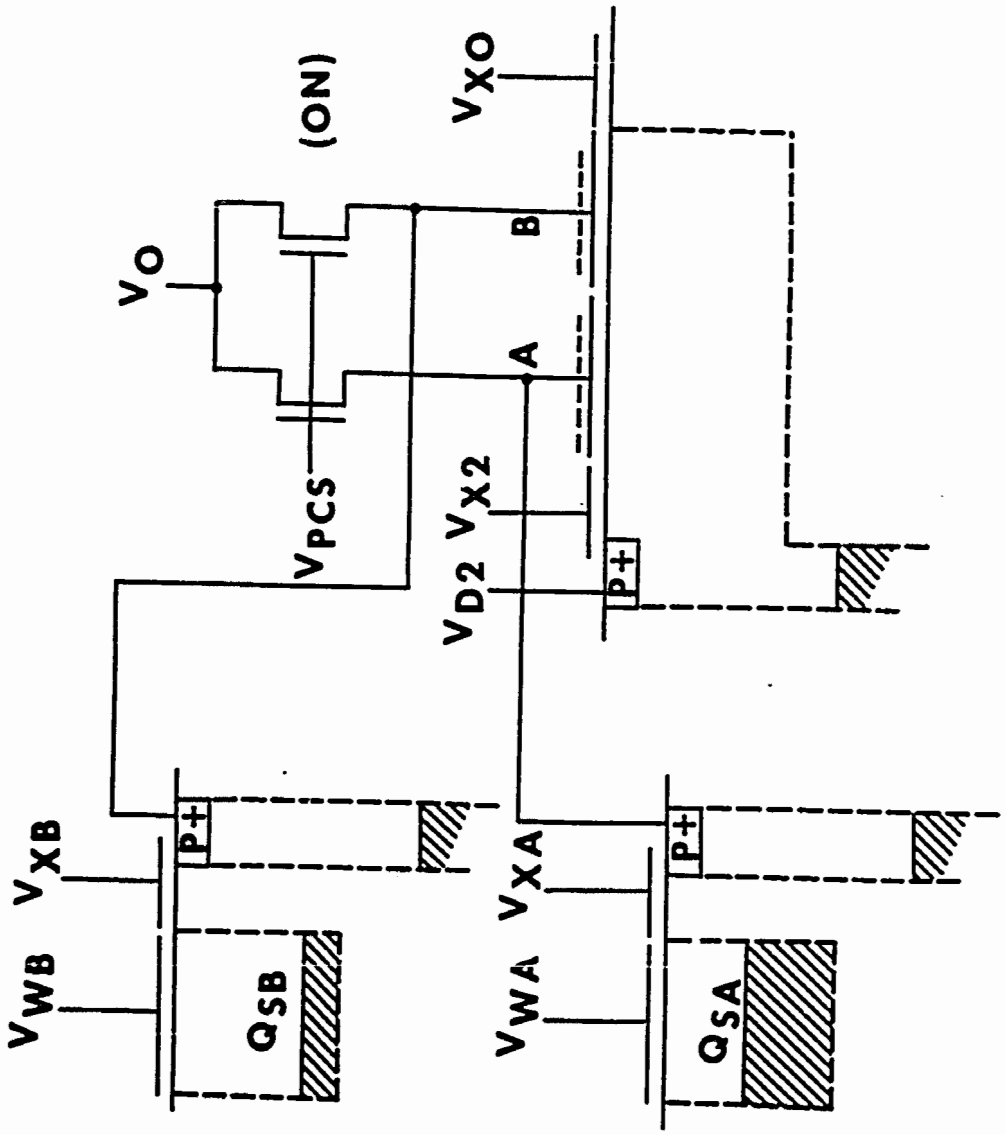
In the sections which follow, the operation of the device will be discussed and the predicted performance analyzed. The device layout and the rationale for the chosen geometry will be covered, including a proposed layout for improved performance. Experimental results obtained from measurement of the fabricated devices will be examined and analyzed, including calibration of the output amplifier. This chapter will end with a summary of the results obtained by this work and conclusions regarding future directions.

5.2 Theory of Operation

Basic Operation

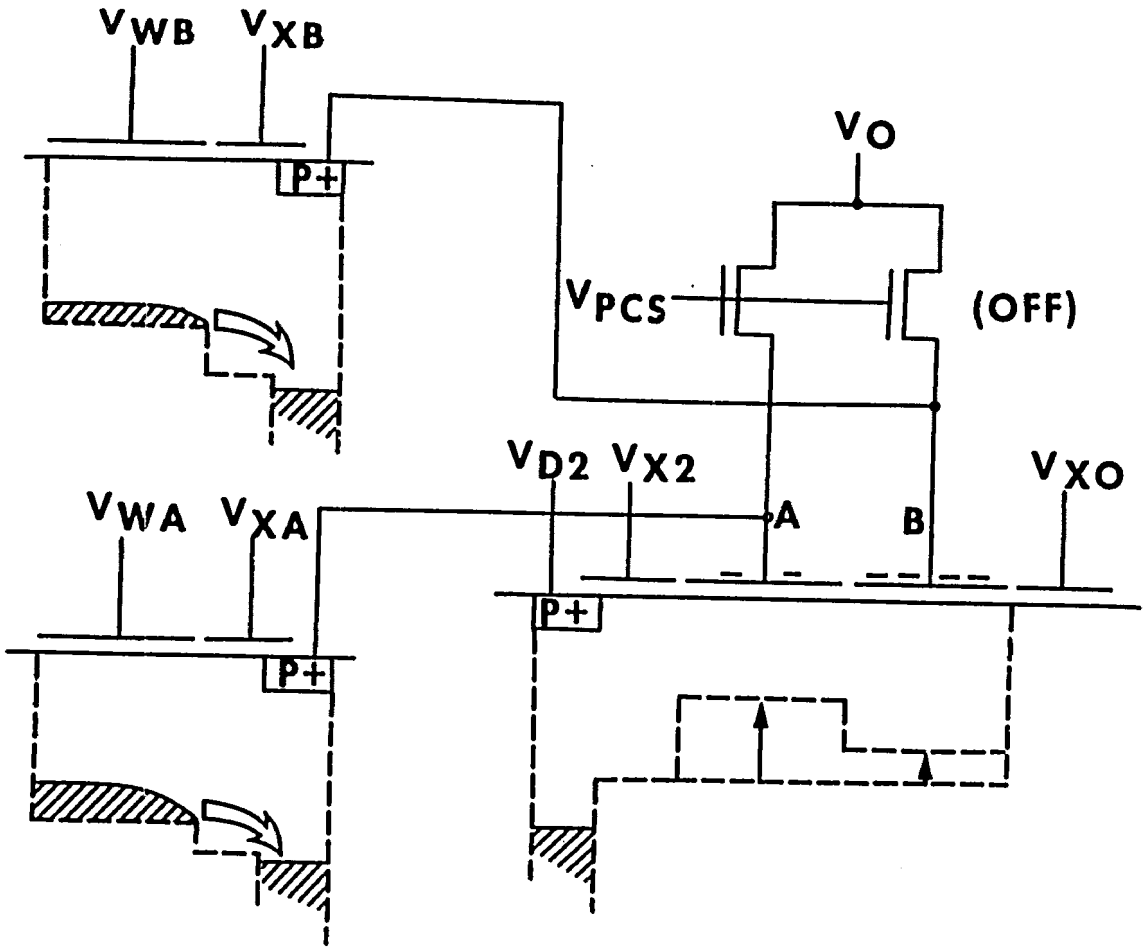
To assist in the discussion of the basic operation of the charge packet differencer circuit, a schematic series of drawings is shown in Fig. 5.2-1. In Fig. 5.2-1a, a cross-section showing two input charge packets Q_{SA} and Q_{SB} in storage wells V_{WA} and V_{WB} isolated from floating p^+ diffusions by output transfer gates V_{XA} and V_{XB} is displayed. The floating diffusions are wired to two electrodes labelled A and B, which are also connected via a MOSFET switch to the precharge potential V_O. Next to electrode A is an isolation gate V_{X2} and an input diode reverse biased by V_{D2}. Adjacent to electrode B is an output transfer gate V_{XO}. These latter four gates, input diode and precharge MOSFET switches comprise the charge packet differencer. In Fig. 5.2-1a, the precharge switches have been momentarily closed by pulsing V_{PCS} to some negative bias larger than $V_O + V_T$, where V_T is the (negative) MOS threshold voltage. When this negative bias is removed and the switches open the circuit, the two differencer electrodes A and B remain precharged at a floating voltage V_O. The negative charge on each gate is Q_A and Q_B respectively.

In Fig. 5.2-1b, the output transfer gates V_{XA} and V_{XB} have been biased to a negative voltage larger than V_{WA} and V_{WB} respectively,

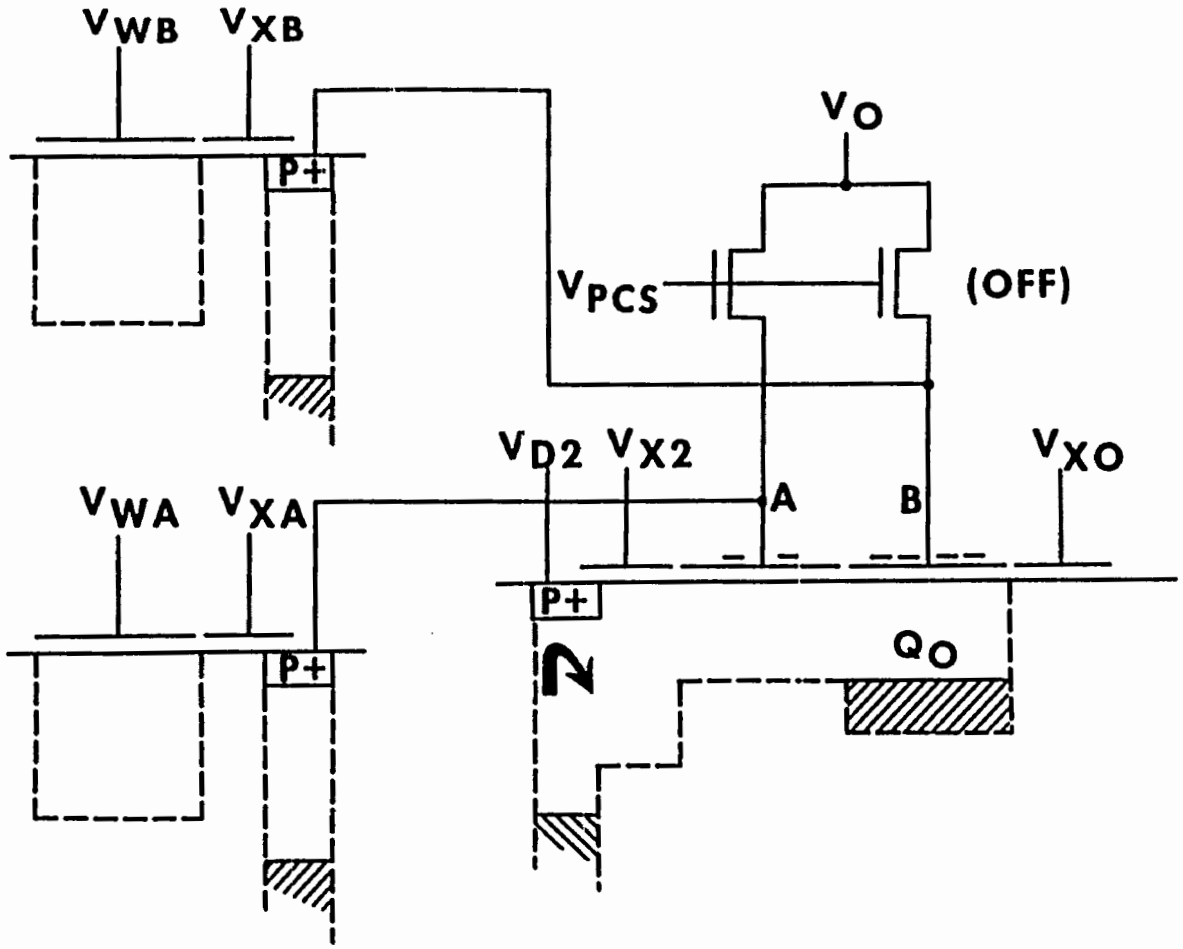


5.2-1 Schematic drawings of charge packet differencer circuit showing the three major phases of operation.

a. Precharge phase. Electrodes A and B are precharged to potential V_o by pulsing V_{PCS}.



b. Gate charge subtraction phase. Signal charge recombines with a portion of the negative gate charge altering the potential wells under electrodes A and B.



c. Charge packet formation phase. Input diode is pulsed creating charge packet Q_o in a surface potential equilibration manner (fill and spill).

causing the positive minority carrier charges $\underline{Q}_{SA}$ and $\underline{Q}_{SB}$ to be collected by the floating diffusions. The positive minority carrier charge recombines with a portion of the negative charge on electrodes A and B, resulting in a reduced charge on each electrode. To ensure complete charge transfer, the precharge voltage must be large enough to ensure that the floating diffusion bias V_O after recombination is still larger than $V_{WA} + V_T$, and $V_{WB} + V_T$. The new negative charges on each differencer electrode $\underline{Q}_A$ and $\underline{Q}_B$ are given by

$$\underline{Q}_A = \underline{Q}_A + \underline{Q}_{SA} \qquad 5.2\text{-}1a$$

$$\underline{Q}_B = \underline{Q}_B + \underline{Q}_{SB} \qquad 5.2\text{-}1b$$

This gate charge subtraction process is very similar to the operation of the output charge packet sense amplifier discussed in appendix D.

In Fig. 5.2-1c, the input diode (V_{D2}) has been momentarily forward biased with respect to the semiconductor surface under electrodes A and B, filling the potential wells under these gates with minority carrier charge. When the diode is returned to a reverse bias condition, a portion of the injected carriers spill out from the potential wells back to the diode and the rest are trapped in the potential well under electrode B. This process is wholly analogous to the Tompsett surface potential equilibration method of creating a charge packet (as discussed in chapter 2) except in this device the gate charge is fixed rather than the gate voltage.

The charge trapped under electrode B can be determined using Gauss' Law in the following manner. Since the spill process described above terminates when the surface potential under the two electrodes are equal, Eq. 2.3-10a is used to argue that the depletion layer charge areal densities under the two electrodes are equal. This assumes that there is a negligible potential drop across the minority carrier

inversion layer under electrode B, which is reasonable for gate oxides in the 500 Å thickness range under moderate bias conditions. It is also necessary to assume identical doping profiles under each electrode, though the doping profile need not be constant. The charge on the gate electrode, charge in the oxide, interface trapped charge, depletion layer charge and in the case of electrode B, the charge in the potential well is summed, and assuming zero electric field outside the MOS sandwich, one obtains from Gauss' Law:

$$\underline{Q}_A/A_A + Q_{OXA} + Q_{ITA} + Q_{DA} = 0 \qquad 5.2\text{-}2$$

$$\underline{Q}_B/A_B + Q_{OXB} + Q_{ITB} + Q_{DB} + \underline{Q}_O/A_B = 0 \qquad 5.2\text{-}3$$

where A_A and A_B are the areas of electrodes A and B, Q_{OXA}, Q_{OXB}, Q_{ITA} and Q_{ITB} are the fixed oxide charge and interface trapped charge areal densities on electrodes A and B, Q_{DA} and Q_{DB} are the depletion layer charge areal densities under electrodes A and B, and $\underline{Q}_O$ is the minority carrier charge trapped in the potential well under electrode B.

The depletion layer charge densities in Eqs. 5.2-2 and 5.2-3 are evaluated and the charge packet Q_O solved for. It is assumed (as is reasonable) that the oxide and interface trapped charge in the two MOS capacitors are identical. This yields

$$\underline{Q}_o = \frac{A_B}{A_A} \cdot \underline{Q}'_A - \underline{Q}'_B \qquad 5.2\text{-}4$$

which relates the magnitude of the captured charge packet to the charge on the two electrodes. This latter charge can be rewritten using Eq. 5.2-1, and by recognizing that in the precharged state, the areal gate charge densities must be equal (i.e. $\underline{Q}_A/A_A = \underline{Q}_B/A_B$). Thus,

$$\underline{Q}_o = \frac{A_B}{A_A} \underline{Q}_{SA} - \underline{Q}_{SB} \qquad 5.2\text{-}5$$

Noting that a negative number of minority carriers cannot be captured, the chapter 2 notation is adopted so that

$$\underline{Q}_o = \left[\frac{A_B}{A_A} \underline{Q}_{SA} - \underline{Q}_{SB} \right] \qquad 5.2\text{-}6$$

This is the basic equation of operation of the charge packet differencer.

If the differencer electrodes are of equal area, the most useful differencer operation is obtained, that of unity gain. Such a characteristic is desired for the charge-coupled computer. Note that if the B channel input charge packet is zero, the differencer acts as a charge packet replicator. Such a replicator might be designed with unity gain (equal electrode areas), as an attenuator ($A_B < A_A$) or as an amplifier ($A_B > A_A$). The latter situation is particularly intriguing since it may be well suited for low noise amplification of small charge packets, however this possibility is not pursued here.

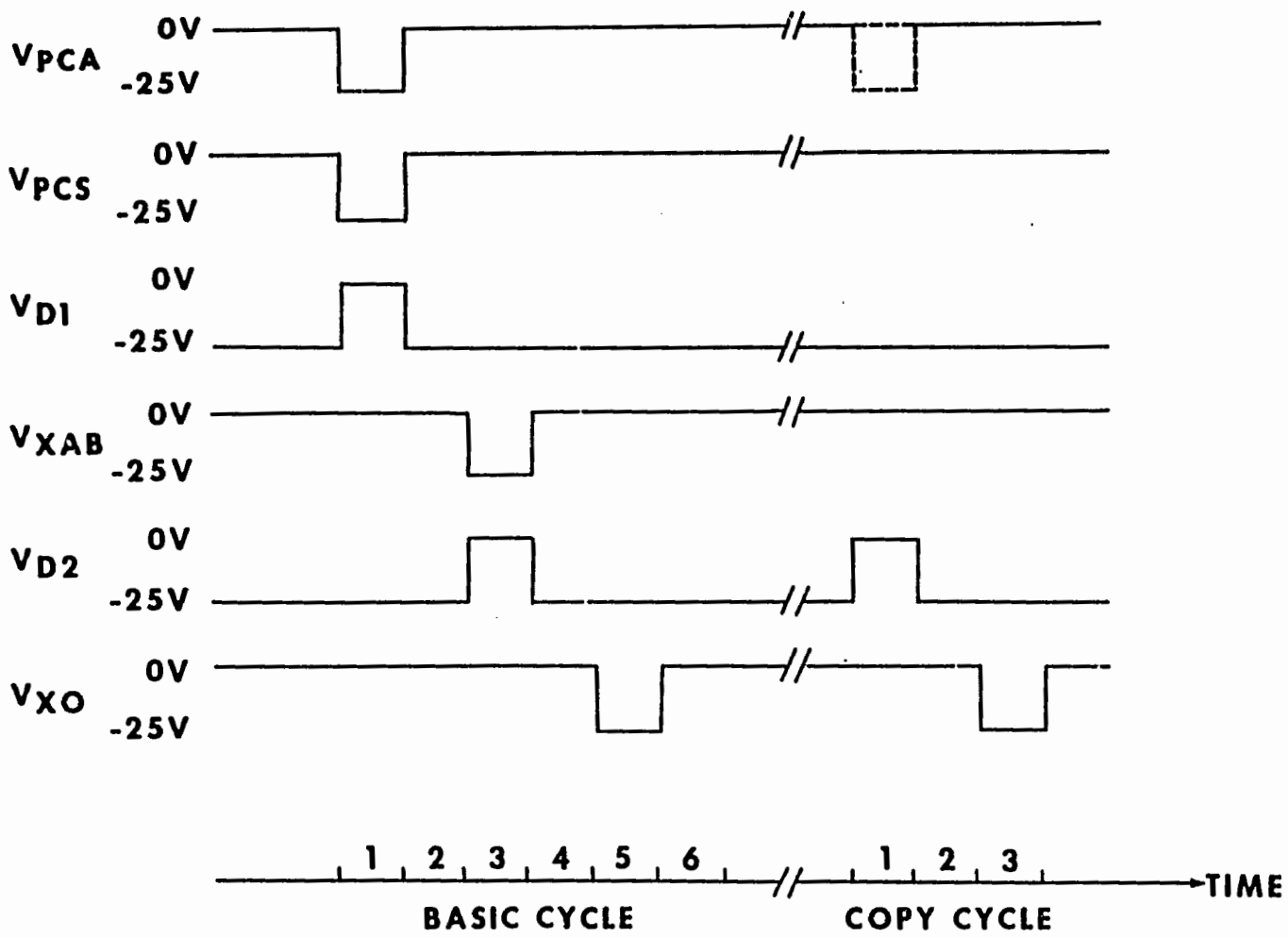
In a final phase not illustrated in Fig. 5.2-1, the newly formed output charge packet $\underline{Q}_O$ may be transferred from the differencer in the usual CCD fashion. At this stage, the differencer may be reset by pulsing the MOSFET precharge switches. Another option is to pulse the diode instead, to fill the potential well under electrode B a second time. In this multiple cycle mode of operation, the differencer (or replicator) acts as a charge packet copying machine. This feature is particularly useful if subsequent circuits utilizing the created charge packet are destructive in nature. The re-generated charge packets may

5.2-2 Timing diagram for charge packet differencer operation. Basic cycle is

1. Precharge phase and prepare input charge packets.
2. Gate charge subtraction phase and differencer fill.
3. Differencer spill.
4. Transfer output charge packet.

Copy cycle is:

1. Reset output amplifier (optional) and differencer fill.
2. Differencer spill.
3. Transfer output charge packet.


V_{PCA}
0V
-25V
V_{PCS}
0V
-25V
V_{D1}
0V
-25V
V_{XAB}
0V
-25V
V_{D2}
0V
-25V
V_{XO}
0V
-25V
1 2 3 4 5 6
1 2 3
TIME
BASIC CYCLE
COPY CYCLE

also be summed to form a larger charge packet which is a discrete multiple of the original. Note also that more than one input operation can take place, with the differencer electrode acting as a summation node during the gate charge subtraction process.

A timing diagram for the basic operation of the differencer circuit is shown in Fig. 5.2-2. Although more difficult to conceptually visualize, the gate charge subtraction phase (pulsing V_{XA}, V_{XB}) can be coincident with the differencer's fill phase (pulsing V_{D2}). This timing reduces the total time required to complete the differencing operation.

Charge Handling Capability

The maximum charge which can be input into the differencer depends upon the precharge bias V_0, and the area of the electrodes. The A electrode is more critical than the B electrode. If too much charge is subtracted from the B electrode during the input phase, the output charge packet will be zero. However, if too much charge is subtracted from the A electrode, the output will saturate, or even be reduced, depending upon the timing. The device saturates when the effective voltage on the A electrode becomes comparable to the MOS threshold voltage V_T. Once the A electrode is less than V_T, the fill and spill operation is hampered.

The maximum precharge voltage which can be applied to electrode A is determined by the breakdown strength of the deep depleted silicon under the electrode. If the breakdown field strength is $\mathcal{E}_{BR}$, the maximum precharge voltage is given by (assuming no interface trapped charge)

$$V_o^{MAX} - V_{FB} = \frac{\varepsilon_s \varepsilon_o}{C_{ox}} \mathcal{E}_{BR} + \frac{\varepsilon_s \varepsilon_o}{2qN_D} \mathcal{E}_{BR}^2 \qquad 5.2\text{-}7$$

For $N_D = 1 \times 10^{16}/cm^3$, the silicon breakdown strength is approximately

2.5×10^5 V/cm [Goetzberger and Nicollian, 1967]. For $d_{ox} = 550$ Å, the maximum precharge voltage is 24 volts. The maximum gate charge is then given by Gauss' Law as

$$\underline{Q}_A/A_A + \varepsilon_s \varepsilon_o \mathcal{E}_{BR} = 0 \qquad 5.2\text{-}8$$

After the maximum input charge has been subtracted from the gate, the effective voltage is V_T. The new gate charge is given according to

$$\underline{Q}_A'/A_A + Q_{DA}' = 0 \qquad 5.2\text{-}9$$

where Q_{DA}, the minimum depletion layer charge density is obtained by setting the surface potential equal to $2\phi_n$ such that

$$Q_{DA}' = (4\phi_n q \varepsilon_s \varepsilon_o N_D)^{1/2} \qquad 5.2\text{-}10$$

The maximum input charge packet is the difference between $\underline{Q}_A$ and $\underline{Q}_A'$ and is

$$\underline{Q}_{SA}^{MAX} = A_A (\varepsilon_o \varepsilon_s \mathcal{E}_{BR} - 4\phi_n q \varepsilon_s \varepsilon_o N_D) \qquad 5.2\text{-}11$$

Evaluating this quantity gives a value of $\underline{Q}_{SA}^{MAX}/A_A$ of 0.21 $\mu C/cm^2$. The maximum output charge packet is determined by Eq. 5.2-6 and is

$$\underline{Q}_o^{MAX} = A_B (\varepsilon_o \varepsilon_s \mathcal{E}_{BR} - 4\phi_n q \varepsilon_s \varepsilon_o N_D) \qquad 5.2\text{-}12$$

and has the same value $\underline{Q}_O^{MAX}/A_B = 0.21$ $\mu C/cm^2$.

If the differencer is not operated at full capacity, the maximum input signal charge is given by

$$\underline{Q}_{SA}^{MAX} = A_A(\varepsilon_o \varepsilon_s \hat{\mathcal{E}} - 4\phi_n q \varepsilon_s \varepsilon_o N_D) \qquad 5.2\text{-}13$$

where

$$\hat{\mathcal{E}} = \left[\left(\frac{qN_D}{C_{ox}}\right)^2 - \frac{2qN_D(V_o - V_{FB})}{\varepsilon_s \varepsilon_o}\right]^{1/2} - \frac{qN_D}{C_{ox}} \qquad 5.2\text{-}14$$

The maximum output charge packet is also given by Eq. 5.2-13 with A_B replacing A_A.

Effect of Stray Capacitance

The effect of stray capacitance is primarily to attenuate the output of the charge packet differencer. Because the stray capacitance is a function of effective differencer electrode voltage, the associated change in attenuation factor introduces non-linearity into the transfer characteristic. The stray capacitance can be divided into three major components. The first is the stray capacitance due to the floating p^+ diffusions. The second is due to the underlap capacitance between the floating p^+ diffusions and adjacent transfer electrodes. The third component is due to stray wiring capacitances. Because the wiring can be done over thick oxide, it can be quite small. The underlap capacitance can be reduced by using a self-aligned implanted diffusion process. The floating p^+ diffusion can be minimized by reducing the area of these regions.

To quantitatively calculate the effect of stray capacitance one begins by writing a statement of charge conservation on the differencer electrodes:

$$\underline{Q}'_A + \underline{Q}'_{SCA} = \underline{Q}_A + \underline{Q}_{SCA} + \underline{Q}_{SA} \qquad 5.2\text{-}15a$$

$$\underline{Q}'_B + \underline{Q}'_{SCB} = \underline{Q}_B + \underline{Q}_{SCB} + \underline{Q}_{SB} \qquad 5.2\text{-}15b$$

where Q_{SCA} and Q_{SCB} are the negative charges stored on the stray capacitance. Substituting these relationships into Eq. 5.2-4 yields

$$Q_0 = \frac{A_B}{A_A} Q_{SA} - Q_{SB} - \Delta Q_0 \qquad 5.2\text{-}16$$

where

$$\Delta Q_0 = \frac{A_B}{A_A} (Q'_{SCA} - Q_{SCA}) - (Q'_{SCB} - Q_{SCB}) \qquad 5.2\text{-}17$$

The difference in charge stored on the stray capacitance can be expressed approximately by the product of the differential stray capacitance (C_{SCA}, C_{SCB}) and the change in differencer electrode effective voltage.

$$Q'_{SCA} - Q_{SCA} \approx C_{SCA} (V'_A - V_A) \qquad 5.2\text{-}18a$$

$$Q'_{SCB} - Q_{SCB} \approx C_{SCB} (V'_B - V_B) \qquad 5.2\text{-}18b$$

Substituting these relationships into 5.2-17 yields

$$\Delta Q_0 = \frac{A_B}{A_A} C_{SCA} (V'_A - V_A) - C_{SCB} (V'_B - V_B) \qquad 5.2\text{-}19$$

If the electrode areas are equal, and if the differential stray capacitances are equal (C_{SC}), Eq. 5.2-19 can be simplified to

$$\Delta Q_0 = C_{SC} (V'_A - V'_B) \qquad 5.2\text{-}20$$

since V_A and V_B are equal.

Finally, recognizing that

$$-Q_0 = C_{OX} (V'_B - V'_A) \qquad 5.2\text{-}21$$

one obtains

$$\Delta Q_0 = \frac{C_{SC}}{C_{OX}} Q_0 \qquad 5.2\text{-}22$$

and Eq. 5.2-16 becomes

$$Q_0 = \alpha_{SC} \left[Q_{SA} - Q_{SB} \right] \qquad 5.2\text{-}23$$

where the stray capacitance attenuation factor α_{SC} is defined as

$$\alpha_{SC} \triangleq (1 + C_{SC}/C_{OX})^{-1} \qquad 5.2\text{-}24$$

Although, in general, the stray capacitance attenuation effect is undesirable, it is possible to convert this to a feature. If the stray capacitance can be controlled by a bias voltage, or more ideally, by a charge packet (perhaps under floating gate capacitor), it may be possible to construct a device which does charge packet differencing and multiplication (division). This possibility is not pursued further here, but it may be worthy of future investigation.

Effect of PN Junction Leakage Current

The floating p^+ diffusions in the charge packet differencer are, of course, reverse biased PN junctions in which a non-zero reverse 'leakage' current flows. This current discharges the differencer electrodes in the same manner discussed in appendix D regarding amplifier droop. The discharge rate is simply

$$-\dot{Q}_A = J_{LA} \cdot A_{DA} \qquad 5.2\text{-}25a$$

$$-\dot{Q}_B = J_{LB} \cdot A_{DB} \qquad 5.2\text{-}25b$$

where J_{LA} and J_{LB} are the current densities through the floating diffusions with areas A_{DA} and A_{DB} respectively. These currents depend upon the quality of the diffusions and the effective differencer electrode bias. The higher the bias, the larger the current. For poor quality diffusions, soft breakdown of the junction may occur for large precharge biases leading to large and unpredictable current levels.

If an interval Δt elapses between the time the electrodes are precharged and the time the output charge packet is formed, the total discharge is given by

$$\Delta Q_A = - J_{LA} A_{DA} \Delta t \qquad 5.2\text{-}26a$$

and

$$\Delta Q_B = - J_{LB} A_{DB} \Delta t \qquad 5.2\text{-}26b$$

assuming a constant discharge current during the interval. The output

charge packet, assuming equal floating diffusion areas, is determined by

$$Q_o = \alpha_{SC} \left[Q_{SA} - Q_{SB} - A_D \Delta t \, (J_{LA} - J_{LB}) \right] \qquad 5.2\text{-}27$$

It is interesting to note that if the discharge currents are equal, there is a net null effect on the output charge packet except that the saturation level will be reduced. However, in most cases of interest the B electrode will have a larger effective voltage than the A electrode, and J_{LB} will be larger than J_{LA}. The consequence will be increased attenuation for increased output charge packets, and increased attenuation for longer intervals. The latter effect will be particularly pronounced for poor junctions when the differencer is operating in a multiple cycle copier mode.

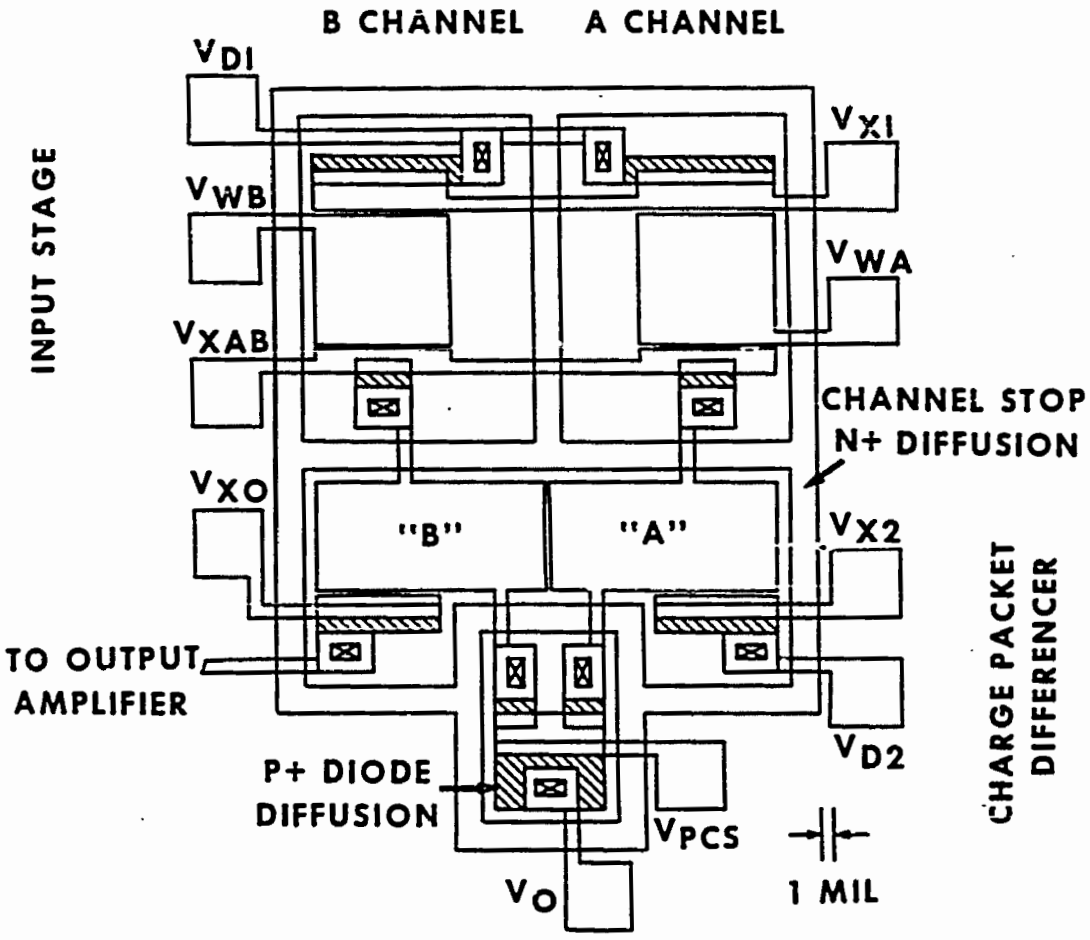
Effect of Dark Current

The differencer circuit is virtually immune to bucket dark current in the same way as the Tompsett surface potential equilibration method is immune. Dark current collected in either the spill bucket or the metering well will either result in premature formation of the charge packet, or be spilled toward the input diode if the metering well is already full. The exception to this immunity occurs during the output transfer phase when dark current integrated over the output transfer phase interval may be included in the output packet.

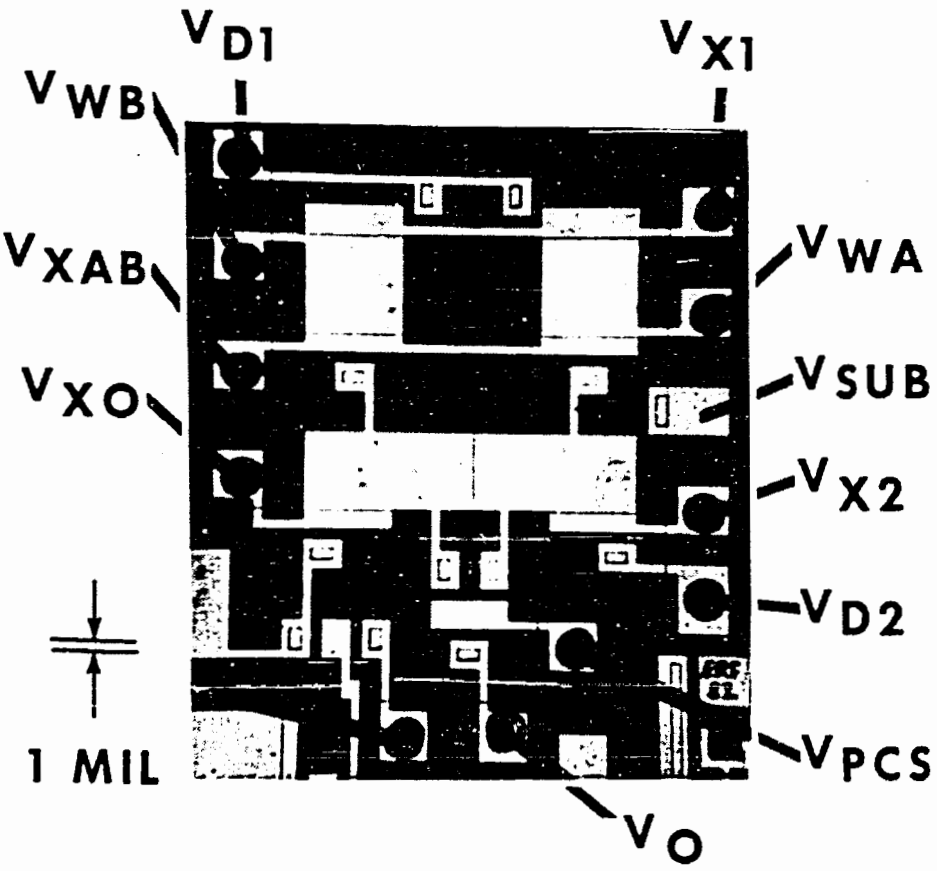
5.3 Device Layout and Fabrication

Yale 8205 Layout

The layout of the charge packet differencing experiment fabricated on the Yale 8205 test chip is shown in Fig. 5.3-1, and a photograph of an actual device is shown in Fig. 5.3-2. There are three major stages to this layout; the input charge packet formation stage, the differencer



5.3-1 Layout of fabricated experimental device showing input stage (A and B channels) and charge packet differencer. Interelectrode gap size exaggerated.



DEVICE 7B20

5.3-2 Photograph of fabricated device (7B20).

circuit, and the output stage. The latter is discussed in detail in appendix D.

The input stage consists of two parallel but isolated fill and spill charge packet formation channels, A and B. Each channel has an input diode which are connected in parallel and biased by V_{D1}. The input diodes are gated by spill electrodes which are also connected in parallel and biased by V_{XI}. Each channel has an independent metering well biased by V_{WA} and V_{WB}. The metering well sizes are 101 $mils^2$ each. Using these two electrodes and input diode, each channel can produce signal charges Q_{SA} and Q_{SB} for input into the charge packet differencer. Each input structure terminates with an output transfer gate controlled by V_{XAB} followed by a floating p^+ diffusion. The area of the floating diffusion is 20 $mils^2$ which includes a 4 $mils^2$ underlap with the output transfer gate. As will be discussed in section 5.5, this underlap constitutes a major portion of the differencer electrodes' stray capacitance.

The floating p^+ diffusion is connected via a 1 mil wide wire over an isolation channel stop to the charge packet differencer electrode. The area of each differencer electrode is 138 $mils^2$. Each differencer electrode is also connected to a floating diffusion of the precharge MOSFET (also isolated by the channel stop diffusion). The area of this diffusion is 18 $mils^2$. The A differencer electrode is adjacent to an input gate (V_{X2}) and input diode (V_{D2}). The B differencer electrode is juxtaposed an output gate (V_{XO}) and a floating diffusion wired to the output sense amplifier MOSFET gate (as described in appendix D).

The circuit shown in Fig. 5.3-1 was designed to be compact, reduce bonding sites, and incorporate as much symmetry as possible. The latter was chosen for balance so that parasitic effects might tend to cancel.

In fact, although the bonding pad count was an issue, as was compaction, these self-imposed constraints led to a design which, with hindsight, was not optimized. The circuit performs as designed and certainly demonstrates the basic principles of its operation. However, the author feels it might be instructive for future investigators if a few words are devoted to a critique of this layout.

The most serious error committed in this layout was that the A and B input channels, and A and B precharge switches were tied in parallel. Although this reduces the bonding pad count, experimentally it makes it impossible to independently bias each differencer electrode. This type of operation is important when testing of the chip is underway and faults are being traced. For example, with the present layout there is no way to test for a short between the two differencer gates.

A second oversight was that the W/L ratio of the differencer electrodes was made less than unity. This was convenient from a compaction point of view, but the effect this layout has on charge transfer rates was underestimated. As a consequence, the differencer operates much slower than it might were the layout optimized.

Finally, the stray capacitance effects were not fully analyzed at the time of layout so that floating p^+ diffusion capacitance, underlap capacitance and wiring capacitance were not reduced to their minimum values. Optimizing these parameters would have led to a greater demonstrable accuracy in the Yale 8205 experimental device.

Although the above issues are glaringly obvious in the course of retrospection, the reader should take note that when the Yale 8205 chip was laid out, primary concern was concerned with broader questions, such as if the channel stop isolation technique would work, or if the sub-micron gap technology would work, or even if the predicted operation of

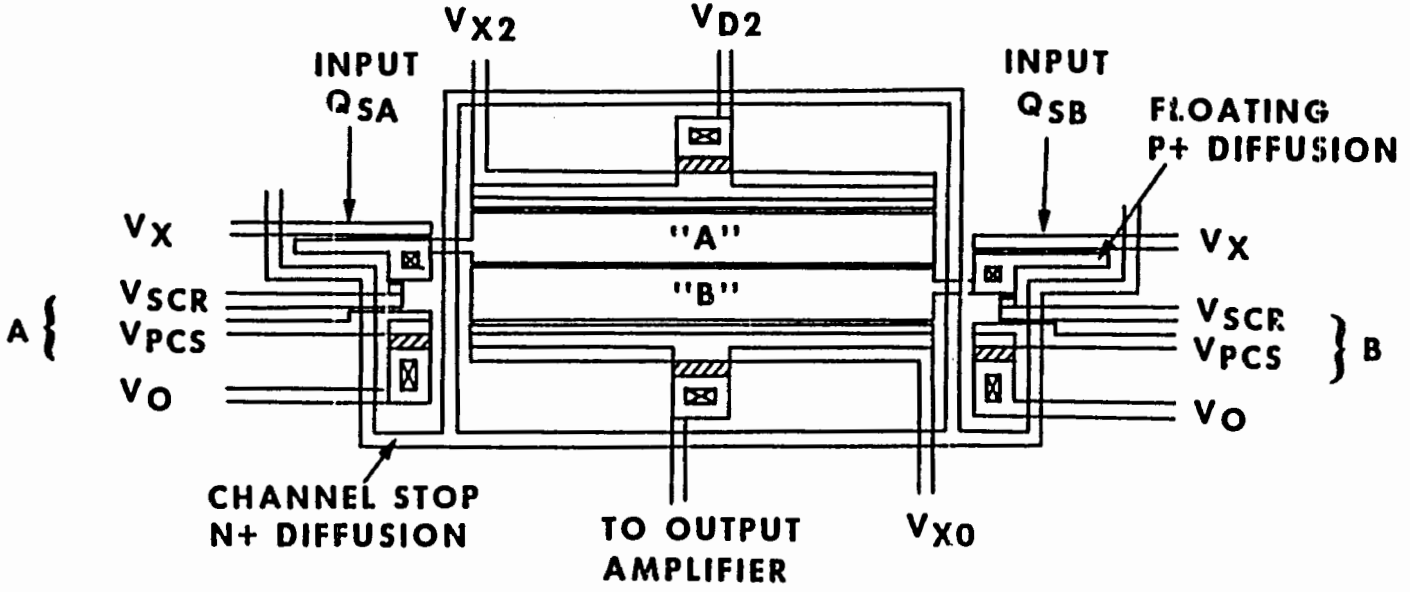
the differencer was indeed correct. Lesser details such as minimizing stray capacitance escaped attention at the time.

Proposed Improved Layout

In light of the observations made in the preceeding paragraphs, an improved layout suitable for fabrication at Yale without process modification is presented in Fig. 5.3-3. The changes include stretching out the differencer electrodes to achieve a greater W/L ratio without changing the area. There is no reason not to continue to stretch the geometry other than that the sub-micron gap yield is reduced for longer gaps. In the improved layout, the W/L ratio has been increased from its previous value of approximately 2/3 to 17/2, an improvement factor of over an order of magnitude. The layout could probably be stretched another factor of two. As proposed, the charge transfer time will be improved by more than an order of magnitude since the transfer time for diffusion limited transfer scales as L^2 [Carnes et al. 1972].

Another change is that the stray capacitance has been reduced as much as possible. In the improved layout, the channel stop diffusion (whose width has been cut in half) is covered with thick oxide, reducing wiring capacitance. The floating p^+ diffusion area has been reduced by using the same diffusion for both the precharge switch and input charge collector. The underlap capacitance has been significantly reduced by eliminating the underlap for input charge transfer which was unnecessary, and decreasing the size of the precharge MOSFET switch, which was overdesigned in the Yale 8205 layout.

The third improvement (which may be unnecessary) was to change the precharge MOSFET into a dual gate device. In this new design, the gate closest to the floating p^+ diffusion acts as a screen gate which is d.c. biased, and the second gate performs the MOSFET switch function as



5.3-3 Proposed improved layout of charge packet differencer.

before. The purpose of the screen gate is to reduce 'reset noise' when the differencer electrodes are precharged to potential V_D. This is analogous to the output amplifier screen gate function described in Beynon and Lamb (1980, pg. 192). The cost of the screen gate is an added bias in the circuit and increased effective floating diffusion area. The utility of such a screen gate may not be apparent in a large geometry device when there are a billion or so electrons in the bucket, but for scaled geometries, the number of electrons decreases as L^2. In these cases, the screen gate may perform a necessary noise reduction function.

5.4 Experimental Results and Analysis

Output Amplifier Calibration

In appendix D, a simple theory describing the operation of the source-follower charge packet output sense amplifier was developed. The change in output voltage of the source-follower configured MOSFET was found to be related to a charge packet Q_{SIG} according to

$$\Delta V_{OUT} = Q_{SIG}/C_{OUT} \qquad 5.4\text{-}1$$

where C_{OUT} is the transcapacitance of the output MOSFET. The transcapacitance can be written in terms of the lumped output MOSFET gate capacitance C_{AG} and the amplifier gain α_A such that

$$C_{OUT} = \alpha_A \, C_{AG} \qquad 5.4\text{-}2$$

By measuring the amplifier gain α_A and transcapacitance C_{OUT}, the physical MOSFET gate capacitance can be ascertained.

To calculate the gain of the source-follower configuration, the precharge MOSFET switch was maintained in the closed or on condition and the voltage V_{OA} was used to control the output MOSFET gate voltage. The output voltage V_{OUT} was measured as a function of V_{OA} for several values

of load resistance. This measurement for device 7B6 is shown in Fig. 5.4-1. The gain of the source-follower amplifier is experimentally determined from the slope of the transfer characteristic using

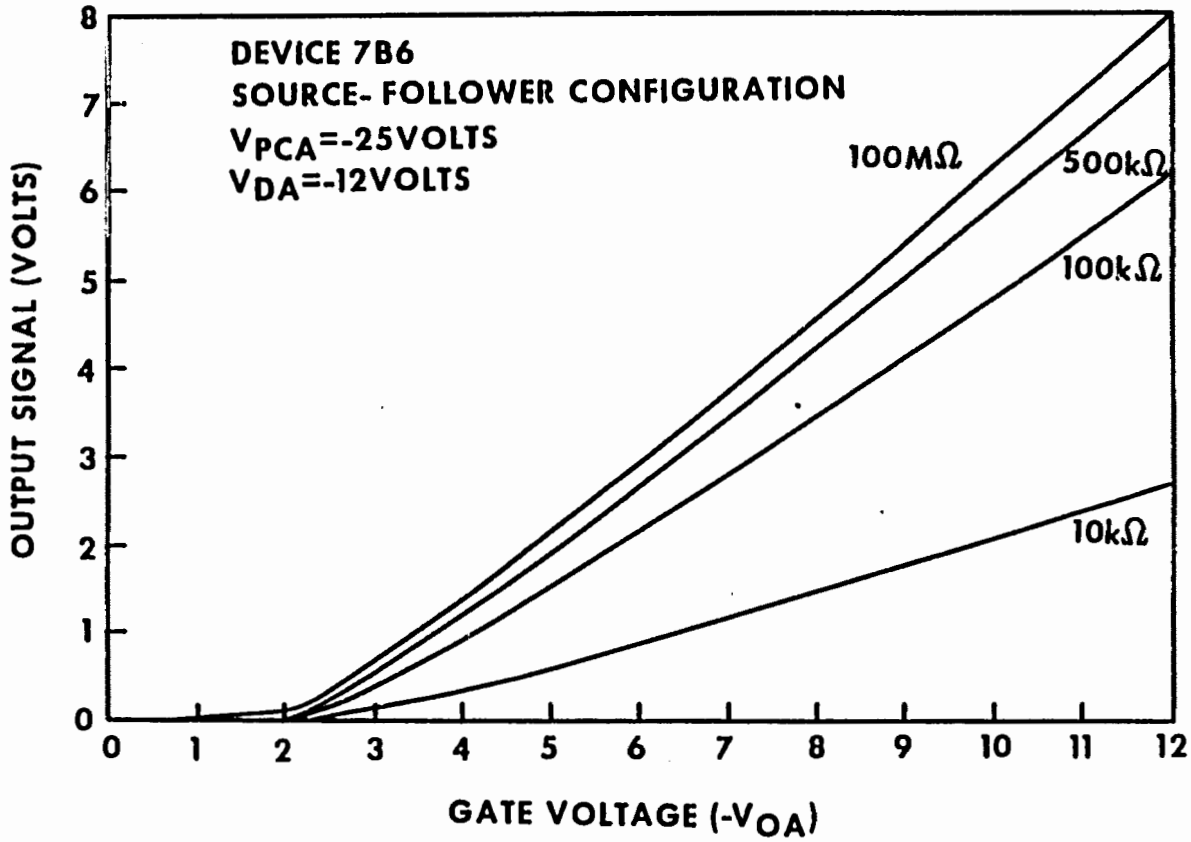
$$\alpha_A = dV_{OUT}/dV_{OA} \qquad 5.4\text{-}3$$

It is seen from the displayed measurement that the gain improves for increasing load resistances. From appendix D, note that

$$\alpha_A = 1 - \left[\frac{2R_L\mu C_{ox}W}{L}\ (V_{OA} - V_T) + 1\right]^{-1/2} \qquad 5.4\text{-}4$$

so that this dependence is expected. In order to cause the observed transfer characteristic to agree with Eq. 5.4-4, an effective value of W/L = 3-5 must be used, which gives good agreement over the range of load resistances examined. This same effective W/L value was observed when the normal (grounded source) I-V characteristics of the output MOSFET were analyzed. However, from the layout shown in appendix D, the design value of W/L was 83. Such a discrepancy can be explained in terms of high impedance inter-digitated source and drain fingers in the output MOSFET. The ohmic voltage generated in these fingers would be sufficient to cause most of the device current to flow at the base of the fingers. Indeed, the measured sheet resistance of the diffused p^+ layer for wafer 7B was 2.1 kΩ/□ due to a defective dopant source. (This problem was corrected and in lot 9 and the sheet resistance was reduced to 0.015 kΩ/□). For wafer 7B, the consequence of such a high diffused layer sheet resistance was to increase the minimum load resistance required to achieve near unity gain of the output amplifier. For wafer 7B devices, a load resistance of 1 MΩ in parallel with an oscilloscope impedance of 1 MΩ was chosen. From the slope of the 500 kΩ transfer characteristic, a gain α_A of 0.81 is observed.

To determine the output sense amplifier transcapacitance, a



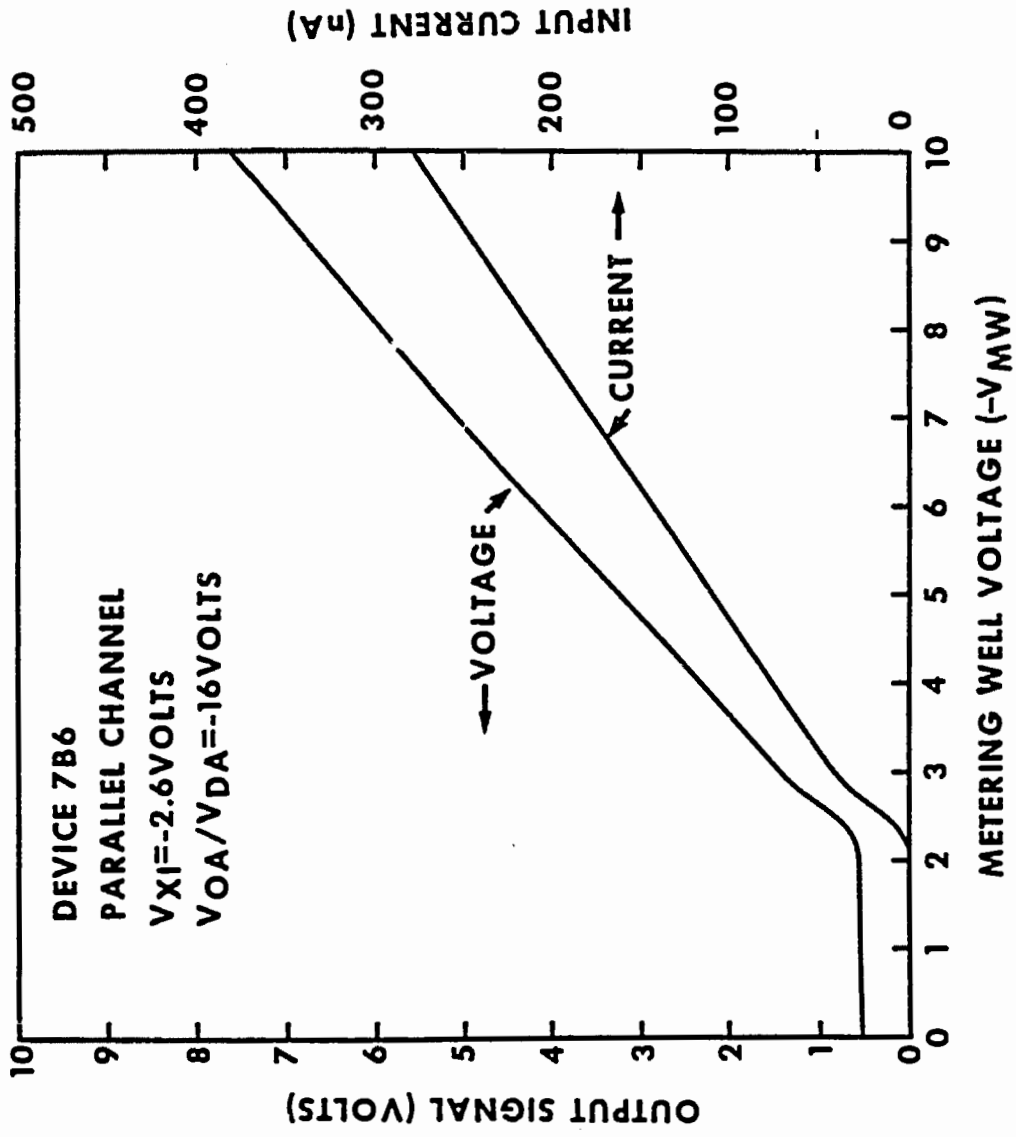
5.4-1 Measurements of source-follower configured output MOSFET for various load resistance.

measurement of output voltage versus input charge packet was performed. For this measurement, the parallel channel of the charge packet initiated MTOS junction switching experiment was utilized. A charge packet was formed using only the fill phase of the normal fill and spill technique. In this mode, the input diode is connected to ground (substrate) through an electrometer operating as an ammeter and the input gate is biased just above threshold so that a small current can flow between the diode and the metering well bucket. This current flows until the metering well bucket is filled and then ceases as the equivalent MOSFET source-drain voltage goes to zero. When the output transfer gate is pulsed, the charge transferred to the output sense amplifier (Q_{SIG}) consists of three components; the charge in the metering well bucket (proportional to the metering well voltage minus V_T), the charge under the input transfer gate (likewise proportional to the input transfer gate bias minus V_T), and the integrated current which flows from the input diode during the transfer interval. The input transfer gate bias is adjusted so that this latter component is small yet the current sufficient to fill the bucket in between output transfer cycles.

The total charge transferred during each cycle, when averaged over the cycle period T, appears as a d.c. current through the input diode as measured by the ammeter. The current sensed by the ammeter is

$$I = Q_{SIG}/T \qquad 5.4\text{-}5$$

This current as a function of metering well voltage (V_{MW}) is shown in Fig. 5.4-2. The current over the metering well voltage range -5 to -10 volts is reasonably linear and has a equivalent slope $dQ_{SIG}/\ dV_{MW}$ of approximately 34.0 pC/volt with a cycle period of one millisecond. This is close to what is predicted from the oxide capacitance.



5.4-2 Calibration of output charge packet sense amplfier by injection of charge packet from parallel channel. Measured current is ratio of injected charge packet to cycle period (one millisecond). Output signal of amplifier referenced to reset level.

Also shown in Fig. 5.4-2 is the change in output voltage (referenced from the reset level and hereafter referred to as the output signal) of the charge packet sense amplifier as a function of metering well voltage. The slope of this transfer characteristic dV_{OUT}/dV_{MW} is 0.90. The transcapacitance is simply calculated from the ratio of the two above slopes as

$$C_{OUT} = (dQ_{SIG}/dV_{MW}) \; / \; (dV_{OUT}/dV_{MW}) \qquad 5.4\text{-}6$$

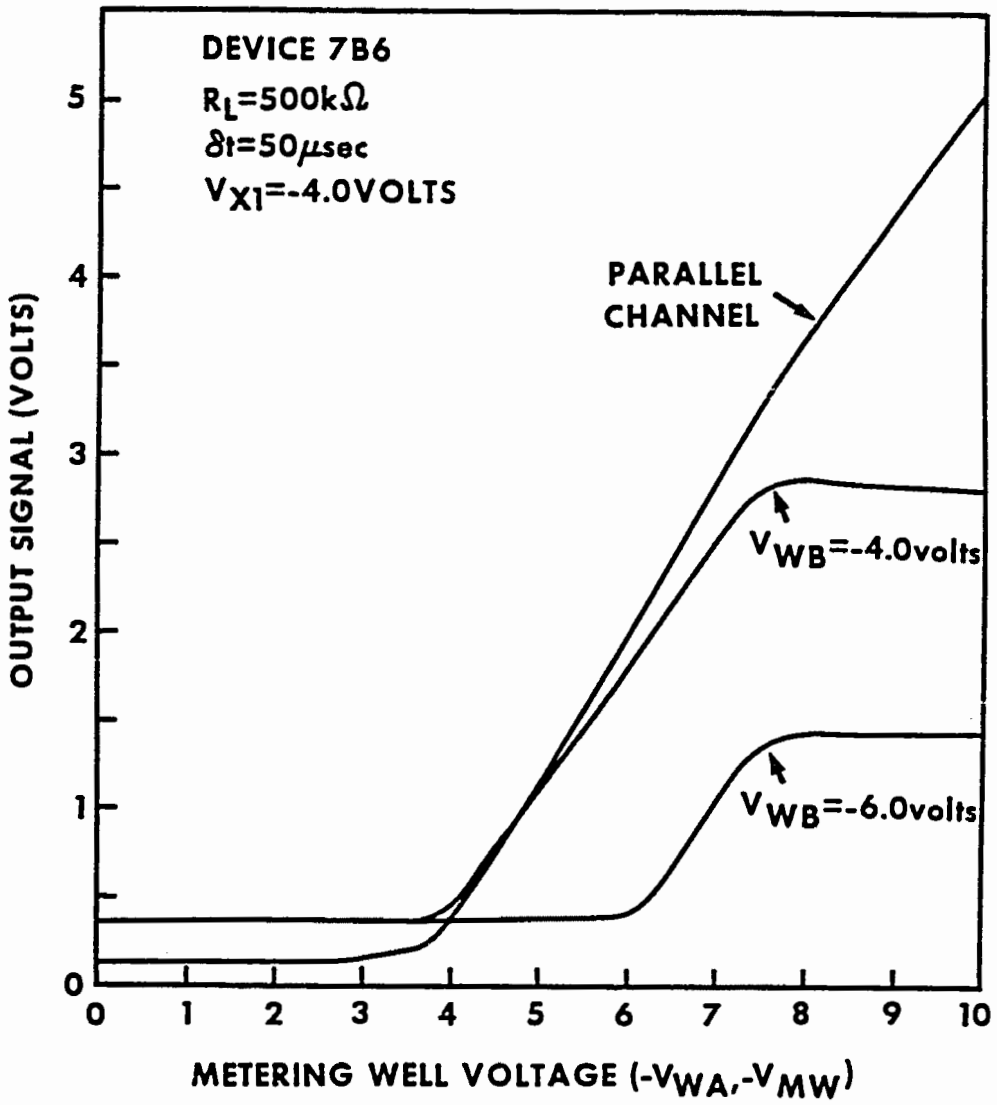
and has a value of 37.7 pC/volt for device 7B6.

Combining the amplifier gain measurement and transcapacitance measurement yields an inferred lumped output MOSFET gate capacitance (using Eq. 5.4-2) of 46.5 pF. This can be compared to a predicted value of 43 pF as computed in appendix D using the fabricated characteristics of wafer 7B. The results are reassuringly close.

Charge Packet Replicator

The charge-packet initiated MTOS junction switching experiments' parallel channel was again utilized to test the charge packet replicator mode of operation of the charge packet differencer experiment. The parallel channel was clocked to perform charge packet formation using the fill and spill technique with the input transfer gate biased at -4.00 volts. The output signal was measured as a function of metering well voltage (V_{MW}) as shown for device 7B6 in Fig. 5.4-3. The non-zero value of the output signal for no charge packet (V_{MW} smaller than -4.0 volts) is due to a combination of non-ideal behavior including amplifier droop and dark current collected during the output transfer phase. The slope of the transfer characteristic is 0.87, which when multiplied by the output transcapacitance indicates that input charge packet size is 33 pC per metering well volt.

The charge packet differencer operating in a replicator mode was also



5.4-3 Charge packet replicator operation compared to parallel channel input as discussed in text.

measured on device 7B6. In this mode, V_{WB} was set equal to the input transfer gate bias which was equal to -4.0 volts. The A input channel metering well bias was ramped from zero to -10.0 volts and the created charge packet was fed into the differencer. The output signal from the differencer was measured as a function of V_{WA} as shown in Fig. 5.4-3. Again, a similar non-zero output signal is seen for V_{WA} smaller than -4.0 volts, which is attributed to the non-ideal behavior discussed previously but amplified due to the longer overall cycle time of the differencing operation. The slope of the transfer characteristic indicates that the differencer is producing 27 pC per A channel metering well volt. From the previous measurement, the replicator ideally would be producing 33 pC per metering well volt so that differencer has a gain (i.e. attenuation factor) of 0.82.

This attenuation factor can be explained by invoking the stray capacitance effect discussed in section 5.2, as exemplified by Eq. 5.2-23. To estimate the stray capacitance we add the three major components. First, the underlap capacitance between the floating p^+ diffusions and adjacent electrodes contributes a capacitance of 0.38 $pF/mils^2$ multiplied by 7 $mils^2$ or 2.7 pF. The floating p^+ diffusion capacitance (to substrate) was calculated in appendix D and estimated to be approximately 60 $fF/mils^2$ yielding 2.2 pF. Finally, the stray wiring capacitance is dominated by where the wire crosses the channel stop diffusion over gate oxide which has a capacitance of 0.38 $pF/mils^2$ multiplied by 4 $mils^2$ or 1.5 pF. The total stray capacitance is 6.4 pF.

The effective oxide capacitance of the differencer electrode is 45 pF. Using Eq. 5.3-9, the predicted attenuation factor is 0.87. The agreement between the experimentally observed attenuation factor and this predicted value is reasonable, but the difference indicates that the stray capacitance may be larger than actually calculated.

When the B input channel is used, the device is no longer operating in the replicator mode, but it is convenient to discuss the effect briefly. Applying V_{WB} = -6.00 volt results in the formation of a two volt bucket and charge packet of 66 pC. Using Eq. 5.2-23 and the output transcapacitance, the output signal should be reduced by 1.45 volts. In Fig. 5.4-3, the results of this measurement are shown and it can be seen that indeed the transfer characteristic is shifted down by the predicted amount indicating proper operation of the differencer circuit. Note that the knee of the transfer characteristic is shifted two volts to the right but the 'zero' level is unaffected. Finally, as a preview to an upcoming sub-section, observe that the output saturates for approximately a four volt A channel input bucket independently of the B channel input bucket size.

Charge Packet Differncer Diagnostic Procedure

Prior to discussing the results of testing a fully functional charge packet differencer, it is useful to discuss the techniques developed to diagnose modes of differencer failure. The procedure outlined below can also serve as a vehicle for fine tuning the circuit. The overall plan is to work backwards beginning from the output amplifier and heading towards the input circuits.

1. Amplifier source-follower test

    a. Apply -20 volt bias to V_{DA}
    b. Apply -25 volt bias to V_{PCA}
    c. Apply -20 volt bias to V_{OA}

In the course of applying this latter bias, the output voltage V_{OUT} (measured across a suitable resistor) should follow and track V_{OA} once V_{OA} exceeds V_T (see Fig. 5.4-1).

    d. Operate V_{PCA} in a pulse mode (0, -25 volts)

Output amplifier should exhibit reset clock feedthrough effects and perhaps droop. Droop should not exceed 10-20 mV/μsec.

    e. Turn off V_{PCA} (zero volts)

Output voltage should gradually return to zero, depending upon droop rate.

f. Return V_{PCA} to normal pulse mode operation

2. Differencer potential well test

a. Apply -25 volt bias to V_{PCS}
b. Apply -20 volt bias to V_0
c. Apply -4 volt bias to V_{X2}
d. Apply pulse to V_{XO} (0, -25 volts)
e. Apply pulse to V_{D2} (0, -25 volts)

Output amplifier should display large (possibly saturated) response. Reducing V_0 should reduce output signal which should go to zero when V_0 is smaller than V_{X2}.

f. Apply -20 volt bias to V_0
g. Apply pulse to V_{PCS} (0, -25 volts)

Output amplifier should display large (possibly saturated) response, but not as large as when V_{PCS} was held 'on'. Reducing V_o should reduce output signal which should go to zero when V_0 is smaller than V_{X2}.

h. Apply -20 volt bias to V_0

Check that increasing V_{X2} decreases the output signal which should go to zero when V_{X2} is increased to -20 volts.

i. Apply -20 volt bias to V_{X2}

3. Replicator mode test

a. Apply -4 volt bias to V_{X1}
b. Apply -7 volt bias to V_{WA}
c. Apply pulse to V_{XAB} (0, -25 volts)
d. Apply pulse to V_{D1} (0, -25 volts)

Output amplifier should display an output signal. Increasing V_{WA} should increase output signal. Decreasing V_{WA} below V_{X1} should result in output signal decreasing to zero.

e. Apply -6 volt bias to V_{XA}

Increasing V_{X1} should cause output signal to decrease which should go to zero when V_{X1} is equal to V_{WA}.

f. Apply -4 volt bias to V_{X1}

4. Differencer Test

Increasing V_{WB} from -4 volts to -6 volts should cause decreasing output signal which should go to zero when V_{WB} is equal to -6 volts. Increasing V_{WA} should increase output signal which should be cancelled by an equal increase in V_{WB}.

5. Fine Tuning

   a. Apply -4 volt bias to V_{WB}
   b. Apply low frequency triangular wave to V_{WA} (-4 to -8 volts)

Oscilloscope display of output synchronized by triangular wave should show a reasonably linear transfer characteristic. Increasing V_{WB} should uniformly decrease envelope (see Fig. 5.4-3). Adjust cycle time, all dc biases, clocked biases' high and low levels, and transition rates until performance is optimized with respect to linearity, dynamic range, zero standoff, clock feedthrough and transfer efficiency.

This last instruction is deliberately vague. Tuning a CCD circuit is an acquired skill which appears to be a black magic art to the inexperienced. This simple differencer experiment can be considered to be a black box with 14 control inputs and one output, the latter of which must be optimized through the 36 parameters which characterize the 14 inputs. The challenge of tuning the circuit is compounded because there are many local optima in parameter space. Only by considering the device physics of the internal workings of the black box (grey box), can one hope to achieve the global optimum in a finite period of time.

Charge Packet Differencer Performance Verification

The charge packet differencer creates a charge packet Q_O equal to the difference of two input charge packets Q_{SA} and Q_{SB} according to

$$Q_O = \left| Q_{SA} - Q_{SB} \right| \qquad 5.4\text{-}7$$

For the Yale 8205 experimental structure, the input charge packets are proportional to the input channel metering well voltages V_{WA} and V_{WB} such that

$$-Q_{SA} = C_{OX} (V_{WA} - V_{X1}) \qquad 5.4\text{-}8a$$

and $$-Q_{SB} = C_{OX} (V_{WB} - V_{X1}) \qquad 5.4\text{-}8b$$

The output transconductance relates the output charge packet to the output signal according to

$$\Delta V_{OUT} = Q_O / C_{OUT} \qquad 5.4\text{-}9$$

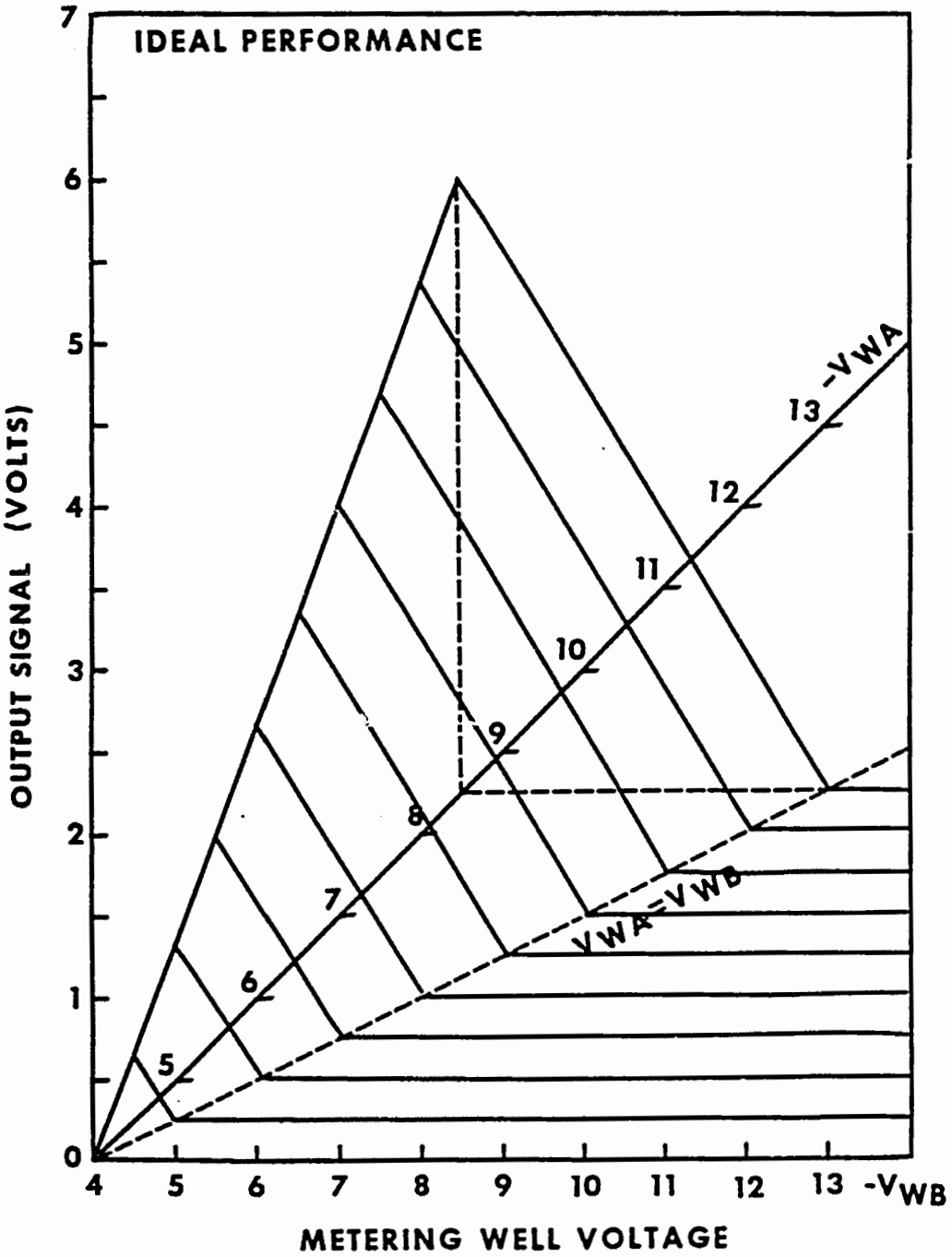
Combining these equations and including the stray capacitance yields the relationship between input channel metering well voltages and output signal,

$$\Delta V_{OUT} = \alpha_{SC} \frac{C_{ox}}{C_{OUT}} (V_{WA} - V_{WB}) \qquad 5.4\text{-}10$$

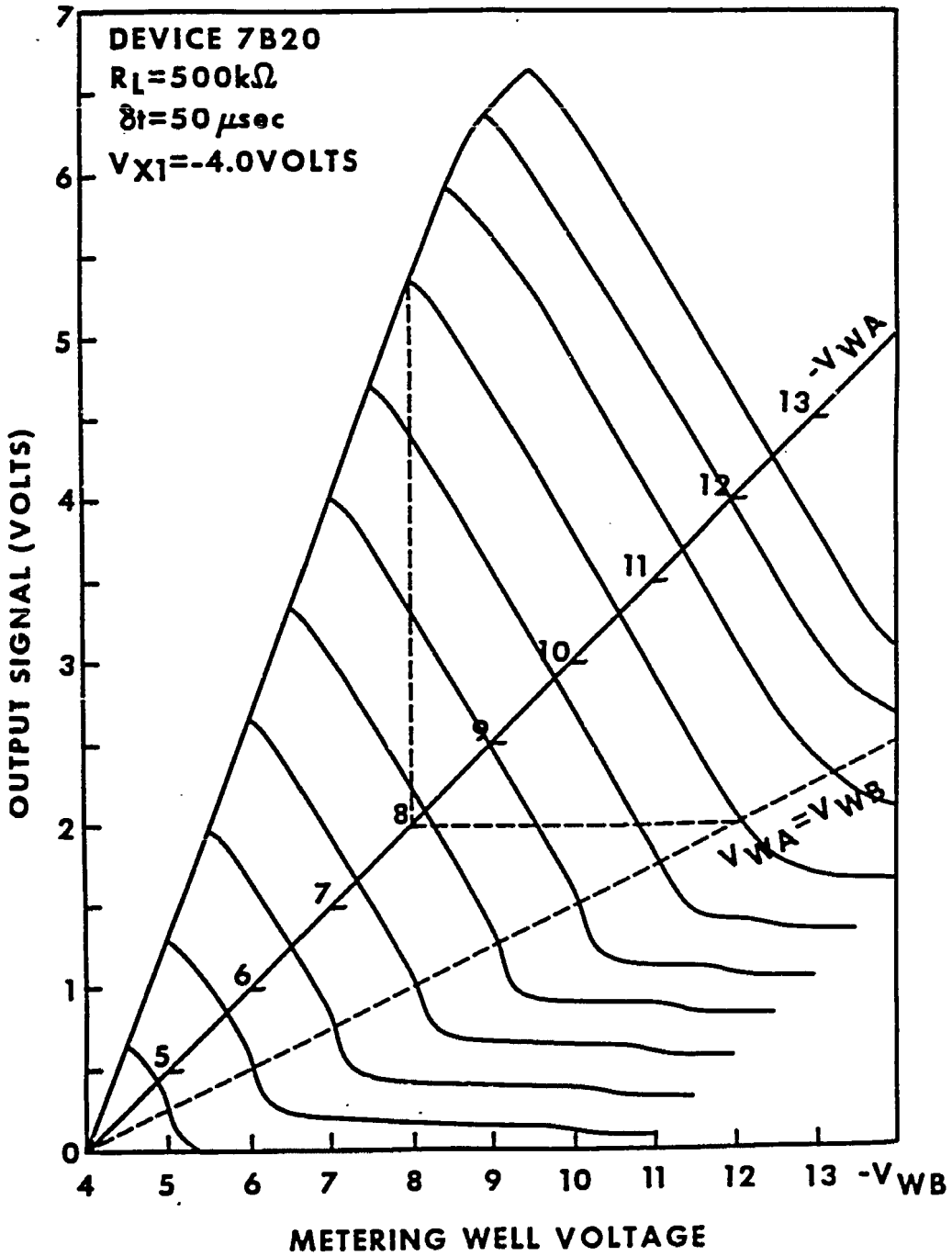
where the effect of stray capacitance has been included. This last equation is valid for V_{WA} and V_{WB} larger than V_{X1}, and V_{WA} larger than V_{WB}. The output signal should be zero otherwise. A plot of the predicted performance of the device using the parameter values determined previously is shown in Fig. 5.4-4. The plot shows a planar surface which intersects the ordinate axes plane along the line $V_{WA} = V_{WB}$.

The semi-empirical predicted performance may be compared to a plot showing the measured transfer characteristics of device 7B20, which is shown in Fig. 5.4-5. The applied biases are summarized in Table 5.4-1. In Fig. 5.4-5, the output signal is referenced to the case when V_{X1}, V_{WA} and V_{WB} are biased at -4.0 volts. The output signal appears to be negative when V_{WB} is biased above V_{WA}. This is because increasing V_{WB} helps suppress dark current collected under the B channel differencer electrode through Q_{SB} and the reduced effective electrode bias.

In the active region of the transfer characteristic we can define the



5.4-4 Ideal performance transfer characteristics of differencer circuit. Output gain calculated using experimentally observed values.



5.4-5 Transfer characteristic surface of experimental charge packet differencer circuit. Output signal referenced to reset level shifted by 0.8 volts.

Table 5.4-1

Differencer Set-up Voltages etc.

Device		7B20
Fundamental clock period	δt	50 μsec
Output MOSFET load resistance	R_L	500 kΩ
Output MOSFET drain bias	V_{DA}	-20 volts
Output MOSFET precharged gate bias	V_{OA}	-20 volts
Precharge MOSFET pulse (amplifier)	V_{PCA}	0,-25 volts
Differencer precharge bias	V_O	-20 volts
Precharge MOSFET pulse (differencer)	V_{PCS}	0,-25 volts
Input isolation gate bias	V_{X2}	-20 volts
Input diode pulse (differencer)	V_{D2}	-25,0 volts
Output transfer gate pulse (differencer)	V_{XO}	0,-25 volts
Input diode pulse (input stage)	V_{D1}	-25,0 volts
Input transfer gate bias	V_{X1}	-4 volts
Output transfer gate pulse (input stage)	V_{XAB}	0,-25 volts

differencer output gains α_{DA} and α_{DB} as

$$\alpha_{DA} = \left.\frac{\partial V_{OUT}}{\partial V_{WA}}\right|_{V_{WB}} \qquad 5.4\text{-}11a$$

and

$$\alpha_{DB} = \left.\frac{\partial V_{OUT}}{\partial V_{WB}}\right|_{V_{WA}} \qquad 5.4\text{-}11b$$

which according to Eq. 5.4-10 should be equal to each other and

$$\alpha_{DA} = \alpha_{DB} = \alpha_{SC}\, C_{OX}/C_{OUT} \qquad 5.4\text{-}12$$

which is equal to 0.91 using the parameters evaluated for device 7B6. For device 7B20, the gains are $\alpha_{DA} = 0.89$ and $\alpha_{DB} = 0.86$ depending slightly on the operating region in which they are evaluated. The difference between the 7B6 predicted gains and the 7B20 observed gains is insignificant and it is concluded that the device is operating properly.

The transfer characteristic indicates that the differencer saturates for an A channel bucket larger than approximately 5 volts. From device 7B6, this corresponds to 165 pC. Using Eq. 5.2-13, the maximum charge packet prior to saturation is 162 pC so that the agreement, although somewhat fortuitous, validates the understanding of the saturation mechanism.

The transfer characteristic of Fig. 5.4-5 demonstrates the successful implementation of the charge packet differencing concept introduced in this chapter. This transfer characteristic will be revisited in the discussion of device linearity in the next subsection.

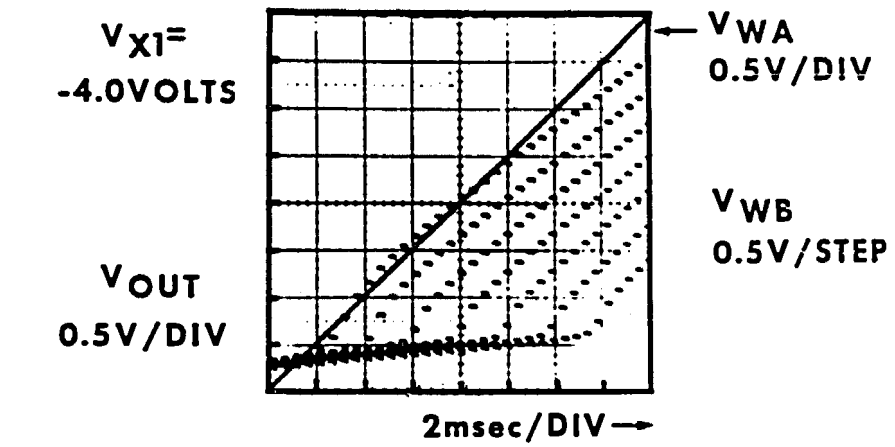
In Figs. 5.4-6a and 5.4-6b, multiple exposure photographs showing the raw output signal from the differencer are displayed. In the first

5.4-6 Multiple exposure photographs showing actual output of charge packet differencer circuit as sensed by output amplifier.

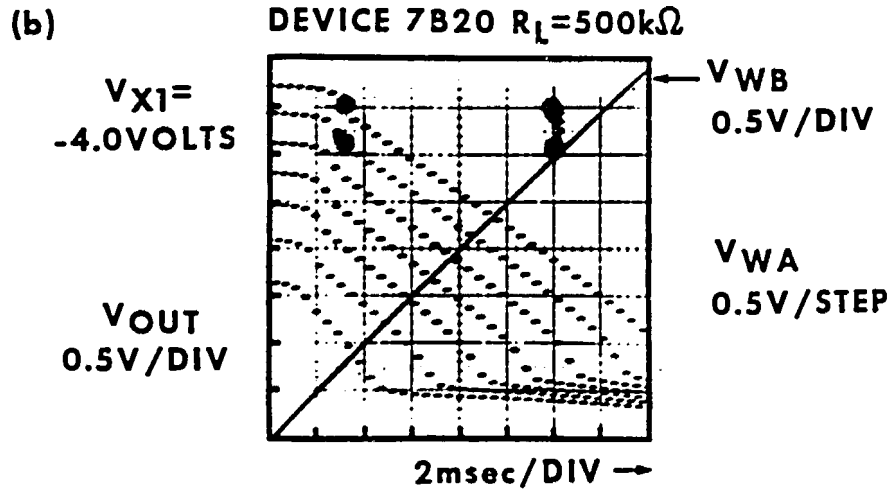
a. Output as a function of channel A input charge packet. V_{WA} ramped from -8.5 volts to -3.5 volts left to right. V_{WB} stepped from -5.0 volts to -8.0 volts in -0.5 volt increments beginning from maximum response.

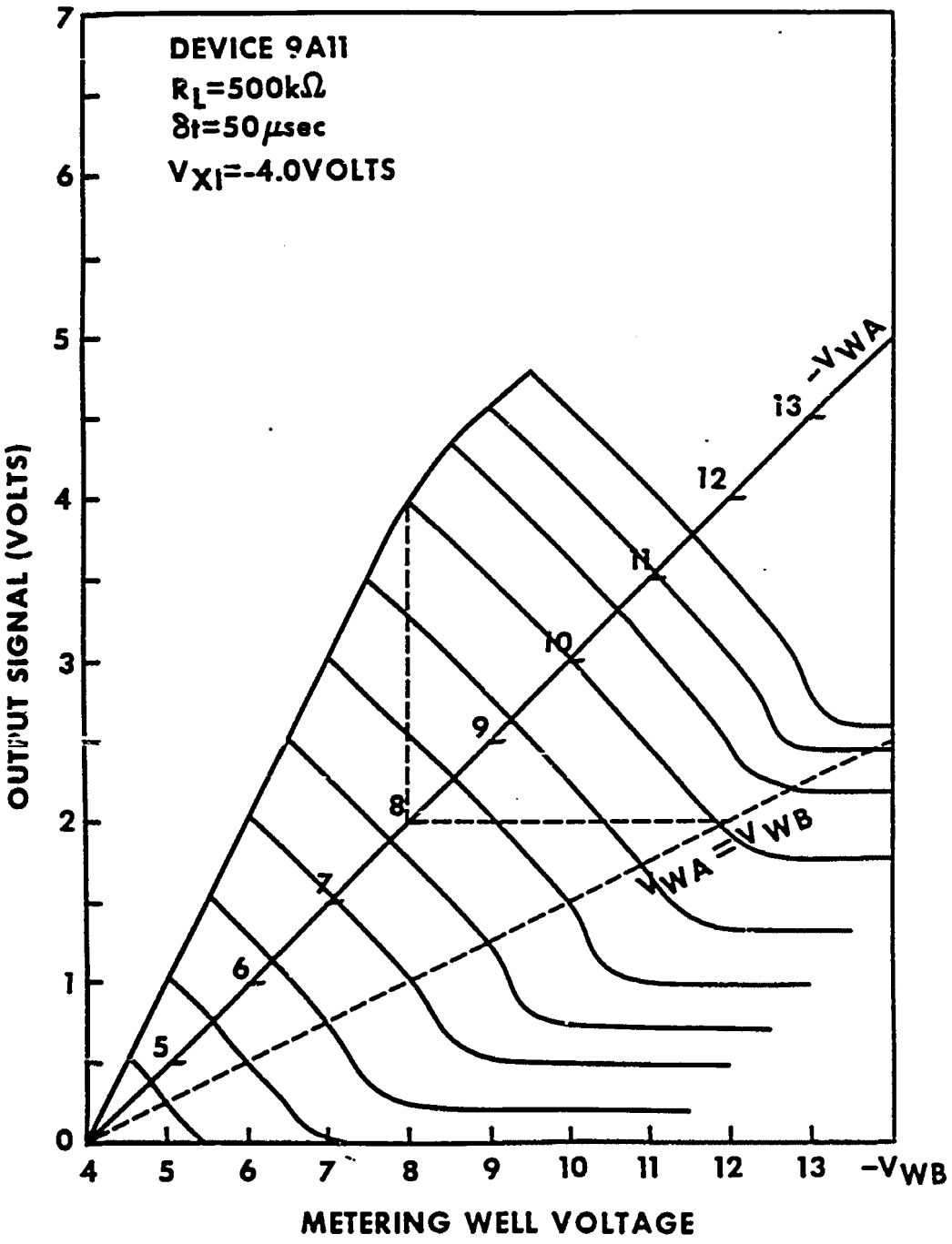
b. Output as a function of channel B input charge packet. V_{WB} ramped from -3.5 volts to -8.5 volts left to right. V_{WA} stepped from -8.0 volts to -5.0 volts in 0.5 volt increments beginning from maximum response.

(a) DEVICE 7B20 R_L=500kΩ



(b) DEVICE 7B20 R_L=500kΩ





5.4-7 Transfer characteristic surface of experimental charge packet differencer. Output signal referenced to reset level shifted by 0.6 volts.

photograph, V_{WA} (solid trace) is slowly increased from -3.5 volts to -8.5 volts from left to right while V_{WB} is kept at constant value. The output signal (which appears as a short dash in these photographs) increases as V_{WA} increases, and decreases as V_{WB} is stepped from -5.0 volts to -8.0 volts in 0.5 increments.

In the second photograph, V_{WB} (solid trace) is slowly increased from -3.5 volts to -8.5 volts from left to right as V_{WA} is kept constant. The output signal decreases with increasing V_{WB} and increases as V_{WA} is stepped from -5.0 volts to -8.0 volts. This second photo is quite similar to Fig. 5.4-5 except that the viewing angle is perpendicular to the V_{WB} axis.

In Fig. 5.4-7, the successful operation of part 9A11 is shown. As expected, this device shows many of the same features as device 7B20 (Fig. 5.4-5) and its behavior is not analyzed here. Since it was fabricated using a slightly different process from that employed during the manufacturing of device 7B20, its operating characteristics are expected to be somewhat different. This is partially exhibited by the lower gain observed in the transfer characteristic of part 9A11.

Linearity

The linearity of the transfer characteristic shown in Fig. 5.4-5 is quite good except close to the bias conditions $V_{WA} = V_{X1}$, $V_{WB} = V_{X1}$, and $V_{WA} = V_{WB}$, where the transfer characteristic is slightly rounded. This distortion is introduced by the fill and spill operation for shallow capture wells. For example, a similar distortion is observed in Fig. 5.4-3 where the MTOS experiments' parallel channel was used to create a charge packet using the fill and spill technique. Ideally, the fill and spill technique should result in a sharp slope discontinuity at $V_{MW} = V_{XI}$. Instead, the large geometry circuit implemented on the Yale 8205

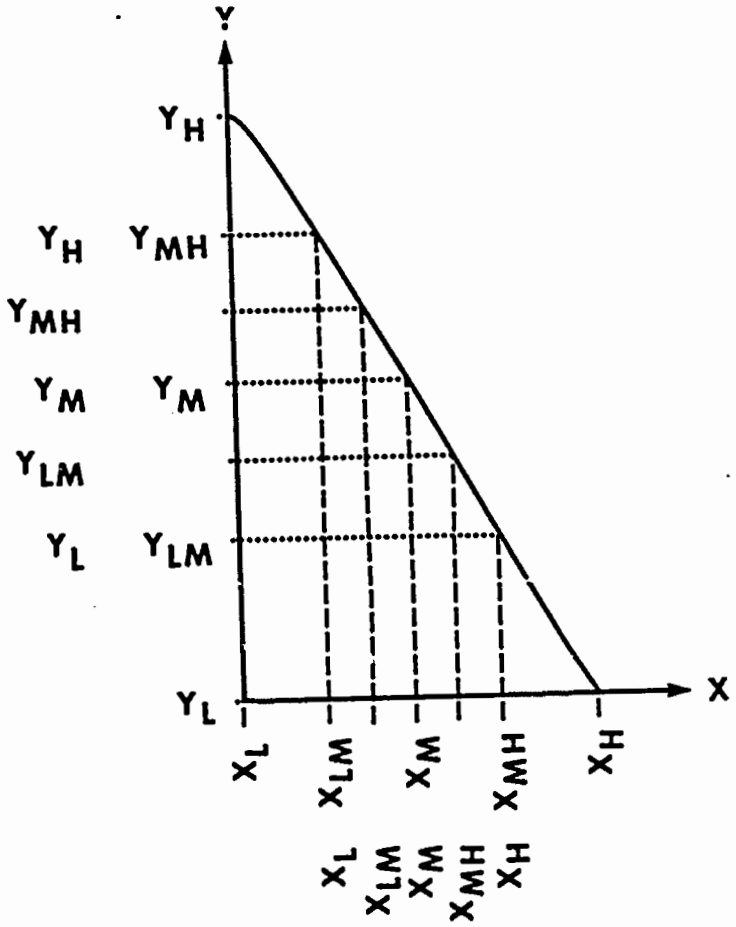
chip produces a rounded knee transfer characteristic. The rounding is probably due to charge transfer inefficiency during the spill phase. A small charge packet is effectively captured due to the slow spill process even though it is energetically favorable for that charge to be spilled out the input diode. Such rounding effects are expected to be reduced for smaller geometry devices.

Charge transfer inefficiency during the spill phase of the input circuit is responsible for the distortion near $V_{WB} = V_{XI}$ and $V_{WA} = V_{XI}$ in the 7B20 transfer characteristics. Charge transfer inefficiency during the spill phase of the differencer circuit is responsible for linearity distortion near $V_{WB} = V_{WA}$.

In characterizing the differencer linearity, four cases are considered. The linearity is measured over two different operating ranges. One is the full operating range which includes the rounding effects discussed above. The second is a reduced range which does not include the rounding effects. The linearity is thus measured for a constant value of V_{WA} over two operating ranges of V_{WB}, and for a constant value of V_{WB} over two operating ranges of V_{WA}.

To quantify the non-linearity, a method introduced by Epsley (1933) as described by Millman and Halkias (1972) is used. The technique is to partition the operating range into four equal sections in the input parameter (V_{WA} or V_{WB}). The four sections are defined by five metering well voltages which are labelled X_L, X_{LM}, X_M, X_{MH} and X_H in order from lowest to highest. Each input voltage is associated with an output voltage (Y_L, Y_{LM}, Y_M, Y_{MH}, Y_H). For the constant V_{WA} case (-8.0 volts), the partitioned transfer characteristic is displayed in Fig. 5.4-8. Partitioning for both the full interval and the sub interval is shown.

In the method of Espley, it is assumed that an input sine wave is



5.4-8 Partitioning of actual transfer characteristic for non-linearity characterization. See text.

applied to the circuit and a distorted sine wave is output. The distorted output signal is assumed to contain the five Fourier components B_0, B_1, B_2, B_3 and B_4 in order of increasing frequency. These five component amplitudes are determined from the transfer characteristic in terms of the five output voltages according to the following relationships:

$$B_0 = \frac{1}{6}\ (Y_H + 2Y_{MH} + 2Y_{LM} + Y_L) - Y_M \qquad 5.4\text{-}13a$$

$$B_1 = \frac{1}{3}\ (Y_H + Y_{MH} - Y_{LM} - Y_L) \qquad 5.4\text{-}13b$$

$$B_2 = \frac{1}{4}\ (Y_H - 2Y_M + Y_L) \qquad 5.4\text{-}13c$$

$$B_3 = \frac{1}{6}\ (Y_H - 2Y_{MH} + 2Y_{LM} - Y_L) \qquad 5.4\text{-}13d$$

$$B_4 = \frac{1}{12}\ (Y_H - 4Y_{MH} + 6Y_M - 4Y_{LM} + Y_L) \qquad 5.4\text{-}13e$$

The harmonic distortion is defined for each component as

$$D_2 = \left|B_2/B_1\right| \qquad 5.4\text{-}14a$$

$$D_3 = \left|B_3/B_1\right| \qquad 5.4\text{-}14b$$

$$D_4 = \left|B_4/B_1\right| \qquad 5.4\text{-}14c$$

The total distortion is defined as

$$D_T = (D_2^2 + D_3^2 + D_4^2)^{1/2} \qquad 5.4\text{-}15$$

For device 7B20, the four cases considered were

1. V_{WA} = -8.0 volts, V_{WB} = -4.0 to -8.0 volts
2. V_{WA} = -8.0 volts, V_{WB} = -5.0 to -7.0 volts
3. V_{WB} = -5.0 volts, V_{WA} = -4.0 to -8.0 volts
4. V_{WB} = -5.0 volts, V_{WA} = -6.0 to -8.0 volts

In case 1, distortion is expected due to the rounding effects discussed above. In case 3, the same distortion is expected with additional distortion due to saturation effects. The results of these

Table 5.4-2

Linearity Distortion

Case	1	2	3	4
Y_H	6.70	5.35	6.38	5.37
Y_{MH}	5.35	4.50	5.37	4.52
Y_M	3.63	3.63	3.68	3.68
Y_{LM}	1.85	2.75	1.93	2.81
Y_L	0.20	1.85	0.19	1.93
B_0	-0.08	-0.013	-0.I5	-0.02
B_1	3.33	1.75	3.21	1.71
B_2	0.09	0.015	-0.02	-0.015
B_3	0.08	0.000	-0.115	0.0033
B_4	-0.01	-0.0016	-0.05	0.005
D_2	0.027	0.0085	0.062	0.0087
D_2 (dB)	-31dB	-41dB	-24dB	-41dB
D_3	0.025	0.0000	0.036	0.0019
D_3 (dB)	-32dB	-	-29dB	-54dB
D_4	0.003	0.0009	0.015	0.0029
D_4 (dB)	-50dB	-61dB	-36dB	-51dB
D_T	0.037	0.0086	0.073	0.0094
D_T (dB)	-29dB	-41dB	-23dB	-41dB

measurements and calculations are displayed in Table 5.4-2. For the first case, the total distortion is -28 dB (20 $\log_{10} D_T$) and for the third case, the distortion level was increased to -23 dB. However, in the sub internal operating ranges, the distortion was much less and measured to be -41 dB for each case.

Noise and Dynamic Range

The noise of part 7B20 was examined by displaying the output signal during actual operation on an oscilloscope. Although such a method does not reveal the spectral properties of the noise nor does it provide an accurate determination of the RMS noise value, it does yield an upper bound on the magnitude of the noise. From such an examination, the noise envelope was found to have a maximum value of 1 millivolt in width, with a 60 Hz envelope of 2 mV.

The observed noise level is most likely due to the experimental set-up which was designed to reduce noise levels to the millivolt range. For measuring noise levels in the microvolt range in such a complex test station, special elaborate precautions must be taken which were not observed in this work since noise levels were not a fundamental issue. The noise was examined to place a lower bound on the dynamic range of the fabricated devices.

A millivolt noise signal corresponds to a 37 fC equivalent noise charge packet. The observed saturation level was 165 pC so that the dynamic range, which is the ratio of these charge packets is over 4400, corresponding to 73 dB for a signal to noise ratio (SNR) of unity. The practical dynamic range might be defined for an SNR of 10, or 53 dB.

To compare this figure to the range expected were test station noise removed, various noise sources are estimated following Carnes and Kosonocky (1972) and Barbe (1975). The noise is calculated in terms of

an RMS number of carriers and converted to output signal by multiplying by q/C_{OUT}. First, at the input, the noise introduced by the A and B input channels is given by

$$\bar{v}_{IN} \approx (kTC_{MW})^{1/2}/C_{OUT} \qquad 5.4\text{-}16$$

which is of the order of 10 μV each. The differencer itself also has a fill and spill noise of the same size. Dark current (which is zero-standoff signal) contributes a noise which is approximately equal to the square root of the zero-standoff voltage equivalent number of carriers (assuming a Poisson distributed generation rate). For part 7B20, which has a zero-standoff signal of 0.8 volts, this yields approximately 60 μV.

Interface trapping noise is given by

$$\bar{v}_{IT} = q(1.4kTD_{it}A_E)^{1/2}/C_{OUT} \qquad 5.4\text{-}17$$

where A_E is the total electrode area in the circuit and D_{it} is the active interface trap density. Assuming $D_{it} = 1 \times 10^{10}$ $(cm^2 - eV)^{-1}$, the noise is 5 μV. Another source of noise during charge transfer is charge transfer inefficiency noise. This noise is given by

$$\bar{v}_{CTI} = q\left[2\varepsilon n_T \bar{N}\right]^{1/2} / C_{OUT} \qquad 5.4\text{-}18$$

where ε is the fractional loss per transfer, n_T is the number of transfers, and $\bar{N}$ is the average number of carriers. Using a modest value of $\varepsilon = 0.05$ yields approximately 60 μV.

The output amplifier reset noise during the precharge phase is

$$\bar{v}_R = (kTC_{AG})^{1/2}/C_{OUT} \qquad 5.4\text{-}19$$

which is of the order of 10 μV. The Johnson noise from the large load resistance is

$$\bar{v}_J = (4kTR_L\Delta B)^{1/2} \qquad 5.4\text{-}20$$

The bandwidth is determined by the 25 pF oscilloscope input capacitance yielding a noise level of 25 μV.

The total noise expected for this CCD circuit is therefore

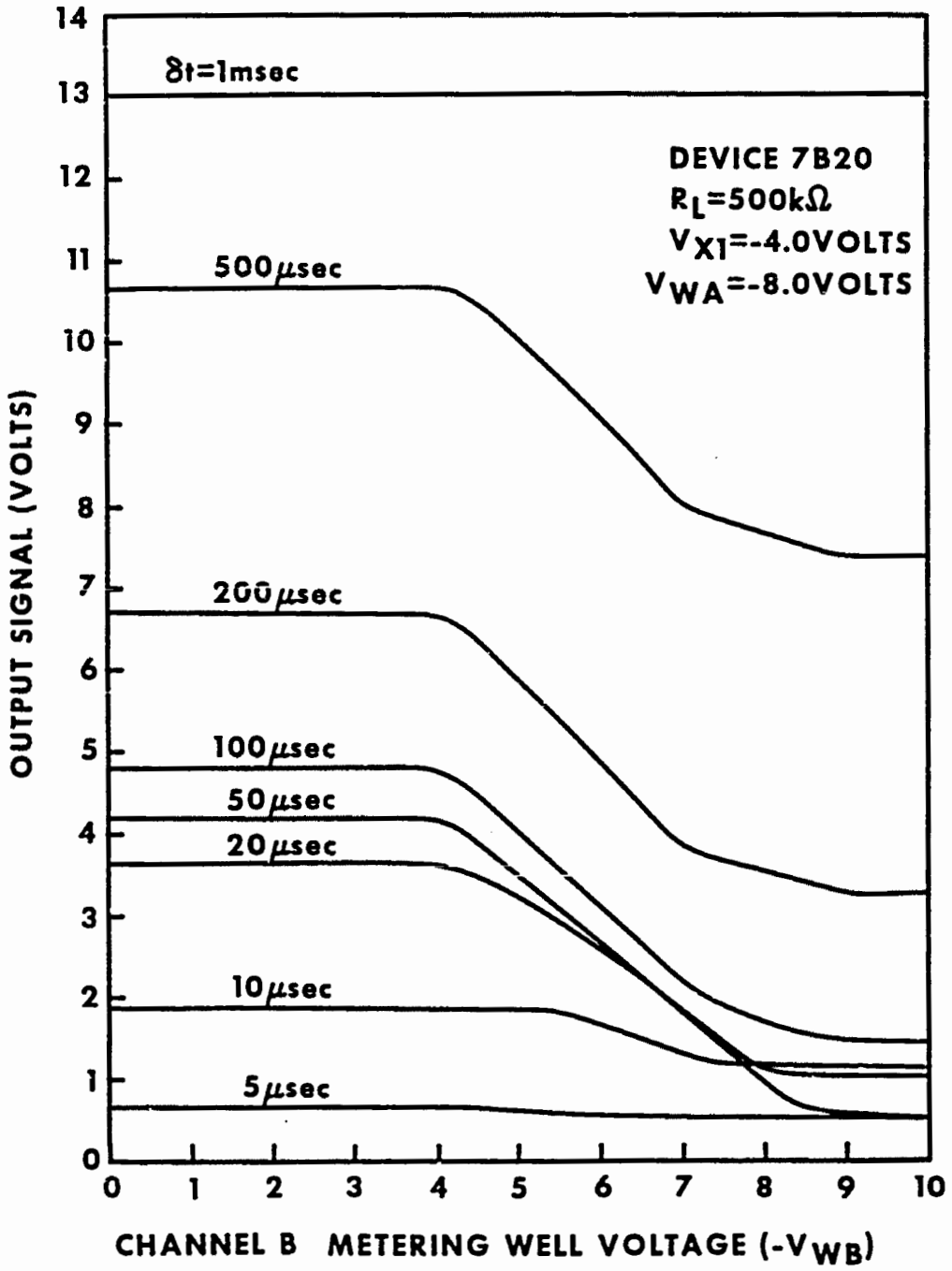
$$\bar{v}_T = (3\bar{v}_{IN}^2 + \bar{v}_{TH}^2 + \bar{v}_{IT}^2 + \bar{v}_{CTI}^2 + \bar{v}_R^2 + \bar{v}_J^2)^{1/2} \qquad 5.4\text{-}21$$

which is approximately 90 μV, or a noise equivalent charge packet of 3.4 fC. The projected maximum dynamic range for an SNR of 1 for this device is given by the ratio of the maximum saturation charge packet of 186 pC to the noise equivalent charge packet which is 5.5×10^4 or 81 dB.

Operating Speed

The effect of varying the fundamental clock period δt from 1000 μsec down to 5 μsec is shown in Fig. 5.4-9. In this experiment, the A channel input charge packet is kept constant (V_{WA} = -8.0 volts) and the B channel charge packet size is slowly ramped (V_{WB} = 0 to -10 volts). The input transfer gate V_{X1} was biased at -4.0 volts and the other d.c. and clocked biases were as in Table 5.4-1.

The effect of varying the operating speed can be divided into two major consequences. For slow speeds (fundamental clock periods longer than 100 μsec), the consequence is increased dark current and amplifier droop resulting in a significant zero-standoff signal. However, the slope of the transfer characteristic is not affected. For clock periods shorter than 50 μsec, charge transfer inefficiency degrades the gain of



5.4-9 Output of charge packet differencer circuit as a function of B channel input for various values of fundamental clock period δt.

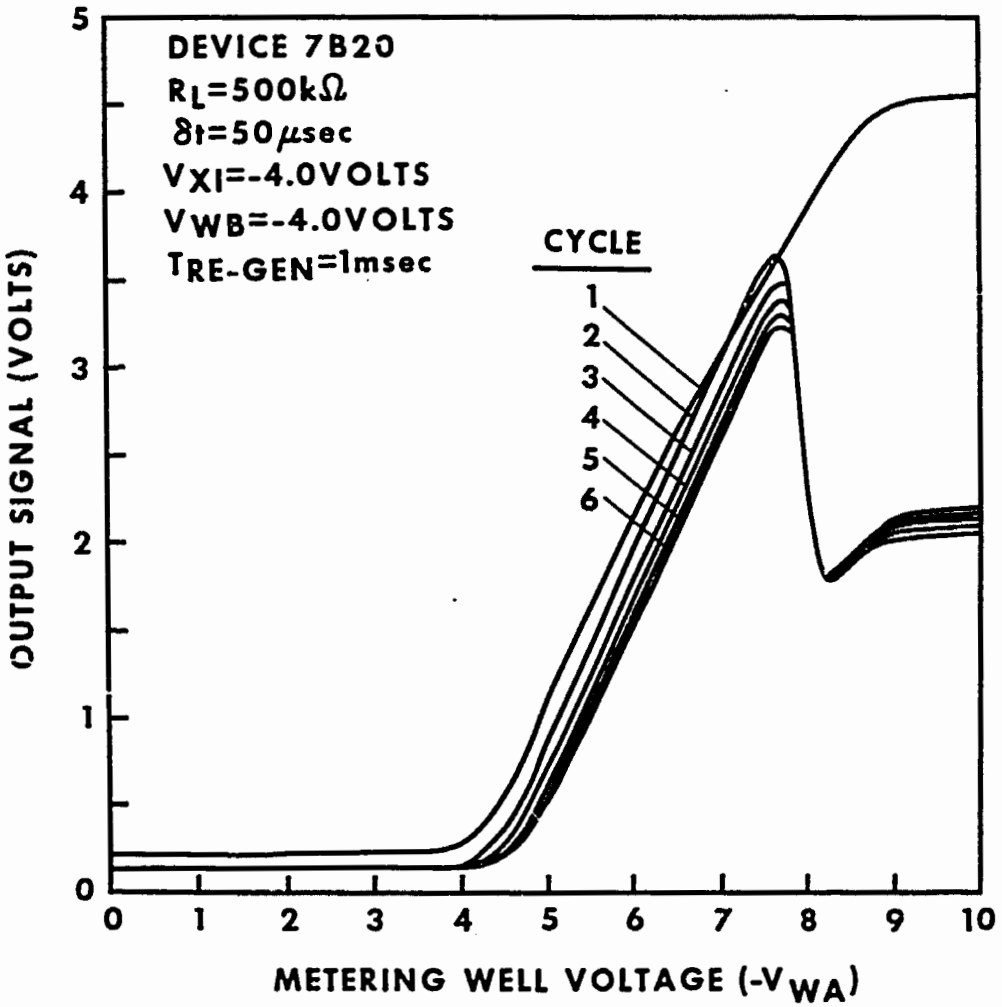
the transfer characteristic. At a clock period of 5 μsec, the transfer characteristic is nearly flat. Using a fundamental clock period of 50 μsec, it takes 300 μsec to complete one full differencing operation and to transfer the output charge packet into subsequent circuitry.

Multiple Cycle Operation

If the timing of the circuit is appended (see Fig. 5.2-2), it is possible to re-generate the output charge packet several times without repeating the gate charge subtraction operation. In this multiple cycle mode, a fill and spill operation is performed subsequent to the transfer of the output charge packet, and a copy of the output charge packet is re-generated.

To measure the effectiveness of this multiple cycle mode, the 7B20 device was biased for replicator mode operation with the re-generation timing cycles appended. The fundamental clock period was 62 μsec and the interval between each charge packet formation was one millisecond. During each cycle, the output amplifier was reset to its precharged state. The output signal is shown as a function of input charge packet size (V_{WA}) for six copy cycles in Fig. 5.4-10.

The overall behavior indicates that the differencer is indeed capable of re-generating output charge packets over multiple cycles. The slope of the transfer characteristic changes slightly from the initial cycle to the copy cycles, but remains nearly constant during the latter. The linearity appears to improve during the re-generation cycles, though the absolute size of the output signal progressively decreases. Most of these effects can be interpreted in terms of PN junction leakage current discussed in section 5.2, and exemplified by Eq. 5.2-27. During the replicator mode operation, the B electrode effective voltage may be significantly larger than the A electrode effective voltage, leading to



5.4-10 Output of charge packet differencer circuit as a function of A channel input for sequential regeneration cycles. Re-generation cycle time was one millisecond.

decreased output charge packets with increasing time.

When V_{WA} is larger than -7.0 volts, the effective A electrode voltage is small enough to result in charge transfer inefficiency during the fill phase. As a result, the output charge packet size dramatically drops. To explain why it does not drop to the zero-standoff voltage level is more troublesome, but a plausible explanation is that soft silicon breakdown in the B channel well is occuring, which tends to fill that potential well independently of the input diode. Such a soft breakdown hypothesis is consistent with the large zero-standoff charge packets observed during the output transfer phase.

5.5 Discussion

Accuracy Error

In chapter two the desired features of the charge packet differencer circuit were discussed. The key feature in an analog processing system is that accuracy be maintained. Within that context, it was proposed that an analog computational building block should have an eight bit equivalent accuracy, i.e., that the error be no larger than one part in 256, or -48 dB.

In the previous section, the fabricated devices were characterized in terms of linearity, gain, noise, and offset dynamic range. Each of these characteristics contributes to the total error accrued in one operation. For the fabricated devices, the equivalent observed noise in these circuits (0.04 pC) might suggest that a minimum resolvable charge packet (MRCP) size of 0.50 pC for an SNR of 12.5 be chosen. For a dynamic range of 256:1, the maximum charge packet which the differencer must need to handle is 128 pC, well below the experimentally observed saturation level of 165 pC.

The linearity distortion contributes to the accuracy error. For example, if the best fitting straight line is chosen through the five measured points of the first case listed in Table 5.4-2, the slope is 0.825. At the point where the actual transfer characteristic deviates the most from this best fitting line, the corresponding error is 6.5 pC. For the second case of Table 5.4-2, the slope is 0.88 and the maximum error in this subinterval corresponds to 0.68 pC. These non-linearity errors should be compared to the MRCP size of 0.50 pC. Thus, the sub-interval non-linearity error is almost tolerable in these large geometry devices.

The non-linearity error is calculated assuming that the slope itself is acceptable. However, non-unity gain caused by stray capacitance is the largest contributor to absolute computational error in the circuit. The maximum error occurs during the replication of a large charge packet. For example, with a stray capacitance attenuation factor of 0.89, replicating a 128 pC charge packet results in an error of 14 pC. To achieve a maximum error smaller than the MRCP, the stray capacitance attenuation factor must be better than 1-1/256 or 0.996. This implies that the stray capacitance must be less than 0.40 percent of the differencer oxide capacitance.

As observed at the output of the charge packet sense amplifier, the zero-standoff voltage also appears as a fixed large offset error. However, most of this error is generated in the course of transferring the charge packet out of the differencer and amplifying it. Therefore, it is not considered as an error in the differencer itself, even though it represents a nuisance.

The improved layout proposed in Fig. 5.3-3 will assist in decreasing the non-linearity error due to charge transfer inefficiency rounding

effects, and partially reduce stray capacitance, especially if the screen gate is omitted. However, even with the screen gate omitted, the stray capacitance will be two percent of the electrode oxide capacitance which is five times too large. The major effect of the improved layout will be to increase the speed of the circuit, which in turn will alleviate the detractive effects of the dark current and PN junction leakage current.

Operating Speed

As was just discussed, the proposed improved layout will increase the speed of the circuit from its present value of 300 μsec/operation (3.3 kops) by perhaps one or two orders of magnitude (∿ 0.1 mops). This is due to the present speed limitation being diffusion limited charge transfer which scales as L^2.

For a charge-coupled computer, the operating speed should be of the order of one million operations per second (1 mop). A major increase in speed comes about if the charge transfer in the channel is due to fringing fields rather diffusion. Such effects dominate in channel lengths under 10 μm which are not practically obtainable at the present time in the Yale facility.

Effects of Scaling and Process Improvement

It is expected that if a suitable processing facility were available, devices meeting the charge-coupled computer's requirements could be realistically manufactured. By decreasing the CCD device geometry to levels comparable to present day commercial devices (for example: W = 50 μm, L = 10 μm) the device speed could easily reach the 1 mops range. The area of such a scaled device is 200 times smaller than the improved layout of Fig. 5.4-3. Thus the noise level in the circuit becomes much more critical, even though the noise equivalent charge packet size

scales with the saturation level charge packet. However, the author's experience is that the charge-coupled computer's noise requirement (3000 RMS noise carriers) is comparable to present day CCD noise levels.

With the transfer efficiencies expected for such a small geometry device (better than 0.999), the linearity error associated with charge transfer inefficiency is expected to be minimal. The attenuating effects of stray capacitance can be improved by using a self-aligned polysilicon gate technology, though wiring capacitance may become more pronounced in system configurations. Any error though could be kept below the MRPC by careful layout and process design.

5.6 Summary and Conclusions

This chapter began with a discussion of the basic theory of operation of the charge packet differencer circuit. To the best knowledge of the author, and confirmed by a patent novelty search, the differencer circuit is unique in its operation and compactness, and offers distinct advantages over present charge packet differencing techniques. Aside from compactness, the advantages of the circuit are inherent linearity and communication over structures in the silicon substrate. The speed of this new circuit is projected to be comparable to present CCD differencing circuits with accuracy unsurpassed by devices occupying significantly more real estate.

In addition to the basic theory, non-ideal behavior was also discussed such as the attenuating effects of stray capacitance. Also covered was the detractive effects of PN junction leakage current and bucket dark current.

The layout of an experimental realization of the proposed differencer circuit which was fabricated at Yale was presented. The layout was critiqued using hindsight and an improved layout suitable for

fabrication without process alteration was described.

The experimental devices fabricated at Yale were measured and characterized. Although the relatively large geometries employed in the Yale process due to the limited facilities were a priori acknowledged to yield less than optimal performance, the fabricated devices functioned well. The major source of accuracy error was due to stray capacitance, though the effects of charge transfer inefficiency, dark current and PN junction leakage current also contributed to the total error. By applying the theory presented early in the chapter, the inaccuracies described above were quantitatively analyzed and found to conform to expectations in a self-consistent way.

Finally, the actual device performance and projected performance in consideration of scaling and process improvement were compared to the requirements of the charge-coupled computer as discussed in chapter two. The experimental devices have well demonstrated the feasibility of performing charge packet differencing using the proposed CCD circuit.

CHAPTER 6

SUMMARY AND CONCLUSION

There were several advances in engineering and applied science which were made in the course of this work. These can be divided into advances of interest to the scientific community at large, and those which have progressed the state of knowledge in the Yale semiconductor research group. For the former community, the advances are considered to be (in no particular order):

First, the novel concept of a charge-coupled computer was introduced, which shows promise for focal plane array image pre-processing applications. This concept is a technological solution to the fundamental problem of machine vision.

Second, a new charge-coupled device circuit which performs charge packet differencing, replication and/or attenuation in an inherently linear and compact way was invented. A prototype circuit was fabricated and analyzed. This circuit forms the basic analog building block of the charge-coupled computer and will be useful in other more specialized systems. This invention represents an advance in solid-state circuits.

Third, although not heavily emphasized, but of some import to the semiconductor industry, was the successful demonstration of a 33 Å gate oxide enhancement mode MOSFET with well behaved characteristics. Previous ultra-thin oxide MOSFETs have shown irregularity in their performance, which has been eliminated in this work through process improvement. Perhaps to be significant was the observation of mobility degradation in ultra-thin oxide MOSFETs, though a systematic study of this phenomenon was not a part of this research plan.

Fourth, the MTOS junction was experimentally demonstrated to behave as a charge packet threshold detector albeit a slow one. This bistable device was found to be triggered into its high current state if an injected charge packet exceeded a certain critical level. The critical level and transient response of the MTOS junction was measured and analyzed.

Fifth, the incorporation of electronic control circuitry on the same chip as the MTOS junction in an integrated fashion has permitted the measurement of internal MTOS junction currents over an extended and previously inaccessible range of state space using both a charge-coupled mode and a surface potential control mode. The magnitudes of topical physical phenomena such as oxide hole transport and hot electron impact ionization have been ascertained.

Sixth, a new N-type negative resistance was discovered and explained. This negative resistance arises in the coupling of an MTOS junction with a PN junction through an appropriately biased field effect control channel at the silicon surface. The strength of the negative resistance depends upon the MTOS junction bias and coupling gate bias, which control the back diffusion current from a super-inversion layer of minority carriers under the tunnel oxide electrode. This negative resistance may have application in optical or electronic switching devices.

Parochial progress of interest to the Yale semiconductor effort was also made in the course of the research. Such progress is probably not of particular interest to the broader community at large because these solutions and observations are either minor or have been established in industry (undoubtedly several times) but never publicly disclosed or have not yet been adapted to the Yale environment. In the process area, for example, the shadowed evaporation technology for forming sub-micron inter-electrode gaps was successfully transferred from the literature to practice. Dark current was suppressed several orders of magnitude by lowering the wet oxidation temperature. The use of phosphorous doped silicon to getter impurities was successfully employed in the Yale 8205 process. Borofilm was found to yield much higher quality P^+N junctions than borosilicafilm. Post gate oxidation forming gas anneal was found to reduce interface traps. The avoidance of ultrasonic agitation following tunnel oxidation was found to significantly improve MTOS junction durability. Sub-micron spaced tunnel oxide electrodes can be sequentially formed without significant difference between the two junctions' properties. A double etch process was found to yield good

contact vias without compromising tunnel oxide quality. Chips with MTOS devices can be scribed and ultrasonically bonded with 100% device yield.

Progress which was not process related included the modelling of deep depleted MOS capacitor carrier concentrations through the use of a simplifying zero current approximation and an appropriate computer program. Such a model will assist in relating surface potential, oxide voltage, surface carrier concentrations, electric field, depletion depth, recombination/generation currents, bulk diffusion current and device capacitance. Progress was also made in measurement instrumentation as described briefly in appendix E.

Philosophically, perhaps the most significant impact this research has had at Yale is to inject the concept of charge-coupled devices applied to fundamental ultra-thin oxide device studies into the local stream of consciousness.

Several research spin-offs were suggested by preliminary results obtained in this research. The first is related to oxidation induced stacking faults. The electrical activity associated with surface silicon defects and their elastic strain fields such as excess minority carrier recombination or generation rates does not appear to be as well studied as more classic interface traps. A better understanding of these effects may be important to dynamic memory technology and CCDs. Previously, it appears that only the bulk effects of silicon defects have been considered [for example, Chatterjee, et. al. 1979].

A second research spin-off already underway is based on the observation of mobility degradation in ultra-thin oxide MOSFETs. This study has been previously limited by an inability to make high quality ultra-thin gate oxides. For example, the effect of ultrasonic agitation during photolithography had not before been related to annealing

durability. Nor had low impedance source-drain contact vias been formed.

A third as yet unexplored research project is the proximity coupling of tunnel oxide devices which are laterally spaced in the sub-micron range. Prior to this work, a method for making these devices was unavailable (in a local sense). Investigation of such devices may yield results of fundamental or practical interest. For example, proximity coupled MINP bistable junctions have been employed to form a shift register [Kawamura and Yamamoto, 1981].

A fourth research spin-off is actually a continued study of the charge control/surface potential control structures coupled to capacitor structures. However, the tunnel oxide might be replaced by a thicker Fowler-Nordheim tunnel oxide [as suggested by D. DiMaria, private communication] or by other dielectric thin films such as silicon rich SiO_2 or nitrided oxide. This study is also presently underway.

Finally, the study of charge-coupled computing is really only in its infancy. This work has demonstrated the concept's feasibility, but there remains many avenues for future investigation.

In conclusion, this research has spanned a spectrum from computer science and parallel processing through electrical engineering and integrated solid state circuit design to the physics of hot electron impact ionization. Perhaps overly ambitious in its scope, the research has nevertheless permitted an insightful glimpse into the myriad of interrelationships among the scientific disciplines, and perhaps has yielded advances on several fronts.

APPENDIX A

LIST OF SYMBOLS

Differencer gain	α
Output amplifier gain	α_A
Prototype differencer overall gain (A channel)	α_{DA}
Prototype differencer overall gain (B channel)	α_{DB}
Stray capacitance attenuation factor	α_{SC}
Reciprocal thermal voltage	β
Delta function	$\delta(x)$
Permittivity of free space	ε_o
Relative dielectric constant of SiO_2	ε_{ox}
Minority carrier mobility	μ_p
Surface minority carrier effective mobility	μ_p'
Photon frequency	ν
Charge density	ρ
Oxide charge density	ρ_{ox}
Capture cross-section	σ
Interface trap capture cross-section	σ_{it}
Electron trap capture cross-section	σ_n
Hole trap capture cross-section	σ_p
Minority carrier lifetime	τ_p
Average minority carrier lifetime	τ_p
Aluminum - SiO_2 barrier voltage	ϕ_{Al}
Metal - SiO_2 barrier voltage	ϕ_m
Bulk silicon built-in voltage	ϕ_n
Body potential	ϕ_o

Semiconductor - SiO_2 barrier voltage	ϕ_s
Silicon - SiO_2 barrier voltage	ϕ_{Si}
Electrostatic potential	$\psi(x)$
Surface potential	ψ_S
Inversion layer voltage drop	$\delta\psi$
Metering well surface potential	ψ_{SMW}
Input transfer gate surface potential	ψ_{SXI}
Diode-controlled surface potential offset	ψ_S^o
Kane band structure constant	A
Device area	A
Metering well area	A_{MW}
Tunnel oxide area	A_{TO}
Output transfer gate area	A_{XO}
Capacitance	C
Output amplifier floating capacitance	C_{AG}
Floating diffusion capacitance	C_{FD}
High frequency capacitance	C_{HF}
Low frequency capacitance	C_{LF}
Output transcapacitance	C_{OUT}
Oxide capacitance	C_{OX}
Stray capacitance	C_{ST}
Total capacitance	C_T
Interface trap density	D_{it}
Aluminum thickness	d
Oxide thickness	d_{ox}
Minority carrier diffusion coefficient	D_p
Total distortion	D_T

Average tunneling barrier height	$\bar{E}_B$
Semiconductor conduction band edge energy	E_C
Planck electron energy	$\hat{E}$
Metal Fermi energy	E_{FM}
Semiconductor Fermi energy	E_{FS}
Semiconductor band gap energy	E_G
Semiconductor midgap energy	E_i
Impact ionization threshold energy	E_{ii}
Metal Fermi energy	E_m
Optical phonon energy	E_{ph}
Semiconductor valence band edge energy	E_V
Electric field	$\hat{\mathcal{E}}(x)$
Semiconductor breakdown electric field	$\hat{\mathcal{E}}_{BR}$
Oxide electric field	$\hat{\mathcal{E}}_{OX}$
Surface electric field	$\hat{\mathcal{E}}_S$
Generation rate (external)	G_{EXT}
Aluminum etching undercut hypotenuse	h
Planck's constant	h
Planck's constant divided by 2π	$\hbar$
Back diffusion current	I_{diff}
Electron oxide tunneling current fitted parameter	I_{et}^{o}
Electron oxide tunneling current	I_{et}
Hole oxide transport current fitted parameter	I_{ht}^{o}
Hole oxide transport current	I_{ht}
Impact ionization current fitted parameter	I_{ii}^{o}
Impact ionization current	I_{ii}
Lateral injection/extraction current	I_{lat}

PN junction reverse leakage current	I_{leak}
Optical generation current	I_{op}
Recombination current	I_{rec}
Thermal generation current	I_{th}
MOSFET saturation current	I_s
Bulk diffusion dark current density	J_B
Depletion region dark current density	J_D
Back diffusion current density	J_{diff}
Electron oxide tunneling current density	J_{et}
Hole oxide transport current density	J_{ht}
Impact ionization current density	J_{ii}
Lateral injection/extraction current density	J_{lat}
PN junction leakage current density (A channel)	J_{LA}
PN junction leakage current density (B channel)	J_{LB}
Current density fitted parameter	J_0
Optical generation current density	J_{op}
Electron current density vector	$\vec{J}_n$
Hole current density vector	$\vec{J}_p$
Recombination current density	J_{rec}
Surface dark current	J_S
Thermal generation current density	J_{th}
Boltzmann's constant	k
Electron wave vector	$\vec{k}$
Impact ionization scattering length	l_{ii}
Phonon scattering length	l_{ph}
Channel length	L
Metering well length	L_{MW}

Input transfer gate length	L_{XI}
Ratio of scattering lengths	m
Electron effective mass	m_e^*
Hole effective mass	m_h^*
Longitudinal electron effective mass	m_l^*
Light hole effective mass	m_{lh}^*
Transverse electron effective mass	m_t^*
Electron concentration	$n(x)$
Intrinsic carrier concentration	n_i
Bulk electron concentration	n_{no}
Trap density (space and energy)	n_t
Conduction band effective density of states	N_c
Doping concentration	N_D
Trap density (space)	N_t
Total number carriers emitted from traps	$N(t)$
Hole concentration	$p(x)$
Bulk hole concentration	p_{no}
Surface hole concentration	p_s
Total impact ionization probability	P_{ii}
Incremental impact ionization probability	P_n
Barrier penetration probability	P_T
Electron charge	q
Charge on the A channel electrode	Q_A
Charge on the B channel electrode	Q_B
Channel charge	Q_C
Depletion layer charge	Q_D
Depletion layer charge under electrode A	Q_{DA}

Depletion layer charge under electrode B	Q_{DB}
Equilibrium state inversion layer charge	Q_{EQ}
Electron tunneling current fitted parameter	Q_{et}
Hole transport current fitted parameter	Q_{ht}
Impact ionization current fitted parameter	Q_{ii}
Input charge packet	Q_{IN}
Inversion layer charge	Q_{INV}
Rate of change of inversion layer charge	$\dot{Q}_{INV}$
Average inversion layer charge	$\bar{Q}_{INV}$
MTOS OFF state inversion layer charge	Q_{INV}
MTOS ON state inversion layer charge	Q_{INV}
Output charge packet	Q_O
Current fitted parameter	Q_0
Total semiconductor space charge	Q_S
A channel input charge packet	Q_{SA}
B channel input charge packet	Q_{SB}
A channel stray capacitance charge	Q_{SCA}
B channel stray capacitance charge	Q_{SCB}
Signal charge packet	Q_{SIG}
Threshold level charge packet	Q_T
Contact via resistance	r_C
Diffused layer sheet resistance	r_D
Impact ionization scattering rate	r_{ii}
Phonon scattering rate	r_{ph}
Contact string test site resistance	R_C
Diffused resistor value	R_D
Load resistance	R_L

Surface recombination velocity	S_o
Fundamental clock period	δt
MOS capacitor relaxation time	Δt
Self-induced drift time constant	t_o
Aluminum etch time	t_d
Metering well fill time	t_F
MTOS switching OFF time	t^{OFF}
MTOS switching ON time	t^{ON}
Fringing field transit time	t_{TR}
Total aluminum overetch time	t_w
Temperature	T
Robotic reaction time	T_R
Shockley-Read-Hall net recombination rate	U
Charge transfer inefficiency noise voltage	v_{CTI}
Input channel noise voltage	v_{IN}
Interface trap noise voltage	v_{IT}
Johnson noise voltage	v_J
Reset noise voltage	v_R
Total CCD noise voltage	v_T
Carrier thermal velocity	v_{th}
Differencer reset voltage	V_0
Amplifier reset voltage	V_{0A}
Diode or drain voltage	V_D
Input diode voltage	V_{D1}
Input diode voltage	V_{D2}
Amplifier drain voltage	V_{DA}
Flat band voltage	V_{FB}

Gate voltage	V_G
Holding voltage	V_H
Amplifier output voltage	V_{OUT}
Metering well voltage	V_{MW}
Oxide voltage drop	V_{OX}
Breakdown oxide voltage drop	V_{OX}^{BR}
Critical oxide voltage drop	V_{OX}^{CR}
Amplifier pre-charge switch voltage	V_{PCA}
Differencer pre-charge switch voltage	V_{PCS}
Source voltage	V_S
Amplifier source voltage	V_{SA}
Screen gate voltage	V_{SCR}
Threshold voltage	V_T
Effective threshold voltage	V_T'
Tunnel oxide voltage	V_{TO}
A input well voltage	V_{WA}
B input well voltage	V_{WB}
A channel transfer voltage	V_{XA}
B channel transfer voltage	V_{XB}
Joint channel transfer voltage	V_{XAB}
Input transfer gate voltage	V_{XI}
Output transfer gate voltage	V_{XO}
Aluminum etch undercut distance	w
Channel width	W
Metal - semiconductor work function difference	W_{MS}
Depletion layer depth	x_d

ABBREVIATIONS

Charge-coupled device	CCD
Charge transfer device	CTD
Charge transfer efficiency	CTE
Czochralski grown silicon	CZ
Dual in-line package	DIP
Float zone grown silicon	FZ
Metal-nitride-oxide-semiconductor	MNOS
Metal-oxide-semiconductor	MOS
MOS field effect transistor	MOSFET
Metal-tunnel oxide-semiconductor	MTOS
Minimum resolvable charge packet	MRCP
Processing element	PE
Post metallization forming gas anneal	PMFGA
Post metallization nitrogen anneal	PMNA
P-type semiconductor/N-type semiconductor	PN
Post oxidation forming gas anneal	POFGA
Radio frequency	RF
Root-mean-square	RMS
Signal to noise ratio	SNR

APPENDIX B

CARRIER CONCENTRATIONS IN NON-EQUILIBRIUM SEMICONDUCTOR SPACE CHARGE REGIONS USING A ZERO CURRENT APPROXIMATION

Introduction

In this appendix, a zero-current approximation will be developed to determine minority and majority carrier concentrations in a non-equilibrium space charge region. In the theses of Lai (1979) and Teng (1983), an approach is used which comes from the work of Green (1974). In that approach, a clever integral of the current density equations is performed to obtain the surface minority carrier concentration. However, in the course of performing the integral, the current contribution is found to be negligible, and is then set to zero.

In the zero-current approximation introduced in this appendix, the current contribution is initially set to zero. The end results are, of course, similar and perhaps the mathematics are not significantly simplified. However, the technique applied in this appendix is thought to be conceptually straight forward, and parallels the approach taken by Sze (1981) for the equilibrium MOS structure. As such, the results obtained are equally valid in deep depletion, equilibrium, super-inversion and even accumulation.

Paradoxically, the zero-current approximation results are useful in calculating currents in the space-charge region and at its edge. Indeed, the zero-current approximation should not satisfy the casual reader until the full calculation has been performed, and the zero-current approximation validated.

Analysis

In this approach, one begins by writing the current density for holes as

$$J_p = q\mu_p p\mathcal{E} - qD_p \frac{dp}{dx} \qquad \text{B-1}$$

Rearranging, and dividing by qD_p yields

$$\frac{-dp}{dx} = \frac{J_p}{qD_p} - \frac{\mu_p}{D_p} p\mathcal{E} \qquad \text{B-2}$$

In most cases of interest (as will be validated later)

$$\frac{J_p}{q\mu_p p\mathcal{E}} << 1 \qquad \text{B-3}$$

This states that the net hole current is much smaller than the drift (or diffusion) component. Therefore, one can justifiably ignore the net current. This is not saying the net current is negligible, rather, it says that a sizeable current can be accounted for by a slight perturbation of the carrier concentration.

Equation B-2 is rewritten as

$$\frac{d \ln p}{dx} = -\beta \frac{d\psi}{dx} \qquad \text{B-4}$$

where Eq. B-2 has been divided by p, the Einstein relation employed, and the electric field $\mathcal{E}$ rewritten in terms of the electrostatic potential gradient. Integrating Eq. B-4 yields

$$\ln p(x) - \ln p(x_o) = -\beta \left[\psi(x) - \psi(x_o)\right]$$

$$\text{or} \quad p(x) = p(x_o) \exp \left[\beta\psi(x_o) - \beta\psi(x)\right] \qquad \text{B-5a}$$

This familar equation relates a change in hole concentration to a change in electrostatic potential. Similarly, for electrons one obtains

$$n(x) = n(x_o) \exp \left[\beta\psi(x) - \beta\psi(x_o)\right] \qquad \text{B-5b}$$

Attention is now confined to an MOS system on n-type silicon, where x is measured from the SiO_2-Si interface into the bulk. In this system, the electron concentration can be written in terms of the bulk concentration n_{no}

$$n(x) = n_{no} \exp (\beta \psi(x)) \qquad \text{B-6a}$$

where $\psi \to 0$ in the bulk.

The hole concentration, on the other hand, may be written in terms of the surface (x = 0) hole concentration p_s. The reason is that the zero-current approximation fails at the depletion region edge where $\mathcal{E} \to 0$ and p(x) are very small in deep depletion. However, the zero-current approximation is well suited for the inversion region where p_s and ψ_s are large. In this case,

$$p(x) = p_s \exp [\beta(\psi_s - \psi(x))]$$

Following Sze (1981, pg. 366-368), the total semiconductor space-charge Q_s can be calculated in terms of the surface potential ψ_s and the surface hole concentration p_s. The Poisson equation is rewritten as

$$\frac{d^2\psi(x)}{dx^2} = \frac{-\rho(x)}{\epsilon_s \epsilon_o} \qquad \text{B-7}$$

where $$\rho(x) = q\left[p(x) - n(x) - p_{no} + n_{no}\right] \qquad \text{B-8}$$

and where constant doping has been assumed.

Thus,

$$\frac{d^2\psi}{dx^2} = \frac{-q}{\epsilon_s \epsilon_o}\left[p_s e^{\beta(\psi_s - \psi(x))} - n_{no} e^{\beta\psi(x)} - p_{no} + n_{no}\right] \qquad \text{B-9}$$

Although, the zero-current approximation for p(x) is poor, near $\psi \approx 0$, Eq. B-9 is not critically dependent upon p(x), since the second term is strongly dominant.

To calculate the electric field $\mathcal{E}(x)$, one integrates the Poisson equation using the clever technique referenced by Sze. The technique is to first write

$$\mathcal{E}(x)^2 = \int_{\hat{\mathcal{E}}=0}^{\hat{\mathcal{E}}=\mathcal{E}(x)} 2\hat{\mathcal{E}}\, d\hat{\mathcal{E}} = \int_{\hat{\mathcal{E}}=0}^{\hat{\mathcal{E}}=\mathcal{E}(x)} 2\left(\frac{d\psi}{dx}\right) d\left(\frac{d\psi}{dx}\right)$$

Then, write

$$\frac{d\psi}{dx} \cdot d\left(\frac{d\psi}{dx}\right) = d\psi \cdot \frac{d}{dx}\left(\frac{d\psi}{dx}\right) = \frac{d^2\psi}{dx^2}\, d\psi$$

Thus

$$\mathcal{E}(x)^2 = \int_{\hat{\psi}=0}^{\hat{\psi}=\psi(x)} \frac{2q}{\varepsilon_s \varepsilon_o} \left[p_s e^{\beta(\psi_s - \hat{\psi})} - n_{no} e^{\beta\hat{\psi}} - p_{no} + n_{no} \right] d\hat{\psi} \qquad \text{B-10}$$

Performing the integral yields

$$\mathcal{E}(x)^2 = \frac{2q}{\varepsilon_s \varepsilon_o \beta} \left[n_{no} \left(e^{\beta\psi(x) - \beta\psi(x) - 1} \right) + p_s \left(e^{\beta\psi_s - \beta\psi(x)} + \frac{p_{no}}{p_s} \beta\psi(x) - e^{\beta\psi_s} \right) \right] \qquad \text{B-11}$$

This equation is the workhorse for all subsequent analysis and modeling.

The total semiconductor charge is calculated from Gauss' Law as

$$Q(x) = \varepsilon_s \varepsilon_o \mathcal{E}(x) \qquad \text{B-12}$$

so that at the surface,

$$Q_S^2 = \frac{2q\varepsilon_s \varepsilon_o}{\beta} \left[n_{no}(e^{\beta\psi_s} - \beta\psi_s - 1) + p_s(1 + \frac{p_{no}}{p_s}\beta\psi_s - e^{\beta\psi_s}) \right] \qquad \text{B-13}$$

This expression relates the total semiconductor space charge to the surface potential and surface hole concentration.

For an MOS system, the applied bias is divided between a voltage drop across the oxide (V_{OX}) and the drop across the semiconductor (ψ_s) according to

$$V_G - V_{FB} = V_{OX} + \psi_s \qquad \text{B-14}$$

where V_{FB} is the flat-band voltage. The voltage drop across the oxide

is given by

$$V_{OX} = -Q_S/C_{OX} \qquad \text{B-15}$$

where the oxide capacitance C_{OX} is defined in terms of the oxide thickness d_{ox} and relative dielectric constant ε_{ox} as

$$C_{ox} \triangleq \frac{\varepsilon_{ox}\varepsilon_o}{d_{ox}} \qquad \text{B-16}$$

Thus, from Eq. B-13,

$$V_G - V_{FB} = \psi_s \pm \sqrt{\frac{2q\varepsilon_s\varepsilon_o}{\beta C_{ox}^2}\left[n_{no}(e^{\beta\psi_s} - \beta\psi_s - 1) + p_s(1 + \frac{p_{no}}{p_s}\beta\psi_s - e^{\beta\psi_s})\right]} \qquad \text{B-17}$$

For an n-type semiconductor, one chooses (+) for accumulation and (-) for depletion. (In depletion, $V_G < \psi_s < 0$).

Note that in equilibrium (i.e. $pn = n_i^2$), $p(x)\Big|_{\psi=0} = p_{no}$ so that Eq. B-17 becomes

$$V_G - V_{FB} = \psi_s \pm \sqrt{\frac{2q\varepsilon_s\varepsilon_o}{\beta C_{ox}^2}\left[n_{no}(e^{\beta\psi_s} - \beta\psi_s - 1) + p_{no}(e^{-\beta\psi_s} + \beta\psi_s - 1)\right]} \qquad \text{B-18}$$

In deep depletion or inversion, the expression in Eq. B-17 can be reduced by noting that

$$\beta\psi_s << -1 \qquad \text{and} \qquad e^{\beta\psi_s} << 1$$

so that it becomes

$$V_G - V_{FB} = \psi_s - \sqrt{\frac{2q\varepsilon_s\varepsilon_o}{\beta C_{ox}^2}\,(p_s - N_D\beta\psi_s)} \qquad \text{B-19}$$

The semiconductor space charge can be decomposed into two components, the depletion layer charge of ionized donors Q_D, and an inversion layer of minority carriers Q_{INV}, such that

$$Q_S = Q_D + Q_{INV} \qquad \text{B-20}$$

In the case of deep depletion, $Q_{INV} << Q_D$, and Eq. B-19 reduces to

$$V_G - V_{FB} = \psi_s - (-\phi_o\psi_s)^{1/2} \quad ; \quad \phi_o^2 = 2q\varepsilon_s\varepsilon_o N_D/C_{ox}^2 \qquad \text{B-21}$$

This is, in essence, <u>the</u> depletion approximation, since precisely the same result is obtained by assuming a constant net charge density from the surface to a depletion depth x_d, and a zero charge beyond. That equivalent depletion layer depth x_d is given by

$$x_d = \left(\frac{2\varepsilon_s\varepsilon_o}{qN_D}\,\psi_s\right)^{1/2} \qquad \text{B-22}$$

In the case of strong or super-inversion, $Q_{INV} >> Q_D$, and Eq. B-19 is

reduced to

$$V_G - V_{FB} = \psi_s - (2\varepsilon_s \varepsilon_o kTp_s)^{1/2} / C_{ox} \qquad \text{B-23}$$

Here, the inversion layer Q_{INV} can be seen to be

$$Q_{INV} = (2\varepsilon_s \varepsilon_o kTp_s)^{1/2} \qquad \text{B-24}$$

This relates the inversion layer charge to the surface hole concentration.

Although approximate, the expression in Eq. B-24 can be substituted into Eq. B-19, and then the surface potential solved for. The result is

$$\psi_s = V_G - V_{FB} - \phi_o + \sqrt{\phi_o^2 - 2\phi_o(V_G - V_{FB}) + \left(\frac{Q_{INV}}{C_{ox}}\right)^2} \qquad \text{B-25}$$

In contrast, using a sheet charge inversion layer approximation, where ψ_s arises from the depletion layer charge only, one obtains

$$\psi_s = V_G - V_{FB} - \phi_o + \sqrt{\phi_o^2 - 2\phi_o(V_G - V_{FB}) - 2\phi_o \frac{Q_{INV}}{C_{ox}}} + \frac{Q_{INV}}{C_{ox}} \qquad \text{B-26}$$

The difference is due to the accounting of the effect of the inversion layer upon ψ_s. However, for MTOS devices, the numerical difference in calculating ψ_s given V_G and Q_{INV} is negligible, since ϕ_o is just a few millivolts (~3 mV), and to a good approximation

$$-V_{OX} = Q_{INV}/C_{OX} \qquad \text{B-27}$$

The problem with the voltage drop across the inversion layer is that it makes it difficult to relate the surface potential to the small-signal (high frequency) capacitance of the junction. The capacitance is given by

$$C = \left[\frac{d_{ox}}{\varepsilon_{ox}\varepsilon_o} + \frac{x_d}{\varepsilon_s\varepsilon_o}\right]^{-1} \qquad \text{B-28}$$

Here, x_d is a weighted depletion layer depth, which for convenience, can be defined as the depth where $n(x) = N_D/2$. The depletion approximation relates x_d and ψ_s according to Eq. B-22. However, in an MTOS system where Q_{INV} is typically $1 \times 10^{-6} C/cm^2$ or larger, this equation yields erroneous results.

Computer Model

The most effective way to calculate the depletion layer depth is to employ a simple computer model of the semiconductor space charge region. In this model, the surface potential as a function of space-charge is converted to surface potential as a function of depth. Once determined, minority and majority carrier concentration profiles may be calculated and the depletion layer depth determined.

The model begins by setting the gate voltage V_G and the oxide voltage drop V_{OX}. Using Eq. B-14, the surface potential ψ_s is determined. The electric field at the surface is calculated from Eq. B-11.

The model now increments the electrostatic potential ψ by amount $\delta\psi$, typically equal to $(10/\beta)^{-1}$. The corresponding incremental value

of depth δx can be obtained by writing

$$\delta x = \delta\psi / \mathcal{E}(x)$$

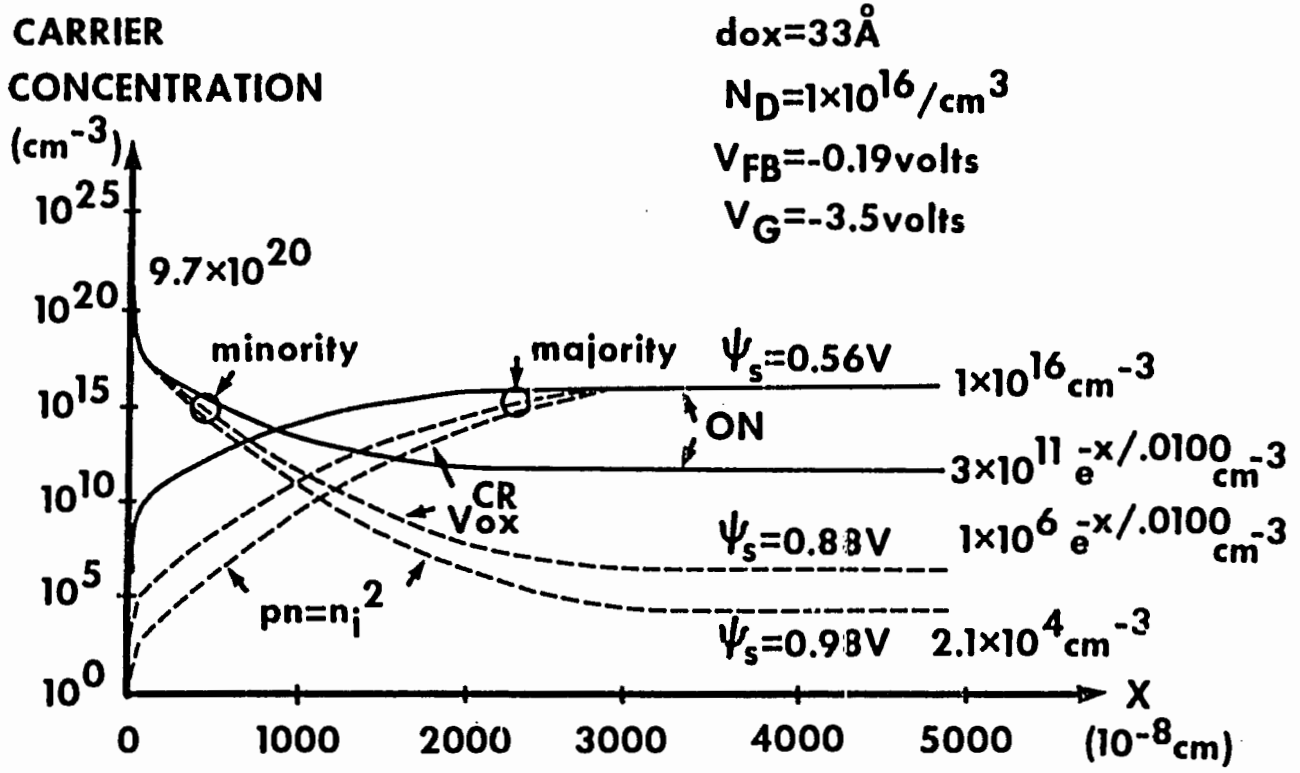
At the new depth $x' = x + \delta x$, the new electric field $\mathcal{E}(x')$ can be determined by using the new electrostatic potential $\psi' = \psi + \delta\psi$ and Eq. B-11.

In this manner, $\psi(x)$ may be numerically computed. From $\psi(x)$, $n(x)$ and $p(x)$ may also be computed, and the depletion depth x_d determined. Simultaneously, the Shockley-Read-Hall recombination current may be computed, by numerically integrating the volume net recombination rate U over the space region. The results of these calculations for the MTOS device are shown in Fig. B-1.

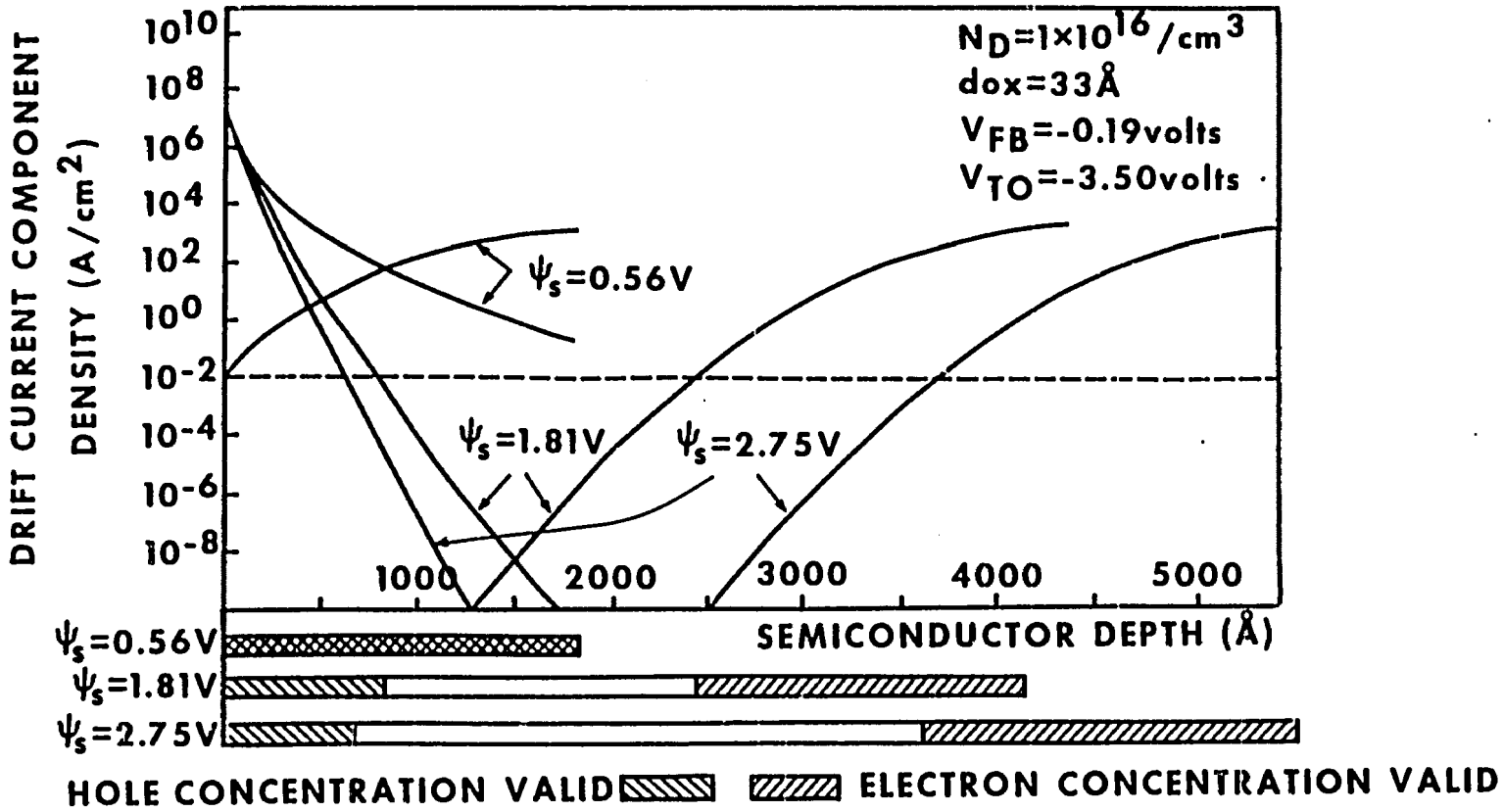
Model Validation

The validity of the zero current approximation is now discussed. As in all cases of validating an approximation, one validates the method with respect to a result. In this case, the surface potential should be shown to be properly related to the device capacitance in an MTOS device where a significant current flows. The procedure is that the region of failure of the zero current approximation (as defined by Eq. B-3) is first identified, and then it is shown that such a failure has negligible effect on the C-ψ_s relationship.

In Fig. B-2, the computer model has been used to plot the drift current components ($q\mu_p p\mathcal{E}$ and $q\mu_n n\mathcal{E}$) in the space-charge region for an n-type semiconductor with constant doping $N_D = 1.0 \times 10^{16}/cm^3$ for various values of surface potential. In actual MTOS device operation ($d_{ox} = 33\text{Å}$, $V_G = -3.5$ volts, $V_{FB} = -0.19$ volts) the total current density does not exceed $1 \times 10^{-2} A/cm^2$. Thus for drift current densities far above this value (say $1 \times 10^{-2} A/cm^2$), the zero-current approximation is valid. From Fig. B-2, the approximation is valid near the surface



B-1 Carrier densities in non-equilibrium MTOS junction space-charge region.



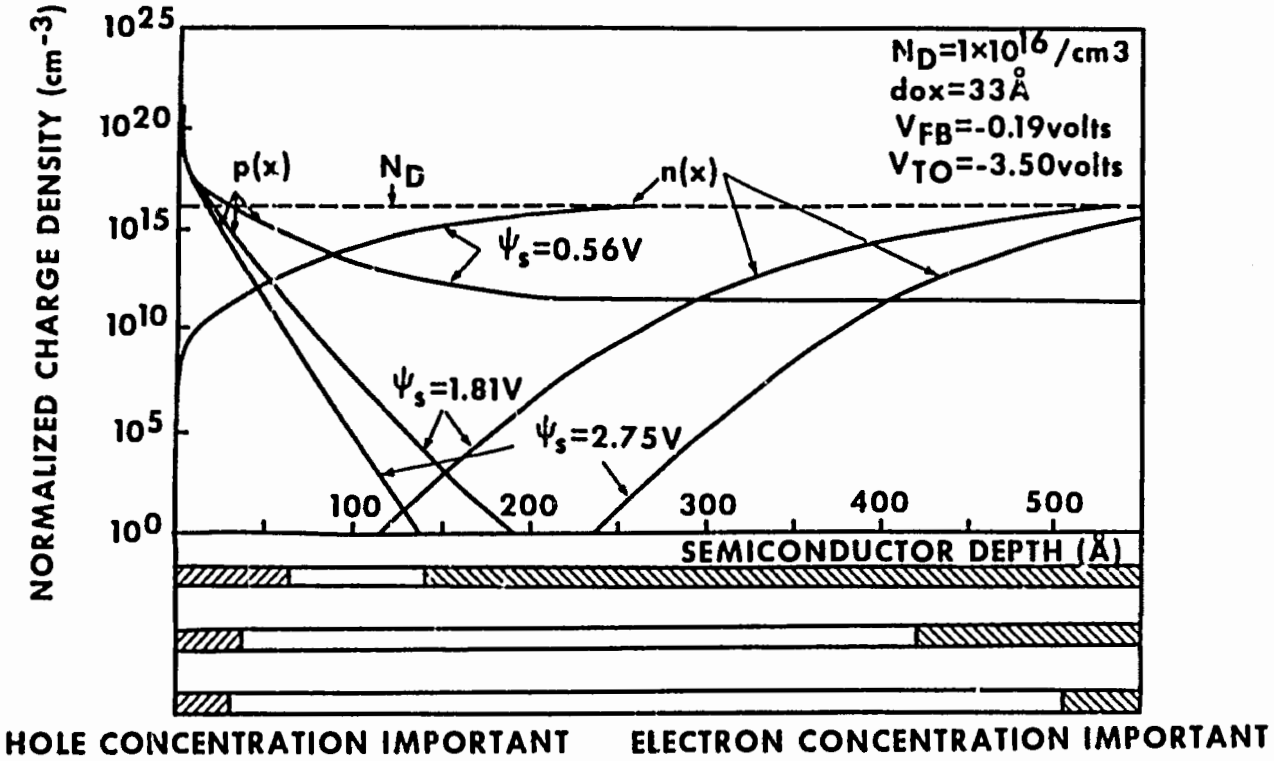
B-2 Drift current component of total current in non-equilibrium MTOS junction using zero current approximation.

for holes, and near the depletion region edge for electrons.

To obtain the C- ψ_s relationship, Poisson's equation was integrated from the bulk to the surface. The zero-current approximation was used to determine the charge density $\rho(x)$ in Poisson's equation. Integrating from the bulk, $\rho(x)$ is at first strongly dominated by $n(x)$, the electron concentration, hence the failure of the zero-current approximation in estimating $p(x)$ in this region is inconsequential (see Fig. B-3). In the space-charge region away from the depletion region edge, the ionized donor concentration N_D determines $\rho(x)$, so that any error in using the zero current approximation for $n(x)$ beyond the depletion region edge becomes irrelevant. Toward the surface, $\rho(x)$ is determined by the hole concentration $p(x)$. However, in this region, $p(x)$ is well described by the zero-current approximation. (This is why it was important to write $p(x)$ in terms of the surface hole concentration p_s). Throughout the space-charge region, the zero current approximation yields a valid estimate of the charge density. This charge density profile in turn yields a valid value of ψ_s, Q_{INV} and Q_D. It is concluded that the C- ψ_s relationship obtained from the computer model is accurate.

Currents

Finally, two currents are calculated using the zero-current approximation. First, note from Fig. B-2 that in super-inversion, the hole concentration is valid up to the depletion region edge. As in a forward biased p-n junction, the hole concentration at the depletion region edge is used to calculate a hole diffusion current from the surface to the bulk. The current is



B-3 Charge components in non-equilibrium MTOS junction.

$$J_{diff} = \left(\frac{q\mu_p kT}{\bar{\tau}_p}\right)^{1/2} \left[p(x_d) - p_{no}\right] \qquad \text{B-30}$$

The total rate of recombination/generation in the space-charge region is calculated using a simplified Shockley-Read-Hall expression[1] for the local net recombination,

$$J_{rec} \approx \int_{x=0}^{x=x_d} \frac{pn-n_i^2}{\bar{\tau}_p(p+n+2n_i)}\, dx \qquad \text{B-31}$$

It is interesting to note that the zero-current approximation yields a constant p-n product, as can be seen from Eqs. B-6. The result is

$$pn = p_s N_D e^{\beta\psi_s} \qquad \text{B-32}$$

Thus, the carrier concentrations are always arranged so that throughout the space-charge region, there is always net recombination or generation, but not both. (i.e. there is not net recombination in the inversion layer and net generation in the depletion region). Once the sign of the numerator in Eq. B-31 is established, the magnitude of the local rate is modulated by the p-n sum.

---

[1]The use of a constant $\bar{\tau}$ in the integral is, in general, incorrect, especially as $x \to 0$ and surface traps become important. See chapter 2.

APPENDIX C

HOT TUNNELING ELECTRON INDUCED IMPACT IONIZATION CURRENT IN SILICON

An electron in the silicon conduction band with a few electron volts of kinetic energy is considered 'hot' because its equivalent thermal energy is significantly higher than the majority of conduction electrons. In equilibrium, the conduction electrons are distributed in energy as described by the Fermi free electron gas approximation. That is,

$$n(E) = \frac{6\sqrt{2}}{\pi^2} \cdot \frac{(E-E_C)^{1/2}}{\hbar^3} \cdot (m_l^{*1/2} m_t^*)\left(1+\exp\left(\frac{E-E_F}{kT}\right)\right)^{-1} \qquad C\text{-}1$$

At a few eV above the conduction band edge (say $E - E_c = 2.5 \pm 0.1$ eV) $n(E)\delta E$ in equilibrium, for $N_D = 1 \times 10^{16}$ cm^{-3} is approximately one electron for every six cubic meters of silicon.

In the MTOS device, in the ON state, the current density is in the range 10^{-4} to 10^{-3} A/cm^2. For this flux level, distributed in the same energy range used above, the electron density is of the order of 10^3 to 10^4 per cm^3. These hot electrons 'thermalize' with the silicon through a series of inelastic collisions with the silicon lattice, emitting phonons, and occasionally knocking loose bound electrons. It is the latter process, referred to as impact ionization, which is to be examined.

In this appendix, the approach of Drummond and Moll (1971) will be taken in calculating the total probability that a hot electron of initial energy E above the silicon conduction band, will generate an

electron-hole pair through the impact ionization process. Before proceeding however, it is interesting to discuss a portion of the historical development of this model.

Most models of the impact ionization process have been phenomenological in nature. The models tend to assume (justifiably so) that the hot electron energy must exceed a minimum threshold level below which impact ionization cannot occur. The hot electron energy must be above E_{gap}, the silicon band gap energy, in order to create an electron-hole pair, and higher if momentum is to be conserved. In 1954, Wolff suggested that the threshold energy was 2.3 eV. In 1957, Miller measured the threshold energy to be approximately 1.5 eV, and in the same year, Chynoweth and McKay reported that they obtained a value of 2.25 eV. Both values were inferred from avalanche multiplication experiments in P-N junctions.

In 1961, Shockley deduced that the impact ionization threshold energy was 1.1 eV, the silicon band gap energy, and introduced a simple model to describe the process. In his model, the total impact ionization probability P_{ii} was given by

$$P_{ii} = 1 - \exp\left[-(E - E_{ii})/mE_{ph}\right] \qquad C\text{-}2$$

where E_{ii} is the impact ionization threshold energy, E_{ph} is the average scattering phonon energy (=0.063 eV) and m is the ratio between the mean free phonon scattering length (70 Å) and the mean free impact ionization length, with a value of m = 17. This model was employed and discussed by Lai (1979) in his analysis of the impact ionization phenomena in MTOS junctions. Lai obtained a value of E_{ii} = 2.1 eV and m = 42.

Moll and van Overstraeten (1963) performed further avalanche multiplication experiments and extended Shockley's model. They obtained

E_{ii} = 1.8 eV and m = 20 ± 10. Because Moll wrote an introductory semiconductor physics text book shortly thereafter, (1964), the 1.8 eV threshold energy value has become the textbook value.

To understand why a threshold value in the range 1.8 eV to 2.3 eV is reasonable, a simple calculation conserving energy and momentum is performed. If E_{ii} is the initial kinetic energy of the electron, and it assumed all three particles have the same final kinetic energy E (specious, but in line with Moll, 1964, pg. 215), then conservation of energy yields

$$E_{ii} = E_{gap} + 3E \qquad \text{C-3}$$

Momentum conservation, using a parabolic band model yields

$$\sqrt{2m_e^* E_{ii}} = 2\sqrt{2m_e^* E} + \sqrt{2m_h^* E} \qquad \text{C-4}$$

where m_e^* and m_h^* are the effective masses of an electron and hole respectively. Squaring the latter equation yields

$$2m_e^* E_{ii} = 8m_e^* E + 2m_h^* E + 8(m_e^* m_h^*)^{1/2} E \qquad \text{C-5}$$

or,

$$E_{ii} = \left[4 + \left(\frac{m_h^*}{m_e^*}\right) + 4\left(\frac{m_h^*}{m_e^*}\right)^{1/2} \right] E \qquad \text{C-6}$$

Substituting this into the energy conservation statement yields

$$E_{ii} = E_{gap} + 3E_{ii} \Big/ \left[4 + \frac{m_h^*}{m_e^*} + 4\left(\frac{m_h^*}{m_e^*}\right)^{1/2} \right] \qquad \text{C-7}$$

or,

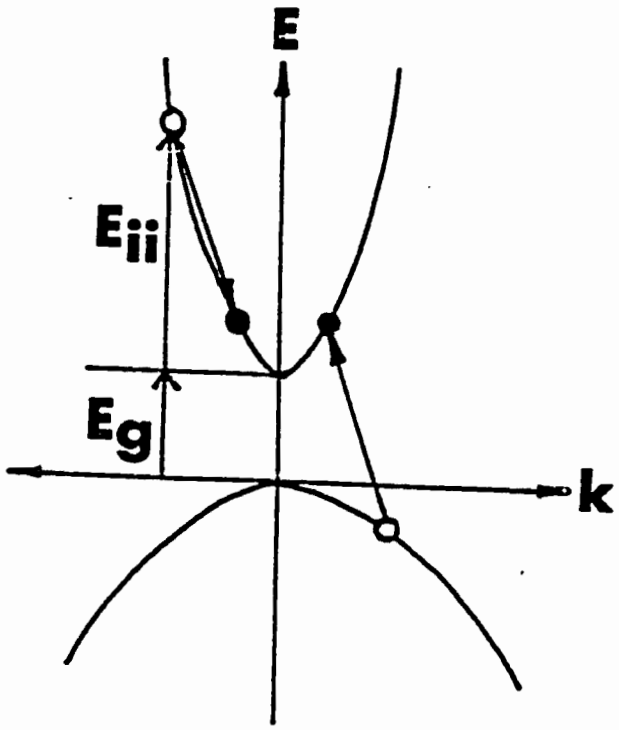
$$E_{ii} = E_{gap} \Big/ \left\{ 1-3 \left[4 + \frac{m_h^*}{m_e^*} + 4\left(\frac{m_h^*}{m_e^*}\right)^{1/2} \right]^{-1} \right\} \qquad \text{C-8}$$

For $m_e^* = m_h^*$, $E_{ii} = \frac{3}{2} E_{gap} = 1.68$ eV, and using $m_e^* = m_t^* = 0.19\ m_o$ and $m_h^* = m_{lh}^* = 0.16\ m_o$ yields $E_{ii} = 1.55\ E_{gap} = 1.74$ eV. However, for an electron tunneling into (100) silicon, $m_e^* = m_l^* = 0.97\ m_o$ and for impact ionization, $m_h^* = m_{lh}^* = 0.16\ m_o$ so that $E_{ii} = 2.1\ E_{gap} = 2.32$ eV. Thus, using parabolic bands with different effective masses for electrons and holes can significantly affect the minimum calculated energy for impact ionization (see Fig. C-1). It is not surprising then that Lai obtained $E_{ii} = 2.1$ eV.

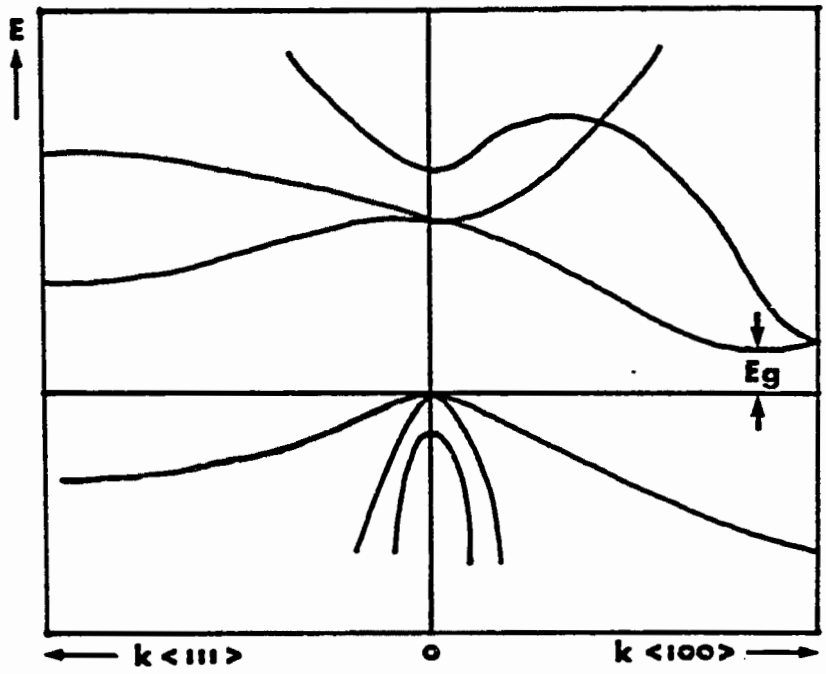
For silicon, an indirect band gap material, a careful calculation should take into account the actual band structure (as shown in Fig. C-2) and occupation probabilities so that transitions like that illustrated in Fig. C-3 are considered. Such a calculation is difficult to do, so that a different approach from that taken by Shockley yields more satisfactory results.

The approach taken by Drummond and Moll (1971) is now discussed. The reader is urged to review this easy to read and well written paper. More complex treatments of the subject have been given since 1971 (for example see Alig, et al., 1980, which unfortunately assumes a free electron band structure) but are more obscure in both the basic physical principles and computational approach.

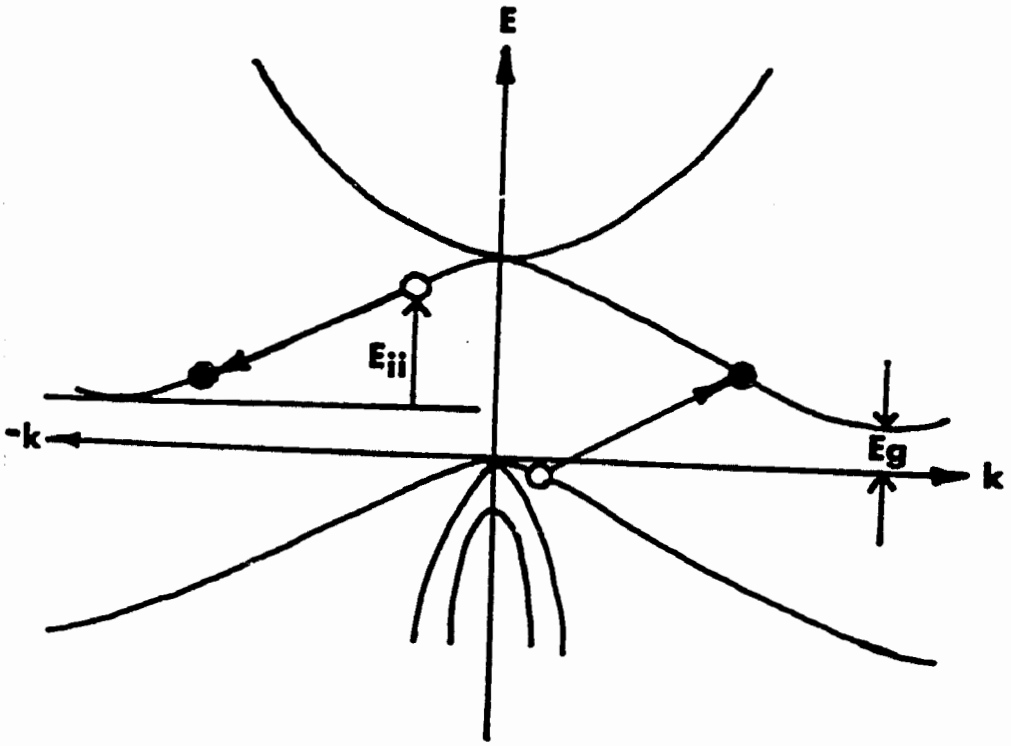
In Drummond and Moll's model, the probability that a hot electron with initial energy E will impact ionize is the sum of the probability that the electron will impact ionize after no phonon loss, plus the probability it will impact ionize after exactly one phonon scattering event, plus the probability it will impact ionize after exactly two



C-1 Energy band diagram illustrating impact ionization.



C-2 Electron dispersion relationship in silicon. (After Cohen and Bergstresser, 1966. Reprinted by permission of the American Institute of Physics)



C-3 Possible impact ionization process in silicon conserving energy and k-vector.

phonon scattering events and so on. Mathematically,

$$P_{ii}(E) = \sum_{n=0}^{N} \bar{P}_{n-1} P_n \qquad \text{C-9}$$

where N is the maximum number of phonon losses which can occur before the electron energy drops below the impact ionization, $\bar{P}_{n-1}$ is the probability that the electron did not impact ionize after the n-1 previous collisions, and P_n is the probability that an electron will impact ionize the silicon after its energy has been reduced by exactly n phonon energy losses.

An electron with energy E scatters through an inelastic process without 'remembering' that it may originally have been at a higher energy, that is, the relaxation process is time invariant. If, during each phonon scattering event, the electron loses energy E_{ph} (=0.063 eV), then

$$[P_n(E)] = [P_{n-1}(E - E_{ph})] \qquad \text{C-10}$$

Electrons with energy E are scattered at a rate $r_{ph}(E)$ by phonons, and at a rate $r_{ii}(E)$ by impact ionization processes. In phonon scattering, the electron may gain or lose energy, but statistically, the loss rate exceeds the gain rate by about an order of magnitude. In addition, the net phonon loss scattering rate is nearly independent of electron energy, and can be taken to be a constant equal to r_{ph}.

The rate of impact ionization scattering on the other hand, is a relatively strong function of energy. According to Drummond and Moll, Kane (1966) has calculated the scattering rate of the impact ionization process as a function of energy using a joint density of states approach and has determined that

$$r_{ii}(E) \propto (E-E_{gap})^{4.2} \qquad \text{C-11}$$

Such a calculation appears to omit momentum conservation explicitly, and the occupation of initial and final states, but does include the silicon band structure.

The probability that a hot electron will be scattered by the impact ionization process, is simply the fraction $r_{ii}/(r_{ph} + r_{ii})$. Thus

$$P_n = \left[1 + \frac{r_{ph}}{r_{ii}}\right]^{-1} = \left[1 + \frac{A}{(E-nE_{ph}-E_{gap})^{4.2}}\right]^{-1} \qquad \text{C-12}$$

The constant A is evaluated by Drummond and Moll by using $r_{ph}/r_{ii} = 2.18$ for E = 5 eV and n = 0, yielding $A = 648\ (e^{-})^{-4.2}$.

Finally, the probability that an electron has not impact ionized the silicon after n-1 collisions is simply

$$\bar{P}_{n-1} = (1 - P_o)\cdot(1 - P_1)\ldots.(1-P_{n-1}) \qquad \text{C-13}$$

Once the total impact ionization probability $P_{ii}(E)$ has been established, the impact ionization current supplied to the MTOS $Si-SiO_2$ interface can be calculated. One writes

$$J_{ii} = \int_{\tilde{E}} J_{et}(E)\, P_{ii}(E)\, dE \qquad \text{C-14}$$

This can be simplified if it is assumed that all the electrons tunnel monoenergetically, which, of course, is only approximate. However, an average impact ionization probability $\bar{P}_{ii}$ can be defined according to

$$\bar{P}_{ii} = \int_E P_{ii}(E) J_{et}(E) dE \; / \; \int_E J_{et}(E) dE \qquad \text{C-15}$$

so that

$$\bar{P}_{ii} = J_{ii}/J_{et} \qquad \text{C-16}$$

It is expected that $P_{ii} \approx P_{ii}\,(E_m - E_c)$ where E_m is the metal Fermi level and E_c is the silicon conduction band edge at the $Si-SiO_2$ interface. The reason is that although the electron tunneling 'centroid' (see Fig. 4.2-3) is below the metal Fermi level, there is a small comparable voltage drop across the inversion layer, which is narrower than the mean free phonon or impact ionization scattering length, thus electrons will gain this energy prior to impact ionization. Such an approximation was supported in part by Lai (1979), who used $\bar{P}_{ii}(E) = P_{ii}\,(E_m - E_c - \delta E)$, where δE was 0.25 eV and E_c was measured using the depletion approximation from capacitance measurements, which calculates the band bending in the silicon up to a point just short of t. inversion layer. The additional drop across the inversion layer is approximately 0.25 - 0.35 eV, as discussed in appendix B. Also, it is believed that the majority of the impact ionizations occur on the order of 50 Å into the silicon. This is supported by experimental evidence that significant band-bending in the depletion region does not affect P_{ii} [D.DiMaria, private communication].

In the present work, $\bar{P}_{ii} = P_{ii}\,(E_m - E_c)$ is used to compare the impact ionization model of Drummond and Moll to the experimental data. The results of this comparison as a function of $E_m - E_c$ over the energy

range probed by MTOS junction transient current technique (2.4 - 2.7 eV) are shown in chapter 4 in Fig. 4.4-9. The results are surprisingly good.

As a final note on the Drummond-Moll model, workers at the University of California, Berkeley, in addition to rediscovering minority carrier transport through thin oxides, have measured the impact ionization probability $\bar{P}_{ii}$ in the range of E = 3 to 5 eV (which extends into the Fowler-Nordheim tunneling regime) and have confirmed the accuracy of the Drummond-Moll model in this energy range [Chang, et al., 1983].

APPENDIX D

CHARGE PACKET OUTPUT SENSE AMPLIFIER

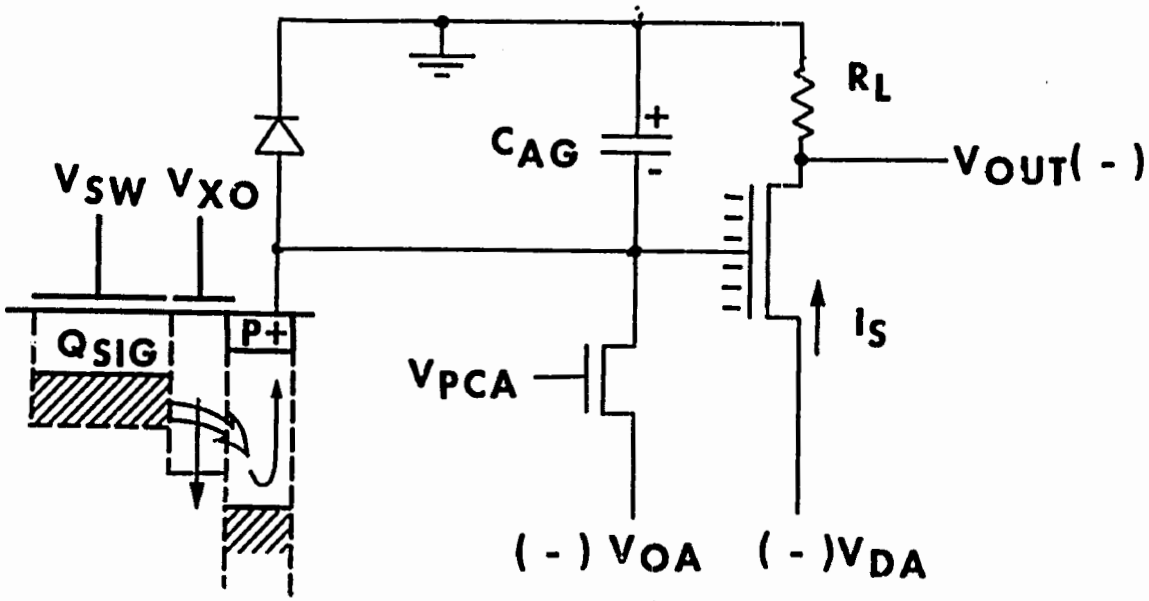
Introduction

There are a number of ways to sense a packet of minority carriers. Some destructive and non-destructive techniques are described by Sequin and Tompsett (1975) and by Beynon and Lamb (1980). The most widely used destructive technique is to use the floating diffusion source-follower amplifier circuit [Kosonocky and Carnes, 1973]. In this method, a source-follower configured MOSFET monitors the voltage of a floating diffusion as it responds to a charge packet of minority carriers. Such a circuit is relatively easy to design and fabricate and can provide a reasonable response (volts/pico-Coulomb). The circuit is also relatively linear. These characteristics motivated the integration of this type of charge packet output sense amplifier on the Yale 8205 test chip.

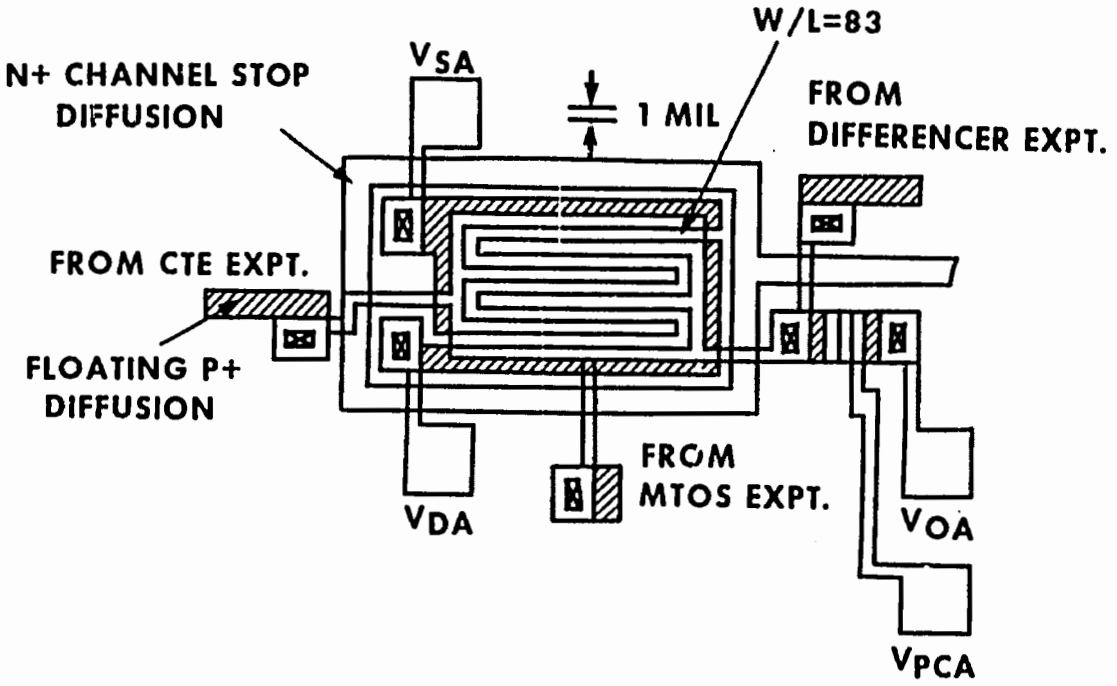
In this appendix, the operating principles and characteristics of the source-follower sense amplifier are discussed and derived. Predicted characteristics are compared to the measured characteristics reported in chapter 5.

Layout and Operation

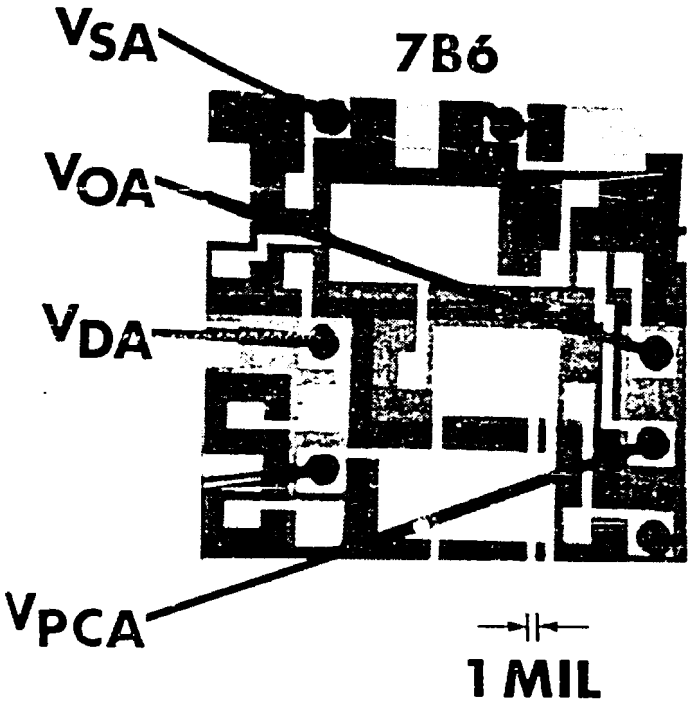
The layout of the output sense amplifier is shown in Fig. D-2. The amplifier consists of two MOSFETs; precharge MOSFET and a large interdigitated output MOSFET. The output MOSFET gate is connected to four floating p^+ diffusions which originate from three experiments and the precharge MOSFET. The total area of the floating diffusions is 100 $mils^2$. The interdigitated source and drain of output MOSFET constitute a channel geometry with W/L equal to 83 if the corners are discounted.



D-1 Schematic illustration of floating diffusion source-follower MOSFET charge packet ouput sense amplifier.



D-2 Layout of amplifier used in the Yale 8205 chip.



D-3 Photograph of actual amplifier tested.

A photograph of a fabricated device is shown in Fig. D-3.

In Fig. D-1 the equivalent circuit of the output sense amplifier is drawn. The operation of the circuit proceeds in two cycles as follows. First, there is a signal charge Q_{SIG} stored in the storage well V_{SW}, which is isolated from the p^+ floating diffusion by V_{XO}. The amplifier is precharged (or reset) by pulsing V_{PCA} on. This causes the output MOSFET gate voltage V_{AG} to be preset to voltage V_{OA} ($V_{PCA} < V_T + V_{OA} < 0$). When V_{PCA} is turned off, the output MOSFET remains charged at voltage V_{AG} due to the negative charge stored on the lumped nodal capacitance C_{AG}. A current flows through the output MOSFET and load resistance R_L, resulting in a non-zero value of V_{OUT}.

The second cycle occurs when the output transfer gate V_{XO} is pulsed on causing the minority carrier signal charge to be collected by the reverse biased p^+ diffusion. This diffusion is floating because its bias is maintained by the charge stored on the nodal capacitance C_{AG}. The minority carriers (positive holes) recombine with a portion of the electrons stored on C_{AG}, reducing the charge stored on the output MOSFET gate by amount Q_{SIG}. Through the capacitance C_{AG}, the effective MOSFET gate voltage is reduced by amount

$$\Delta V_{AG} = Q_{SIG}/C_{AG} \qquad \text{D-1}$$

The resultant change in output MOSFET current is reflected by a reduction in the output voltage ΔV_{OUT}. By choosing a large load resistance, the variation in V_{OUT} can be made to be quite linear in signal charge.

Analysis of the Source-Follower MOSFET Configuration

A simple model of the source-follower configured MOSFET can be used to obtain the conditions of linearity. The MOSFET saturation current I_s can be written approximately as [Streetman, 1972].

$$I_s = \frac{\bar{\mu}C_{ox}W}{2L} (V_G - V_S - V_T)^2 \qquad \text{D-2}$$

where μ is the average minority carrier surface mobility, V_G is the gate voltage, V_S is the source voltage, and V_T the MOS threshold voltage, and assuming that $V_D < V_G - V_T < 0$.

For the source-follower configuration, the source voltage is

$$V_S = I_S \cdot R_L \qquad \text{D-3}$$

Substituting this in Eq. D-2 and expanding yields

$$I_s = \frac{\bar{\mu}C_{ox}W}{2L} \left[(V_G - V_T)^2 - 2I_S R_L (V_G - V_T) + I_S^2 R_L^2 \right] \qquad \text{D-4}$$

This quadratic equation in current can be solved to give

$$I_S = \left(\frac{R_L^2 \mu C_{ox} W}{L} \right)^{-1} \left\{ \frac{R_L \mu C_{ox} W}{L} (V_G - V_T) + 1 - \left[\frac{2 R_L \mu C_{ox} W}{L} (V_G - V_T) + 1 \right]^{1/2} \right\} \qquad \text{D-5}$$

Therefore, the output voltage V_{OUT} behaves according to

$$\frac{dV_{OUT}}{dV_G} = 1 - \left[\frac{2R_L \mu C_{ox} W}{L} (V_G - V_T) + 1 \right]^{-1/2} \qquad \text{D-6}$$

The non-linear part of this last result can be made arbitrarily small by choosing R_L very large such that

$$2R_L \, \mu C_{OX} \frac{W}{L}(V_G - V_T) > > 1$$

or

$$\frac{1}{R_L} < < \quad 2 \, \mu C_{OX} \, \frac{W}{L} \, (V_G - V_T) \qquad \text{D-7}$$

This latter expression states that linear response is obtained if the output MOSFET saturation transconductance is much larger than the load conductance. For example, with W/L equal to 83, C_{OX} = 60 nF/cm^2 (d_{OX} = 575 Å), μ = 150 cm^2/V sec, and $V_G - V_T$ = 10 volts, the output MOSFET transconductance is 0.015 mhos. Thus, choosing R_L > 1 k Ω should satisfy the linearity requirement.

The response of the source-follower configuration to a step input is adversely affected by choosing R_L large due to the gate to source and gate to channel capacitance. To minimize R-C charging and discharging times, R_L should be chosen to be as small as possible. To compromise between these conflicting design criteria, the channel geometry is chosen to give a large value of W/L so that Eq. D7 is satisfied with a minimal value of load resistance. In some designs, the output sense amplifier is made with an integrated active load but this option was not pursued in the Yale 8205 layout.

The transfer characteristics of a fabricated device (7B6) is shown in Fig. 5.4-1. As was discussed in Chapter 3, the diffused p^+ layer sheet resistance was anomalously high for lot 7B due to a defective dopant source. Such high sheet resistance (2.1 kΩ/□) adversely affects the MOSFET properties of the interdigitated output sense amplifier by causing most of the current to flow at the base of the fingers. The effective W/L of this device works out to be only 3- 5 which is only a few percent of the design. Thus the minimum value of R_L required for

linear operation with near unity gain was increased.

Response to a Charge Packet

The source-follower output amplifier once precharged responds to a minority carrier charge packet through the recombination of a portion of the negative gate charge with the positive signal charge. Since the gate charge is reduced, the effective gate is reduced according to Eq. D-1. The lumped nodal capacitance C_{AG} (which is primarily due to the output MOSFET gate capacitance) determines the magnitude of the response. Increasing this capacitance decreases the sensitivity but increases the range. A rule of thumb is that the capacitance should be about twice the gate oxide capacitance of a typical storage bucket.

To calculate the nodal capacitance C_{AG} of a lot 7B device, the components of the lumped capacitance are summed. These components are (in decreasing order of importance) the gate to drain overlap capacitance, the gate to channel capacitance, floating diffusion capacitance, gate to source capacitance, and miscellaneous wiring capacitances.

The gate to drain overlap capacitance is the product of the specific gate oxide capacitance (0.38 pF/mil^2 for lot 7B) and the overlap area (61 mils^2) which is 23.2 pF. The gate to channel capacitance depends upon the amplifier gain. With unity gain it would be zero, but with a gain of 0.81, the capacitance is 0.19 of the grounded source configuration gate to channel capacitance. This latter capacitance for a saturated MOSFET is the product of the channel area with two-thirds of the specific oxide capacitance. With a channel area of 93 mils^2, this total source-follower configured gate to channel capacitance works out to be 6.7 pF. The gate to source capacitance is also 0.19 of the product of overlap area (60 mils^2) and specific oxide capacitance which

is 4.3 pF.

The specific floating diffusion capacitance is calculated by using

$$C_{FD} = \left[\varepsilon_s \varepsilon_o q N_D / 2(V_{BI}+V_{AG})\right]^{1/2} \qquad \text{D-8}$$

where V_{BI} is the built-in diode voltage and V_{AG} is the floating diffusion reverse bias. Although non-linear, the capacitance can be treated as being approximately constant using a nominal value of V_{AG}. Evaluating Eq. D8 for $N_D = 1 \times 10^{16}/cm^3$ and $V_{BI} + V_{AG}$ equal to 10 volts and multiplying by the floating diffusion area (100 $mils^2$) yields a total floating diffusion capacitance of approximately 6 pF.

Finally, miscellaneous wiring capacitances account for about 3 pF. The lumped total capacitance C_{AG} is therefore 43 pF. By calibrating the output amplifier of device 7B6 it was determined experimentally that the lumped capacitance was 46 pF, which is quite close to the predicted value.

The output sense amplifier may be characterized by an output transcapacitance C_{OUT}, which is defined according to

$$C_{OUT} = \partial Q_{SIG} / \partial V_{OUT} \qquad \text{D-9}$$

which relates a differential change in signal charge to a corresponding differential change in output voltage. The output transcapacitance can be written in terms of the source-follower gain α_A and lumped capacitance C_{AG} as

$$C_{OUT} = \alpha_A C_{AG} \qquad \text{D-10}$$

For device 7B6, the calibrated output transcapacitance was determined to be 37.7 pF.

Amplifier Droop and Clock Feedthrough

The floating sense diffusions of the output amplifier are, of course, reverse biased PN junctions in which a non-zero reverse 'leakage' current flows. This current discharges the lumped amplifier capacitance C_{AG} causing the output voltage to droop. For good PN junctions, the discharge can be a slow process compared to the reset rate of the amplifier. The droop rate can be written as

$$\frac{dV_{OUT}}{dt} = \frac{-I_{LEAK}}{C_{OUT}} \qquad \text{D-11}$$

where I_{LEAK} is the PN junction leakage current. For a high quality PN junction leakage current of 10 nA/cm^2, the amplifier droops at 0.2 volts/sec which is not a problem. For poor junctions, the leakage current can be much greater, especially at large (> 10 volts) precharge voltages. At these voltages, soft PN junction can occur leading to severe amplifier droop.

Feedthrough from the precharge (or 'reset') MOSFET can be a problem. When the precharge MOSFET turns off, its channel charge is split between the source and drain. The minority carrier charge which flows toward the source changes the precharged amplifier output voltage by the amount δV_{OUT} given approximately by

$$\delta V_{OUT} \approx \frac{1}{2} C_{ox} \cdot (V_{PCA} - V_T - V_{OA})/C_{OUT} \qquad \text{D-12}$$

where C_{OX} is the total channel gate oxide capacitance of the precharge MOSFET. For C_{OX} = 1.5 pF, V_{PCA} = -25 volts and V_{OA} = -12 volts, the expected feedthrough is 0.2 volts. In practice, the feedthrough may be somewhat larger than this amount.

This d.c. offset due to clock feedthrough by itself is not a problem. More troublesome is that the feedthrough process tends to be noisy and may limit the sensitivity of the amplifier. In commercial circuits, the precharge MOSFET is usually a dual gate device in which the gate closest to the source is biased at a d.c. level and acts as a screen when the other gate is pulsed on and off. In retrospect, implementing such a dual gate precharge MOSFET would have been trivial once the submicron gap technology was established. However, the parallel channel of the charge packet initiated MTOS junction switching experiment can be operated as a tri-gate or dual gate precharge MOSFET for sensitive experiments, though this will result in the likely destruction of the MTOS junction.

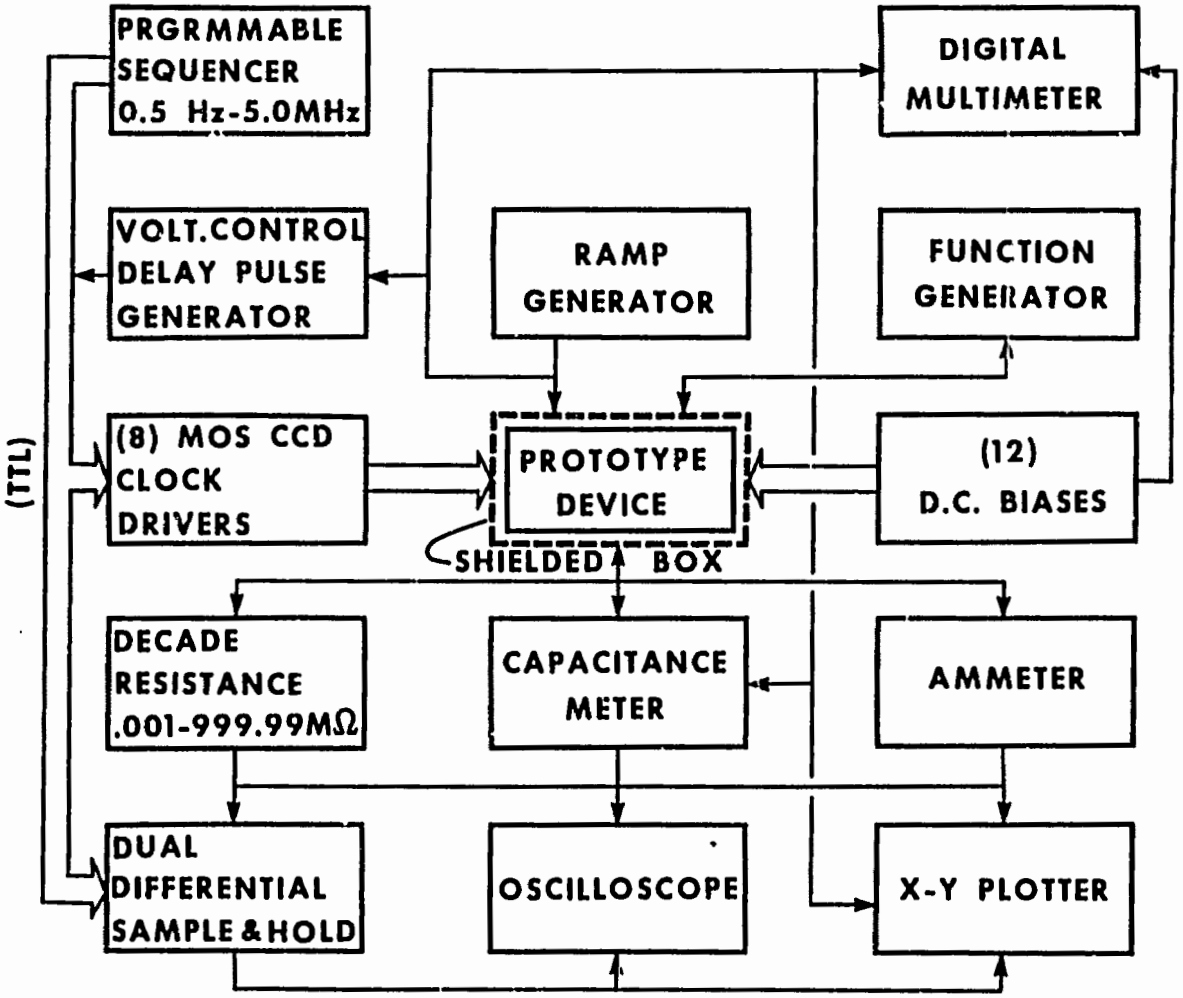
APPENDIX E

INSTRUMENTATION

This appendix summarizes the instrumentation used in measuring the prototype devices. A block diagram which partially illustrates a generic measurement set-up is shown in Fig. E-1, though each measurement requires a uniquely tailored configuration. The instrumentation can be broadly divided into control equipment and sense equipment. Control equipment includes a programmable sequencer, a voltage controlled delay pulse generator, MOS CCD clock drivers, D.C. biases, a ramp generator, and a function generator. Sense equipment includes a digital multimeter, an electrometer (ammeter), a capacitance meter, a decade resistance box, and an oscilloscope. Also utilized is a dual differential sample and hold circuit (which assists in acquiring time dependent signals), and an X-Y plotter.

At the heart of the instrumentation is a shielded box for making connections to the packaged prototype devices. The box has two zero insertion force sockets; one for a 24 pin dual-in-line-package, and the other for a 44 pin leadless carrier. A hinged hatch allows access to the sockets and is battened down by two wing nuts. This hatch provides for light tight and electromagnetically shielded operation of the experimental devices. The sockets are mounted on a custom printed circuit board for fan-out to connection points. Inside a second shielded box butted to the first, co-axial cables connect each socket pin via the printed circuit board to an external BNC female plug. Some of the BNC plugs are shared between the two sockets.

The ramp generator is a custom-design/custom-built instrument providing a slow to high speed low noise ramped voltage output which can



E-1 Block diagram illustrating instrumentation set up.

source or sink up to 200 mA. The function generator is a commercial instrument (Wavetek model 145) which provides square, triangular, and sine wave outputs over a wide range of amplitudes, frequencies, and D.C. level offsets. Also used, but not illustrated in Fig. E-1, are high current ouput D.C. power supplies which are useful for 'burning' out occaisional inter-electrode short circuits, and providing less critical D.C. biases. The twelve D.C. bias controls are actually a bank of ten-turn 10 kΩ potentiometers which deliver a voltage between ground and a D.C. level typically set at 20 volts (2.0 volts/turn). Most of the chip inputs require little (reversed biased PN junction) or no (MOS capacitor) current so that the potentiometer approach suffices for the bias. The output of all the D.C. biases are periodically monitored by a DVM (Tektronix model DM501a).

The clocked biases for testing the prototype devices are delivered by a bank of eight MOS CCD clock drivers (Pulse Instruments model PI-451). These drivers allow for independent control of high and low clock levels, and rise and fall times. The high and low levels can be adjusted through the front panel control knobs, through an external D.C. inputs, or by remote programming. These clock drivers are essentially sophisticated level shifters which take an input TTL signal and convert it to shifted high and low values. The TTL signals used to sequentially drive the clocked biases originate in a custom built but general purpose programmable sequencer. The sequencer consists of a block of high speed static (volatile) memory chips organized as 256 twelve bit words. The memory is read out in word format and the bit values are buffered and delivered in parallel to BNC outputs on a rear panel. The memory is loaded via a front panel switch register and display register. When running, the sequencer steps through the memory at a rate which can be varied with a front panel knob from 0.3 Hz to 5.0 MHz. When the internal program counter address is equal to the value of the front

panel switch register, the program counter is reset to address 000 and the sequence is repeated. In this manner, the clock waveforms delivered to the prototype device can be easily varied both in content and in speed. Although not implemented as a remotely programmable instrument, the programmable sequencer may be easily adapted by future investigators should the need arise. (Indeed, it is somewhat tedious to manually load a long sequence by hand toggling the switch register).

Although there are only eight CCD clock drivers, the additional four bits emanating from the programmable sequencer are useful. For example, two of these bits are used to trigger the dual differential sample and hold circuit discussed below. Another bit is used to trigger the oscilloscope sweep to synchronize the display to the control sequence. Occaisionally, it is also useful to generate a control pulse which is of short duration and placeable anywhere in the clock sequence as determined by a input control voltage. An example of this occurs in the case of converting a waveform on the oscilloscope (e.g. MTOS I-t characteristic) to the X-Y plotter. In this case, the control voltage is used to position a current sampling pulse and simultaneously drive the X-axis of the plotter. The sampled signal (averaged over many waveforms) is input to the Y-axis, thereby completing the waveform translating function.

To sense the output of the prototype circuit, several instruments are employed, though not usually at the same time. For example, for the charge packet differencing experiment, the output of the charge packet sense amplifier is read across a load resistor and displayed on the oscilloscope. The output signal is also fed in parallel to the dual differential sample and hold circuit. For the MTOS junction experiments, the electrometer (as an ammeter), or capacitance meter is more frequently used. For static measurements, the output of the

capacitance meter, the output of the ammeter, or the resistor generated voltage is sent directly to the X-Y plotter. For dynamic measurements, the sample and hold circuit is used to convert the dynamic signal sampled at some instant of time, to a static signal, which in turn is then used in the X-Y plotter.

The decade resistance box is a homemade switched circuit which yields a resistance from 1 kΩ to 999.999 MΩ in 1 kΩ increments. The capacitance meter is a commercial instrument (Boonton model 72B) which measures the small signal capacitance of a device with a 1 MHz 15mV rms probe signal. For dynamic measurements, the response time of the Boonton meter is 1 msec, though response times smaller than 10 msec are never actually required. The ammeter is also a commercial precision instrument (Keithley model 610C electrometer). The response time of this instrument in the low impedance feedback fast mode is of a similar time scale. The oscilloscope is a Hewlett-Packard model 180A unit and the X-Y recorder is made by Houston Instruments (model 2000).

The dual differential sample and hold circuit has proven to be surprisingly versatile. This instrument, which is of the home grown variety, consists of two sample and hold circuits, labelled A and B. Each sample input is buffered by a high impedance (1.5 TΩ) CMOS voltage follower configured operational amplifier. The op amp output drives a switched capacitor and second buffer stage. The switch is a high performance CMOS device and is closed in the sample mode so that the second stage follows the first stage. In this mode, the instrument may be used as an ultra-high impedance buffer. In the hold mode, the switch disconnects the first stage from driving the second, and the capacitor voltage is sensed by the high impedance second stage. Of course, the capacitor charge eventually leaks away, as evidenced by an output voltage droop, but the droop rate is slow, (volts/hour) and is

determined by the capacitor size. Larger capacitors lead to longer charging intervals and discharging intervals (longer acquistion time) and a trade-off must be made. Each channel has a separate (TTL triggered) sample/hold control input so that the same signal may be sampled at two different times. The output of the two parallel second stages are independently sent to the front panel for output, and in addition, they are fed into a differential amplifier. The calibrated differential output is also sent to the front panel. As an example of a typical application, the output of the charge packet output sense amplifier is sent to both the A and B input channels of the dual differential sample and hold circuit. The B channel is sampled following amplifier reset. After the charge packet has been sensed, and the amplifier has responded, the output is sampled by channel A. Since the change in output voltage is what matters in the source-follower circuit, the A-B sample and hold circuit output is proportional to the sensed charge packet.

Perhaps the major problem with measuring charge-coupled devices is test station noise. Precautions were taken in the set up to reduce ground loop noise and to shield against electromagnetic pick-up, and the noise in the test station was reduced to the one millivolt range. However, as indicated in chapter five, the noise from the semiconductor devices themselves is expected to be in the tens of microvolts range. Thus, the Yale set-up, as in many industrial laboratories, continues to be dominated by test station noise. (This is true even for static measurements made in the Yale screened room test area.)

REFERENCES

Alig, R.C., S. Bloom, and C.W. Struck, "Scattering by Ionization and Phonon Emission in Semiconductors", Phys. Rev. B22, 5565 (1980).

Amelio, G.F., M.F. Tompsett, and G.E. Smith, "Experimental Verification of the Charge-Coupled Device Concept", Bell Syst. Tech. J. 49, 593 (1970).

Amelio, G.F., W.J. Bertram, and M.F. Tompsett, "Charge-Coupled Imaging Devices: Design Considerations", IEEE Trans. Electron Devices 18, 986 (1971).

Amelio, G.F., "Computer Modeling of Charge-Coupled Device Characteristics", Bell Syst. Tech. J. 51, 705 (1972).

Antonsson, D., B. Gudmundsson, T. Hedblom, B. Kruse, A. Linge, P. Lord, and T. Ohlsson, "PICAP - A System Approach to Image Processing", IEEE Trans. Comput. 31, 997 (1982).

Barbe, D.F., "Imaging Devices using the Charge-Coupled Concept", Proc. IEEE 63, 38 (1975).

Batcher, K.E., "Design of a Massively Parallel Processor", IEEE Trans. Comput. 29, 836 (1980).

Beynon, J.D.E. and D.R. Lamb, Charge-Coupled Devices and their Applications, McGraw-Hill Book Company, New York (1980).

Briggs, F.A., K-S. Fu, K. Hwang, and B.W. Wah, "PUMPS Architecture for Pattern Analysis and Image Database Management", IEEE Trans. Comput. 31, 969 (1982).

Browne, V.A. and K.D. Perkins, "A Nonoverlaping Gate Charge-Coupling

Technology for Serial Memory and Signal Processing Applications", IEEE Trans. Electron Devices 23, 271 (1976).

Bullis, W.M., "Properties of Gold in Silicon", Solid-State Electron. 9, 143 (1966).

Carnes, J.E., W.F. Kosonocky, and E.G. Ramberg, "Drift-Aiding Fringing Fields in Charge-Coupled Devices", IEEE J. Solid-State Circuits 6, 322 (1971).

Carnes, J.E. and W.F. Kosonocky, "Noise Sources in Charge-Coupled Devices", RCA Rev. 33, 327 (1972).

Carnes, J.E. and W.F. Kosonocky, "Fast-Interface-State Losses in Charge-Coupled Devices", Appl. Phys. Lett. 20, 261 (1972).

Carnes, J.E., W.F. Kosonocky, and E.G. Ramberg, "Free Charge Transfer in Charge-Coupled Devices", IEEE Trans. Electron Devices 19, 798 (1972).

Chang, C., M-S. Liang, C. Hu, and R.W. Brodersen, "Charge Tunneling Related Phenomena in Thin Oxide MOSFETs", IEDM Tech. Digest, Washington, D.C. (1983), p. 194.

Chatterjee, P.K., G.W. Taylor, A.F. Tasch, and H-S. Fu, "Leakage Studies in High-Density Dynamic MOS Memory Devices", IEEE Trans. Electron Devices 26, 564 (1979).

Chin, M-R., "RF Plasma Annealing on MOS Structures", Ph.D. Thesis, Yale University (1981).

Chynoweth, A.G. and K.G. McKay, "Threshold Energy for Electron-Hole Pair Production by Electrons in Silicon", Phys. Rev. 108, 29 (1957).

Cohen, M.L. and T.K. Bergstresser, "Band Structures and Pseudopotential Form Factors for Fourteen Semiconductors of the Diamond and Zinc-blende Structures", Phys. Rev. 141, 789 (1966).

Cellins, D.R., W.C. Rhines, J.B. Barton, S.R. Shortes, R.W. Brodersen, and A.F. Tasch, "Electrical Characteristics of 500-Bit Al-Al_2O_3-Al CCD Shift Registers", Proc. IEEE 62, 282 (1974).

Collins, T.W. and J.N. Churchill, "Exact Modeling of the Transient Response of an MOS Capacitor", IEEE Trans. Electron Devices 22, 90 (1975).

Danielsson, P-E. and S. Levialdi, "Computer Architectures for Pictorial Information Systems", IEEE Computer 14(11), 53 (1981).

Deal, B.E., E.L. MacKenna, and P.L. Castro, "Characteristics of Fast Interface States Associated with SiO_2-Si and Si_3N_4 - SiO_2-Si Structures", J. Electrochem. Soc. 116, 997 (1969).

Dressendorfer, P.V. "Interface and Electron Tunneling Properties of Thin Oxides on Silicon", Ph.D. Thesis, Yale University (1978).

Dressendorfer, P.V., S. K-C. Lai, R.C. Barker, and T-P. Ma, "Processing Dependence of Metal/Tunnel-Oxide/Semiconductor Junctions", Appl. Phys. Lett. 36, 850 (1980).

Drummond, W.E. and J.L. Moll, "Hot Carriers in Si and Ge Radiation Detectors", J. Appl. Phys. 42, 5556 (1971).

Duff, M.J.B., B.M. Jones, and L.J. Townsend, "Parallel Processing Pattern Recognition System UCPR1", Nuclear Instr. and Methods 52, 284 (1967).

Duff, M.J.B., "Parallel Processors for Digital Image Processing" in Advances in Digital Image Processing, P. Stucki, Ed., Plenum Press, New York (1979), p. 265.

Engeler, W.E., J.J. Tiemann, and R.D. Baertsch, "Surface Charge Transport in Silicon", Appl. Phys. Lett. 17, 469 (1970).

Espley, D.C., "The Calculation of Harmonic Production in Thermionic Valves with Resistive Loads", Proc. IRE 21, 1439 (1933).

Fagan, J.L., "Monolithic Implementation of a Fast Fourier Transform", United States Patent 4,085,441 (April, 1978).

Forster, J.H. and H.S. Veloric, "Effect of Variations in Surface Potential on Junction Characteristics", J. Appl. Phys. 30, 906 (1959).

Fossum, E.R., "Charge-Coupled Computing Elements for Parallel Image Processing", Ph.D. Thesis Proposal, Yale University (1981) (unpublished).

Fossum, E.R., "Charge Packet Differencer/Replicator", Patent Disclosure PD82260, Hughes Aircraft Company (Sept., 1982).

Fossum, E.R. and R.C. Barker, "Charge Packet Initiated Switching of Metal/Tunnel-Oxide/Semiconductor Junctions" to be published in IEEE Electron Devices Lett. 5(4), (1984).

Fossum, E.R. and R.C. Barker, "Measurement of Hole Leakage and Impact Ionization Currents in Bistable Metal/Tunnel-Oxide/Semiconductor Junctions", to be published in IEEE Trans. Electron Devices 31, (1984).

Fouse, S.D., G.R. Nudd, and P.A. Nygaard, "Implementation of Image Preprocessing Functions using CCD LSI Circuits", in Proc. of the SPIE 225, IR Image Sensor Tech., (1980), p. 118.

Franz, W., in Handbuk der Physik, 17, 155 edited by S. Flugge, Springer, Berlin (1956).

Goetzberger, A. and W. Shockley, "Metal Precipitates in Silicon PN Junctions", J. Appl. Phys. 31, 1821 (1960).

Goetzberger, A. and E.H. Nicollian, "MOS Avalanche and Tunneling Effects in Silicon Surfaces", J. Appl. Phys. 38, 4582 (1967).

Grant, W.N., "Electron Tunneling in Aluminum-Insulator-Palladium, Silver Junctions", Ph.D. Thesis, Yale University (1970).

Green, M.A., "Properties and Applications of the Metal-Insulator-Semiconductor (MIS) Tunnel Diode", Ph.D. Thesis, McMaster University (1974).

Green, M.A. and J. Shewchun, "Current Multiplication in Metal-Insulator-Semiconductor (MIS) Tunnel Diodes", Solid-State Electron. 17, 349 (1974).

Grinberg, J.L., G.R. Nudd, R.D. Etchells, "A Cellular VLSI Architecture", IEEE Computer 17(1), 69 (1984).

Grove, A.S. and D.J. Fitzgerald, "Surface Effects on PN Junctions: Characteristics of Surface Space Charge Regions under Non-Equilibrium Conditions", Solid-State Electron. 9, 783 (1966).

Grunthaner, F.J., P.J. Grunthaner, R.P. Vasquez, B.F. Lewis, J. Maserjian, and A. Madhukar, "Local Atomic and Electronic Structure of Oxide/GaAs and SiO_2/Si Interfaces using High-Resolution XPS", J. Vac. Sci. Technol. 16, 1443 (1979).

Gruodis, A.J., "One Electron Tunneling through Thin Oxides", Ph.D. Thesis, Yale University (1971).

Hall, R.N., "Electron-Hole Recombination in Germanium", Phys. Rev. 87, 387 (1952).

Hamilton, J.M., J.G. Nash, R.C. Henderson, J.Y. Chu, D.M. Long, and N. Cuk, "High Density Electron Beam Fabricated CCD Memory Components", IEEE Trans. Electron Devices 27, 1655 (1980).

Hayashi, Y., "Switching Phenomena in Thin-Insulator Metal-Insulator-Semiconductor Diodes", Appl. Phys. Lett. 37, 407 (1980).

Hecht, E. and A. Zajac, Optics, Addison-Wesley, Mass (1974) p. 354, p. 140.

Herron, J.M., J. Farley, K. Preston, and H. Sellner, "A General-Purpose

High-Speed Logical Transform Image Processor", IEEE Trans. Comput. 31, 795 (1982).

Hwang, K., "Guest Editor's Introduction: Computer Architectures for Image Processing", IEEE Computer 16(1), 10 (1983).

Iskarlatos, Y., "Inelastic Electron Tunneling Spectroscopy as a Tool in Water Analysis", Ph.D. Thesis, Yale University (1974).

Kane, E.O., J. Phys. Soc. Japan Suppl. 21, 37 (1966).

Kawamura, K. and T. Yamamoto, "A Two-Phase Shift Register using Si Tunnel MIS Switching Diodes", IEEE Trans. Electron Devices 28, 1078 (1981).

Kidode, M., H. Asada, and S. Watanabe, "Local Pattern Processor, PPP", Toshiba Rev. No. 125 (1978).

Kim, C-K. and M. Lenzlinger, "Charge Transfer in Charge-Coupled Devices", J. Appl. Phys. 42, 3586 (1971).

Kim, C-K. and E.H. Snow, "P-Channel Charge-Coupled Devices with Resistive Gate Structure", Appl. Phys. Lett. 20, 514 (1972).

Kosonocky, W.F. and J.E. Carnes, "Two-Phase Charge-Coupled Devices with Overlapping Polysilicon and Aluminum Gates", RCA Rev. 34, 164 (1973).

Kosonocky, W.F., "Charge-Coupled Devices - An Overview" as in Melen and Buss (1977) p. 28.

Kushner, T., A.Y. Wu, and A. Rosenfeld, "Image Processing on ZMOB", IEEE Trans. Comput. 31, 943 (1982).

Lai, S. K-C., "Optoelectronic Properties of Metal/Tunnel-Oxide/Silicon Junctions", Ph.D. Thesis, Yale University (1979).

Lai, S. K-C., P.V. Dressendorfer, T-P. Ma, and R.C. Barker, "Optically Induced Bistable States in Metal/Tunnel-Oxide/Semiconductor (MTOS)

Junctions", Appl. Phys. Lett. 38, 41 (1981).

Lambert, J.L. and M. Reese, "The Gettering of Gold and Copper from Silicon", Solid-State Electron. 11, 1055 (1968).

Lee, H-S. and L.G. Heller, "Charge-Control Method of Charge-Coupled Device Transfer Analysis", IEEE Trans. Electron Devices 19, 1270 (1972).

Luetjen, K. P. Gemmer, and H. Ischen, "FLIP: A Flexible Multiprocessor System for Image Processing", in Proc. Fifth Int'l. Conf. Pattern Recognition, Miami Beach, Florida, (Dec. 1980) 1, 326.

Ma, T-P., "Surface States in Thick Oxide MOS Tunnel Junctions", Ph.D. Thesis, Yale University (1974).

Ma, T-P. and W. H-L. Ma, "Low Pressure RF Annealing: A New Technique to Remove Charge Centers in MIS Dielectrics", Appl. Phys. Lett. 32, 441 (1978).

Ma, T-P., Integrated Microelectronics Laboratory Manual, Yale University (1984).

McCormick, B.H., "The Illinois Pattern Recognition Computer - ILLIAC III", IEEE Trans. Electron. Comput. 12, 791 (1963).

McNutt, M.J. and C.F. Sah, "Determination of the MOS Oxide Capacitance", J. Appl. Phys. 46, 3909 (1975).

Melen, R. and D. Buss, Ed., Charge-Coupled Devices: Technology and Applications, IEEE Press, New York (1977).

Miller, S.L., "Ionization Rates for Holes and Electrons in Silicon", Phys. Rev. 105, 1246 (1957).

Millman, J. and C.C. Halkias, Integrated Electronics: Analog and Digital Circuits and Systems, McGraw-Hill Book Company, New York, (1972)

Moll, J.L. and R. van Overstraeten, "Charge Multiplication in Silicon PN Junctions", Solid-State Electron. 6, 147 (1963).

Moll, J.L., Physics of Semiconductors, McGraw-Hill Book Company, New York, (1964).

Muller, R.S. and T.I. Kamins, Device Electronics for Integrated Circuits, John Wiley and Sons, New York (1977).

Nelson, E.T., C.N. Anagnostopoulos, J.P. Lavine, and B.C. Burkey, "Reduction of Lateral Diffusion of Photoelectrons in Silicon Photodiode Imager Arrays by Internal Gettering", IEEE Trans. Electron Devices 30, 1392 (1983).

Nicollian, E.H., A. Goetzberger, and C.N. Berglund, "Avalanche Injection Currents and Charging Phenomena in Thermal SiO_2", Appl. Phys. Lett. 15, 174 (1969).

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

Nitzan, D., "Assessment of Robotic Sensors" in Workshop on the Research needed to Advance the State of Knowledge in Robotics, April 15-17, 1980, Newport, RI, p. 182.

Nudd, G., "CCDs Image Processing Circuitry" in Proc. Image Understanding Workshop Applications, SAI-78-549-WA (1977).

Nudd, G.R., P.A. Nygaard, G.D. Thurmond, and S.D. Fouse, "A CCD Image Processor for Smart Sensor Application", Proc of the SPIE 155, 15 (1978).

Nudd, G.R., "Image Understanding Architectures" in Proc. of the National Computer Conference, p. 377 (1980).

Okuyama, Y. and T-P. Ma, "High Efficiency Inversion Layer Solar Cells based on Ionizing Radiation - Induced Surface Inversion" to be

published in Appl. Phys. Lett. 44(6), 1984.

Patrin, N.A., "Performance of Very High Density Charge-Coupled Devices", IBM J. Res. Develop. 17, 241 (1973).

Preston, Jr., K., M.J.B. Duff, S. Levialdi, P.E. Norgren, and J-I. Toriwaki, "Basis of Cellular Logic with some Applications in Medical Image Processing", Proc. IEEE 67, 826 (1979).

Preston, K., "Cellular Logic Computers for Pattern Recognition", IEEE Computer 16(1), 36 (1983).

Rabbani, K.S. and D.R. Lamb, "Direct Correlation of Generation Lifetimes Obtained from Pulsed MOS Capacitors and Gated Diode Measurements", Solid-State Electron. 26, 161

Ravi, K.V. and C.J. Varker, "Oxidation-Induced Stacking Faults in Silicon. I. Nucleation Phenomenon", J. Appl. Phys. 45, 263 (1974).

Reddaway, S.F., "The DAP Approach" in Infotech State of the Art Report on Supercomputers, 2, 309 (1979).

Rosenfeld, A. and A.C. Kak, Digital Picture Processing, Academic Press, New York (1976).

Sah, C-T., R.N. Noyce, and W. Shockley, "Carrier Generation and Recombination in PN Junctions and PN Junction Characteristics", Proc. IEEE. 45, 1228 (1957).

Schaefer, D.H. and J.P. Strong, "Tse Computers", Proc. IEEE 65, 129 (1977).

Schaefer, D.H. and J.R. Fisher, "Beyond the Supercomputer", IEEE Spectrum, Mar. 1982, p. 32.

Schaefer, D., "Spatially Parallel Architectures: An Overview", Computer Design, Aug. 1982, p. 117.

Schiff, L.I., Quantum Mechanics (3rd Ed.), McGraw-Hill Book Company, New York (1968).

Schmidt, P.F., "Furnace Contamination and its Remedies", Solid-State Tech. 26(6), 147 (1983).

Schroder, D.K. and H.C. Nathanson, "On the Separation of Bulk and Surface Components of Lifetime using the Pulsed MOS Capacitor", Solid-State Electron. 13, 577 (1970).

Sequin, C.H., D.A. Sealer, F.J. Morris, R.R. Buckley, W.J. Bertram, and M.F. Tompsett, "Charge-Coupled Image Sensing Devices using Three Levels of Polysilicon", 1974 Int'l. Solid-State Circuits Conference Digest of Tech. Papers, p. 24, 25, 218, Philadelphia, PA, Feb. 1974.

Sequin, C.H. and M.F. Tompsett, Charge Transfer Devices, Academic Press, Inc., New York (1975).

Shockley, W. and W.T. Read, "Statistics of the Recombination of Holes and Electrons", Phys. Rev. 87, 835 (1952).

Shockley, W. "Problems Related to PN Junctions in Silicon", Solid-State Electron 2, 35 (1961).

Slotnick, D.L., W.C. Borck, and R.C. McReynolds, "The SOLOMON Computer", Proc. Western Joint Computer Conf., (1962), p. 87.

Sternberg, S.R., "Biomedical Image Processing", IEEE Computer 16(1), 22 (1983).

Streetman, B.G., Solid State Electronic Devices, Prentice-Hall, New Jersey, (1972).

Suzuki, N. and H. Yanai, "Computer Analysis of Surface-Charge Transport between Transfer Electrodes in a Charge-Coupled Device", IEEE Trans. Electron Devices 21, 73 (1974).

Swaroop, R.B., "Effects of Defect Types on Bipolar Product Yield", Solid-

State Tech. 27(1), 177 (1984).

Sze, S.M., Physics of Semiconductor Devices (2nd ed.), John Wiley and Sons, New York (1981).

Tanikawa, K. Y. Ito, and H. Sei, "Evaluation of Dark-Current Non-Uniformity in a Charge-Coupled Device", Appl. Phys. Lett. 28, 285 (1976).

Teng, C. W-H., "Threshold Control and Signal Amplification of the Metal/Tunnel-Oxide/Semiconductor (MTOS) Junction", Ph.D. Thesis, Yale University (1983).

Tompsett, M.F., "The Quantitative Effects of Interface States on the Performance of Charge-Coupled Devices", IEEE Trans. Electron Devices 20, 45 (1973).

Tompsett, M.F., "Surface Potential Equilibration Method of Setting Charge in Charge-Coupled Devices", IEEE Trans. Electron Devices 22, 305 (1975).

Unter, T.F., P.C.T. Roberts, D.R. Lamb, "Correlation of Pulsed MOS Capacitor Measurements with Oxidation Induced Defects", Electron Lett. 13, 93 (1977).

Varker, C.J. and K.V. Ravi, "Oxidation-Induced Stacking Faults in Silicon. II. Electrical Effects in PN Diodes", J. Appl. Phys. 45, 272 (1974).

Wagner, S. and C.N. Berglund, "A Simplified Graphical Evaluation of High-Frequency and Quasistatic Capacitance-Voltage Curves", Rev. Sci. Instrum. 43, 12 (1972).

Wen, et. al. in Digest of Technical Papers of Int'l. Solid-State Circuits Conference, Philadelphia, PA (1975) p. 24.

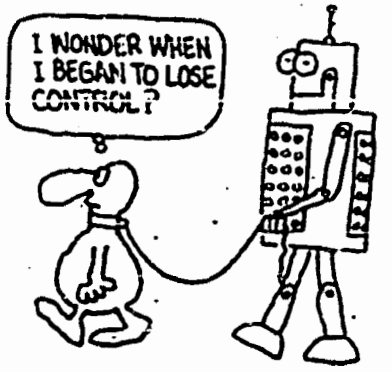
Wolff, P.A., Phys. Rev. 95, 1415 (1954).

Wong, V.S., G.R. Nudd, P.A. Nygaard, P. Grieff, D. Boxenhorn, and B.

Dew, "A CCD Focal Plane Star Tracker" (unpublished)

Young, D.R. and C.M. Osburn, "Minority Carrier Generation Studies in MOS Capacitors on N-Type Silicon", J. Electrochem. Soc. 120, 1578 (1973).

Zerbst, M., Z. Angew. Phys. 22, 30 (1966).



By Mal