A 2X2 ARRAY ANALOG MEMORY IMPLEMENTED WITH A SPECIAL LAYOUT INJECTOR

By

YONG-YOONG CHAI

Bachelor of Electronic Engineering
SoGang University
Seoul, Korea
1985

Master of Electrical Engineering
Oklahoma State University
Stillwater, Oklahoma
1991

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1994

A 2X2 ARRAY ANALOG MEMORY IMPLEMENTED

WITH A SPECIAL LAYOUT INJECTOR

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

PREFACE

This thesis attempts to realize the analog memory implemented with a special shaped injector, and propose the 8 bit programmable analog-to-digital converter. The building blocks were laid out using MAGIC and fabricated using MOSIS service.

I wish to express my deepest appreciation to my advisor, Dr. Louis. G. Johnson. His guidance throughout this research has been invaluable. The helpful advice of the other committee members, Dr. Chriswell Hutchens, Dr. Keith A. Teague, and Dr. John P. Chandler are also gratefully appreciated.

I greatly appreciated the School of Electrical Engineering for the financial support during the study, and MOSIS fabrication services for fabricating my circuits.

Finally, I would also like to thank my parents, Nam-Young Chai and Chun-Ja Lee; my loving wife, Eun-Kyung Chai; proud daughters, Hye-Won and Hye-Min. Their countless sacrifices and prayers have helped me tremendously.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

NOMENCLATURE

J	Current Density
E	Electric Field
α	Characteristic Constant
β	Characteristic Constant
t_{ox}	Oxide Thickness
K_w	Writing Voltage Coupling Ratio
K_e	Erasing Voltage Coupling Ratio
Q_{float}	Floating Gate Voltage
C_{pp}	Interpoly Capacitance between Control Gate and Floating Gate
C_{inj}	Interpoly Capacitance between Injector and Floating Gate
C_g	Gate Capacitance under the Channel of the Measuring Tr.
C_{ox}	Capacitance between Floating Gate and Substrate
V_{pp}	Voltage on Control Gate of EEPROM
V_{inj}	Voltage on Injector of EEPROM
V_{tun}	Voltage Magnitude between Floating Gate and Injector
V_t	Threshold Voltage
V_{ti}	Initial Threshold Voltage
V_{tw}	Threshold Voltage during Writing Operation
V_{te}	Threshold Voltage during Erasing Operation
V_γ	Minimum Voltage Required to charge injection
C_t	Total Capacitance
V_{ds}	Drain to Source Voltage of Transistor

V_{gs}	Gate to Source Voltage of Transistor
I_{ds}	Drain to Source Current of Transistor
Φ_B	Energy Barrier
t	Elapsed Time
ν	Dielectric Relaxation Frequency
k	Boltzman's Constant
T	Temperature

CHAPTER I

INTRODUCTION

Analog Memory

Digital memories have been widely used as a device for storing information due to its reliable, fast, and relatively simple control circuitry. However, towards the end of the decade, the storage capability of the digital memory will be limited by not scaling to smaller linewidths. One way to increase the storage capability of the memory dramatically is to change the type of stored data from digital to analog. In the conventional digital memory several bits are required to represent just one state, whereas the analog memory requires just one cell to characterize the state. In addition to the enhancement of the storage density, the analog memory also contribute to simplifying a typical signal processing system implemented with peripheral devices such as A/D and/or D/A converter.

Recently, some sources [Shimabukuro, 1988][Sackinger, 1988][Ong, 1989][Carley, 1989][Blyth, 1991] have reported that an EEPROM(Electrically Erasable Programmable Read Only Memory) can be used as an analog memory by using its strong non-volatile characteristics. Since there are a lot of applications in this area of the EEPROM including the simple array memory and weight control of synapse in a neural network, there has been a lot of research concerned with the theoretical and experimental non-volatile memory for three

decades. However, some problems such as large device size and high programming voltage in the EEPROM have been major obstacles to advance the technologies in this area. Especially the high voltage required for programming of the EEPROM device leads to failure mechanisms associated with junction breakdown. Many papers[Nozawa, 1986][Lee, 1991][Thomsen, 1991][Durfee, 1992][Sethi, 1992][Montalvo, 1993] concerning about reducing the programming voltage have been published. However, still the voltage is high enough to damage the chip circuitry.

The first objective of this paper is to introduce a special method to reduce the above problems by just simple modification of the layout of the conventional EEPROM. Especially, our trials focus on developing a floating gate MOSFET operated with a lower programming voltage than the conventional EEPROM. The second objective of this paper is to introduce a 2 by 2 analog signal storing device which is expandable to any size memory. The analog array implemented with the above EEPROM cell has been designed to reduce the programming voltage, to enhance the resolution of stored signal, and to simplify the control circuits. Finally an A/D converter implemented with the floating gate MOSFET is presented. Compared to the conventional A/D converters, it consumes less power and increases the resolution.

Chapter Description

In Chapter II, a literature survey is presented which covers different types of non-volatile floating gate MOSFET and analog memories.

In Chapter III, some new proposed floating gate MOSFET layouts are presented. The various injector shapes and the efficiency of each shape are described.

In Chapter IV, the system block diagram and each function of the 2 by 2 analog memory are given.

in Chapter V, an innovative A/D converter implemented with the floating gate MOSFET is proposed.

In Chapter VI, a summary and conclusion is given.

CHAPTER II

LITERATURE SURVEY

Introduction

This chapter describes short and long term analog memories. Especially, we focus on the long term analog memories implemented with floating gate MOSFETs. In addition, the several potentials of the floating gate MOSFETs are introduced. Finally, the basic principles of the floating gate MOSFETs are explained in this chapter.

Analog Memory Paradigm

The memory can be classified into two groups according to a storage time capability: short term and long term memory.

Short Term Analog Memory

Figure 1 shows the sample and hold circuit which temporarily holds charges on the parasitic capacitance. The basic idea of the short term storage device utilizing a parasitic capacitance on a MOS device is very simple; during the sampling process, Vg is high and the capacitor C is charged/discharged with a specific number of electrons. During the subsequent hold process, the charge on the capacitance C will not change with Vg being low. The retention

time of the charge depends on the parasitic resistance and capacitance of the design. In terms of storing desity, this device should be chosen primarily. However, the consistency of this device is very bad due to the irregularity of the parasitic capacitance and resistance. Besides the consistency problem, the retention time of this device is so small that the circuit can not keep the initially stored value during the hold process.

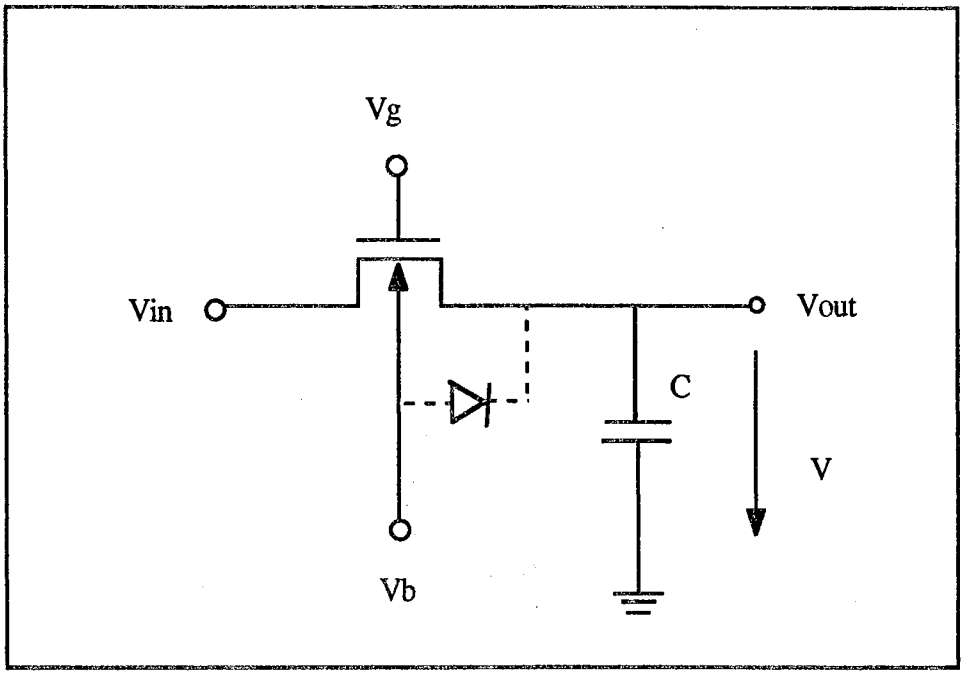


Figure 1. A Sample and Hold Circuit Representing a Simple Analog Memory

The above circuit is replaced with a new version of sample and hold circuit to improve the resolution and the retention property of an analog memory. Figure 2 shows the circuit which consists of two differential amplifiers, a switch, and a MOSFET. The operation of this memory is as follows; during sampling, the capacitance C shown in Figure 2 is charged according to the difference ΔV_1 between V_{in} and V_{out} with V_G high and the switch S_2 being

off. The sampled charge is retained during the subsequent holding process which is activated by switch S_2 being on and V_G being off.

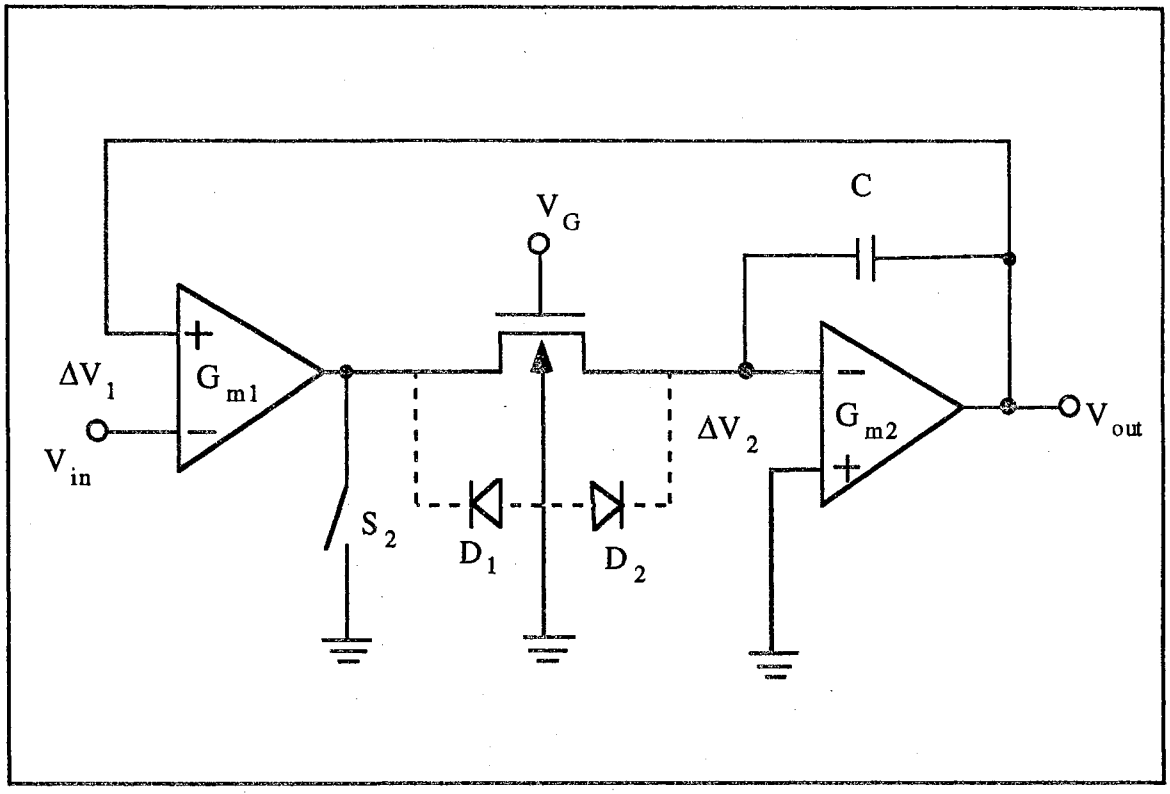


Figure 2. Improved Version of Sample and Hold Circuit in terms of retention time

Then, the V_{out} is equal to the offset voltage of the transconductance G_{m2} due to the virtual ground action of G_{m2}. Thus, the fact that the voltage difference between both nodes of capacitance C is negligible induces the sample and hold circuit to the lower leakage charge storage device than the previous one.

The accuracy and the retention capability of this sample and hold circuit have been proven to be better than the previous circuit. But the accuracy is

deteriorated by the offset voltage of G_{m2}. Additionally, it is a very area consuming technology.

Long Term Analog Memory

A long term storage device, non-volatile memory such as EPROM(Electrically programmable read only memory) and EEPROM has been investigated in many different areas due to its strong holding and trimming properties. Even though the attractiveness of the characteristics on the nonvolatile device is reduced by the limitation to a finite number of programming, variability, and high programming voltage, its long term storage and the high density properties have made it to be the most valuable candidate of a neural network system.

In 1967, Khang and Sze[Khang, 1967] first proposed a floating-gate metal-oxide-semiconductor(FAMOS) shown in Figure 3 as a nonvolatile memory element. In this structure, the charge is transported from the silicon substrate across an insulator layer to a floating metal electrode. The injection into the floating gate is accomplished by a vertical electric field of electrons with excess energy acquired from a high source to drain channel electric field. The hot electrons with sufficient energy will conduct across the oxide barrier and charge the gate. The injection is called a hot electron injection. This device is known as having the first EPROM structure. The typical oxide thickness of FAMOS is so thick that the retention characteristics of this device is excellent. The drawback of this device is that the electron removal from the floating gate is not controllable in this structure. Therefore, to erase the pattern in the memory, the chip must be removed from the circuit and exposed to the UV light.

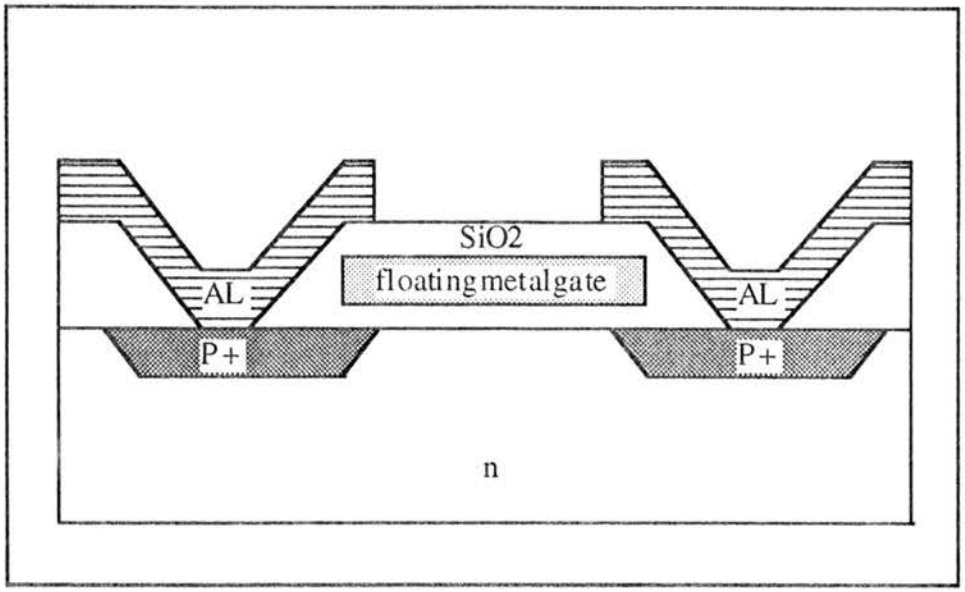


Figure 3. Cross Section of FAMOS Structure

In 1969, Wallmark et al. [Wallmark, 1969] proposed the Metal-nitride-oxide-semiconductor(MNOS) structure shown in Figure 4. The MNOS uses silicon nitride instead of the floating metal gate on the FAMOS. Like the EPROM, it also uses hot electron injection for programming. However, unlike the EPROM, the MNOS has a capability of speedy in-system erasure with a strong electric field due to the fact that the electrode in this structure is electrically accessible. Instead of the hot electron injection during the programming, the electron removal from the floating gate in this device is based on a Fowler-Nordheim tunneling[Lenzlinger, 1969]; a quantum mechanical process in which electrons tunnel through a gate oxide from a conducting channel to a floating gate. Based on the Fowler-Nordheim tunneling, the high electric field will reduce the energy barrier preventing electrons in the silicon or polysilicon from entering SiO_2. As a result, the high programming voltage(around 25V) would be required to move electrons from normal gate-oxide of thickness (40nm) [Carley, 1989]. Hence, the gate oxide layer should be very thin to reduce programming voltage.

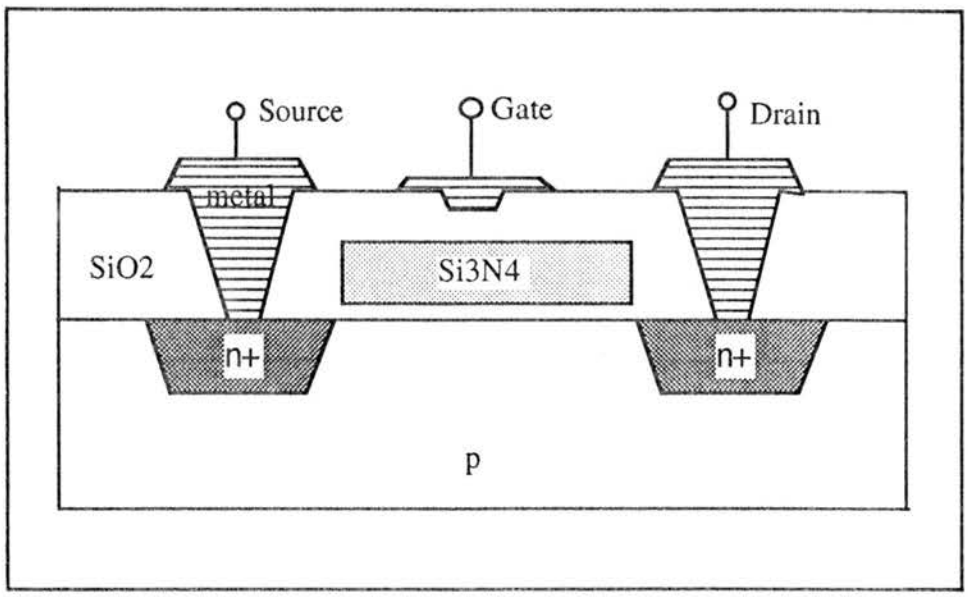


Figure 4. Cross Section of MNOS Structure

The major drawback of this device is that the threshold voltage in this structure is vulnerable to disturbance by even small applied voltage. The data disturbance and the poor retention time of the MNOS due to its thin oxide have prevented the memory from being used widely[Haznedar, 1991].

In 1980, Johnson et al.[Johnson, 1980] developed the first EEPROM(electrically erasable programmable read only memory) device, FLOTOX. The FLOTOX technology uses polysilicon as a floating gate. The general device structure is shown in Figure 5.

While the MNOS uses the Fowler-Nordheim tunneling for only discharging, the FLOTOX employs the same mechanism to increase or decrease charge stored on the floating gate. In order to realize the bidirectional operation based on the mechanism with a reasonable magnitude of programming voltage, the very thin tunnel oxide between the floating gate and the drain shown in Figure 5 is indispensable. The typical tunnel oxide thickness of this device is 100 Å.

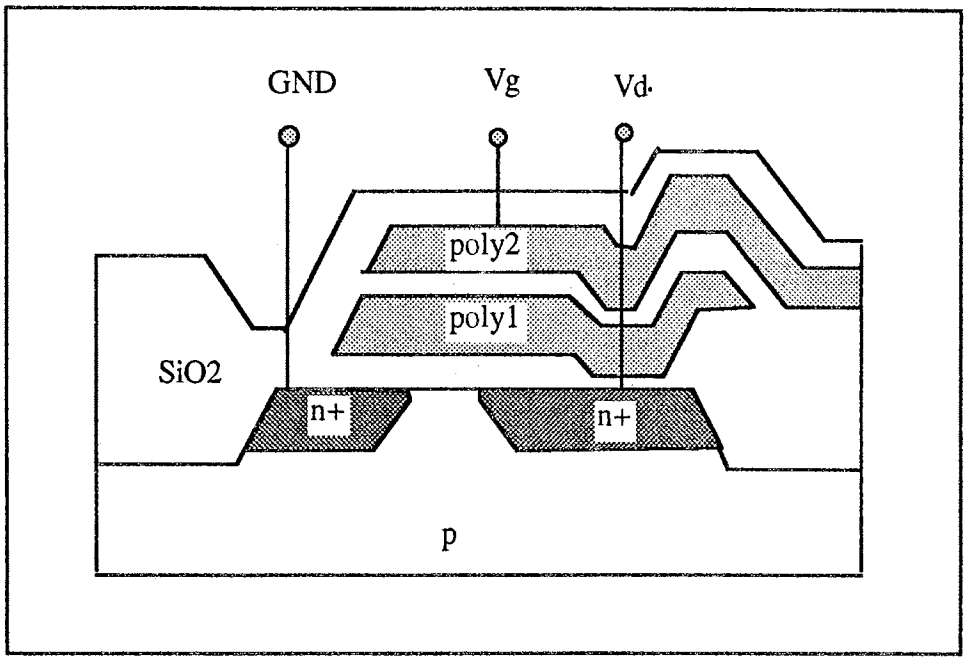


Figure 5. Cross Section of FLOTOX Structure

The thin oxide area must be relatively small to maximize the tunneling injection at a given tunneling oxide thickness and voltage level.

The FLOTOX's programming has been proven to be well-controlled and stable[Mielke, 1987]. However, as mentioned earlier, the Fowler-Nordheim tunneling mechanism basically requires a high programming voltage. The way to reduce the voltage is to decrease the thin oxide thickness. However, the reduction of the oxide thickness deteriorates the retention time as well as yield. The devices utilizing the Fowler-Nordheim tunneling mechanism cannot avoid these problems. The requirement of the high programming voltage implies high electric fields which are close to a junction breakdown. Thus, the breakdown leading to data loss is severe in this structure. Worse than that, the fabrication process for the FLOTOX must be special due to the tiny thin tunneling oxide.

In 1980, Kupec et al.[Kupec, 1980] suggested a textured-polysilicon EEPROM shown in Figure 6. It uses bump structure for tunneling instead of the

thin oxide like FLOTOX. Note that the structure enhances electric field distinctively[Carley, 1989]. Thus the corner placed in the direction of the electron flow can strengthen the electric field for the tunneling injection at a given programming condition. In order to enhance the programming operation bidirectionally with the corners, two different corners are required. In Figure 6, the upper-left corner of the 1st level polysilicon intensifies electric field, and accelerates a charge injection onto the floating gate. To the contrary, electrons get out of the floating gate mainly from the polysilicon level 2 through the lower-left textured layer to the polysilicon level 3.

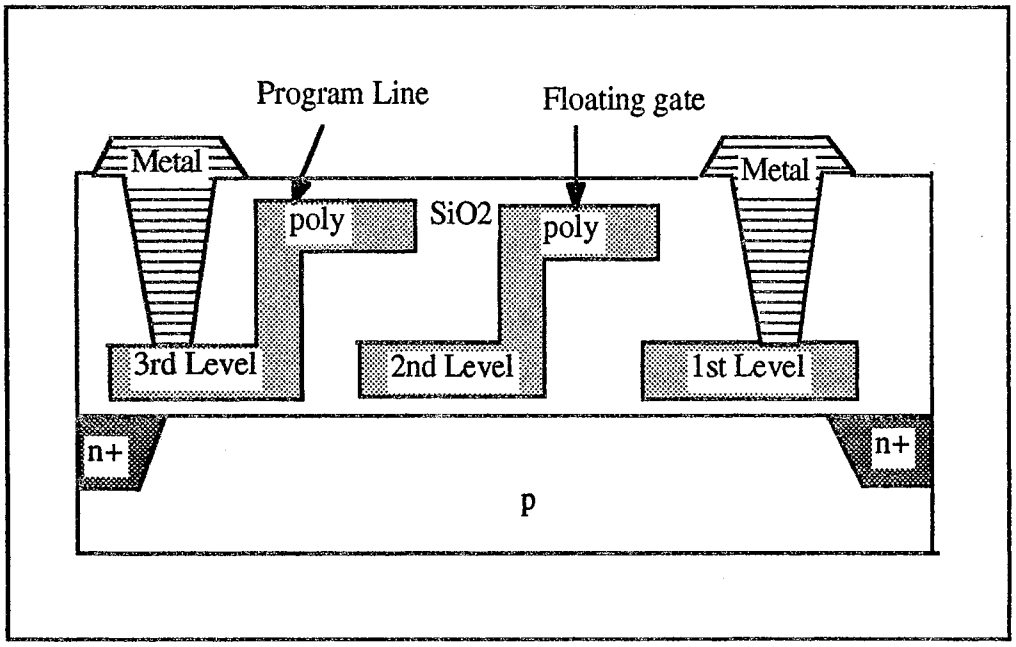


Figure 6. Cross Section of Textured Polysilicon Structure

The advantage of this structure is that it reduces the programming voltage without reducing the thickness of oxide. It is supposed to increase the retention time compared to FLOTOX, and decrease the breakdown effect which is serious

in the FLOTOX structure. But the device requires a special fabrication technology. Additionally, a window closing effect is severe in this structure[Mielke, 1987].

In 1989, Carly[Carley, 1989] introduced a special shape of an EEPROM device shown in Figure 7. Like the textured polysilicon, it relies on a rectangular corner to enhance the local electric field. But it does not require a special fabrication process since this is the same structure of a standard transistor except the drain and the source being connected together. The chip reliability based on the concept is relatively higher than the previous ones, and the removal of the thin oxide from the layout will improve the retention time. However, a problem still exists in this approach. It requires very high programming voltage to induce the proper charge transfer. Therefore, we have to be cautious about the junction breakdown in this layout.

In 1991, Lee et al.[Lee, 1991] introduced a new EEPROM structure which overcomes the above problems. They used a tunneling injection at a polysilicon 1 with polysilicon 2 instead of polysilicon 1 with diffusion layer. It has been previously known that the thermally grown SiO_2 films on amorphous-deposited surface of the polysilicon exhibit surface asperity[Anderson, 1977]. The surface roughness of the polysilicon-SiO_2 interface leads to localized electric field being greater than the average applied field. Therefore, it is clear that the significant emission at the moderate applied field is achieved in this structure with a standard fabrication process. Hence their idea reduces the programming voltage drastically.

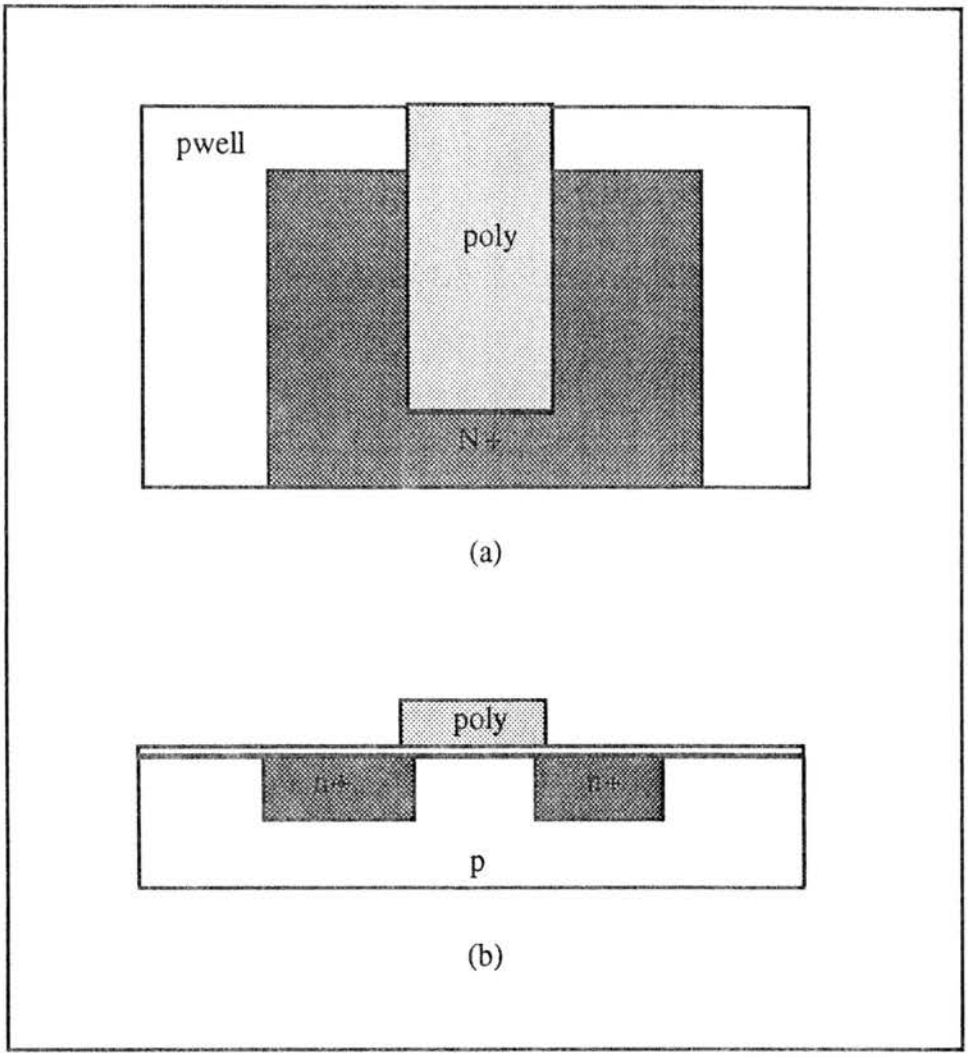


Figure 7. (a) Top View and (b) Cross Section of Carley's EEPROM[Carley, 1989]

In addition, the simple structure of the storing device makes the fabrication process just like conventional CMOS, thereby improving the yield considerably. Moreover, the retention time of the EEPROM is also proved to be better than the FLOTOX due to the thickness of the oxide[Thomsen, 1991]. Fig. 8(a) and (b) shows the top and cross section view of the poly-poly tunneling suggested by Thomsen[Thomsen, 1991].

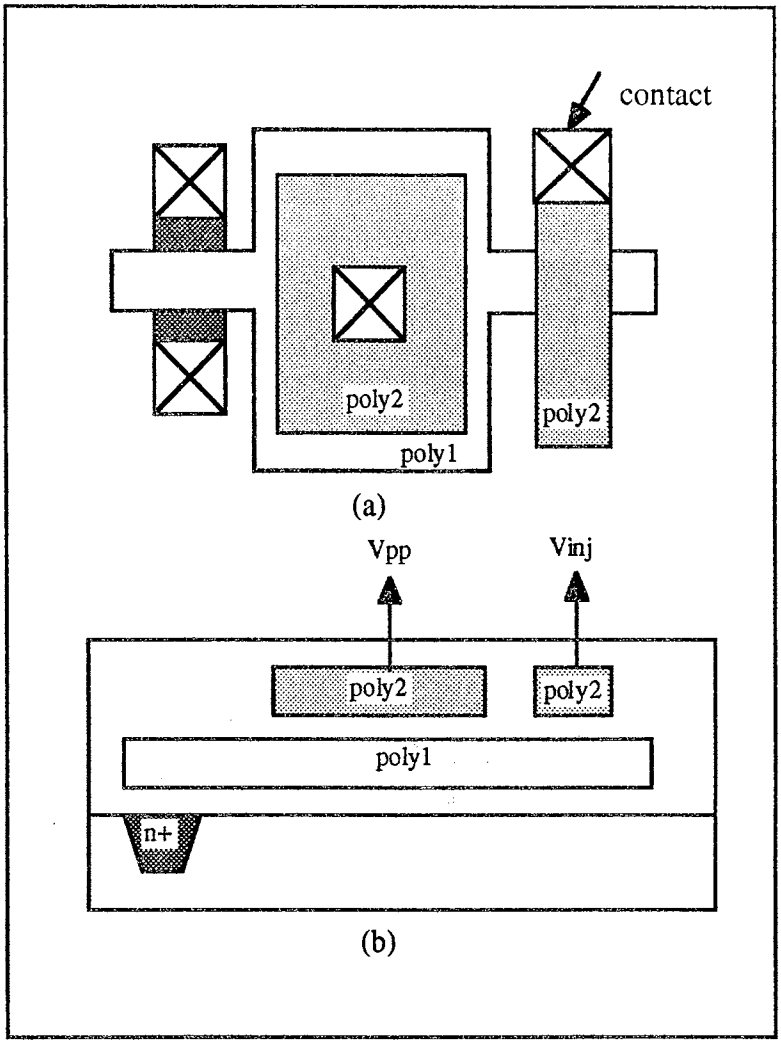


Figure 8. (a) Top View (b) Cross Section of Thomsen's EEPROM[Thomsen, 1991]

Potentials of an Analog Memory

The analog memory now is extensively used as synaptic weight control on a neural network, an adaptive filter, an A/D or D/A converter, and in an offset voltage adjustable circuit of an operational amplifier, etc. due to its interesting characteristics such as continuous values and trimmability.

The MOS analog memory implemented with the long term non-volatile characteristic shown in Figure 9 was provided by Shimabukuro, et al [Shimabukuro, 1988] in 1988.

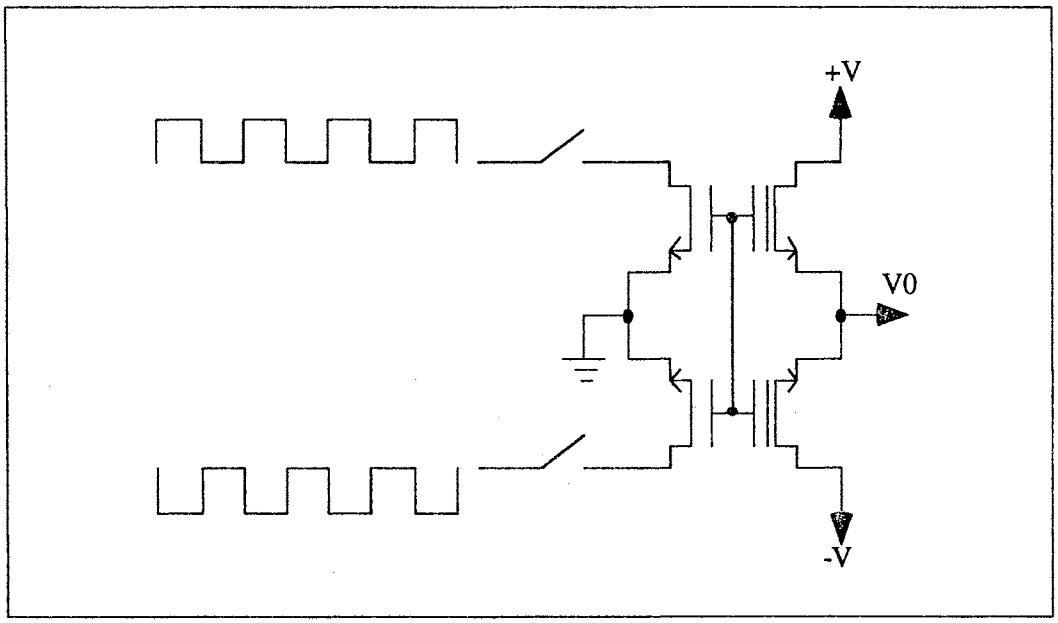


Figure 9. Schematic Diagram of MOS Analog Memory Cell

The cell which is implemented with floating gate MOSFETs using a hot electron injection mechanism was proposed as a device for storing the information of each interconnection weight between processing units in neural networks. The programming is accomplished with positive pulses to the drain of n-channel

transistor or negative pulses on the drain of p-channel transistor to inject holes or electrons, respectively.

In 1988, Sackinger, E. et al. [Sackinger, 1988] presented four different circuits implemented with floating gate MOSFETs. One of them is shown in Figure 10 where the trimming circuit is used for compensating between the two tail current Ics (F1) and between the two differential pairs(F2).

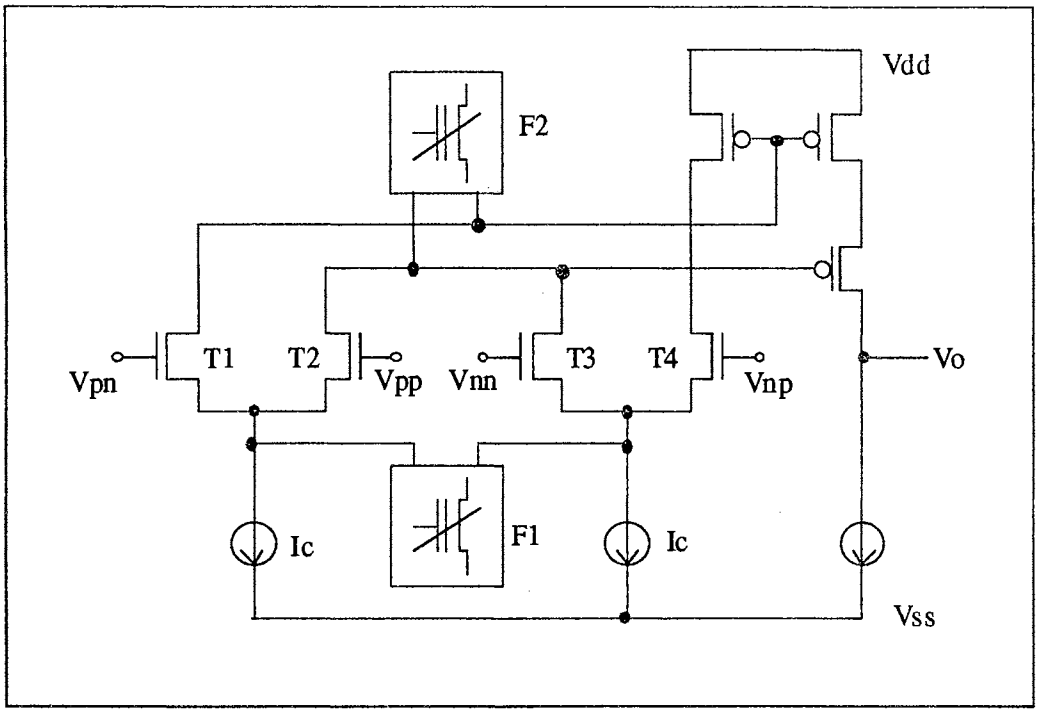


Figure 10. Double Differential Amplifier Implemented with Two Floating Gate MOSFET for Improving a Linearity and Reducing an Input Offset Voltage

They introduced a simple and cost-effective iterative trimming technique for the fast programming, where the programming voltage magnitude is roughly adjusted depending on the difference between desired and current voltage on the floating gate; if the error reduction is below 20 percent or more than 50

percent, the pulse voltage is increased by 1V or decreased by 1V, respectively. Otherwise, the magnitude of the programming voltage would not be changed.

In 1989, Carley, L. R. [Carley, 1989] showed the adjustment of the input offset voltage of MOS differential pairs with an EEPROM device utilizing a local enhancement technique. The input offset voltage being caused by transistor mismatches deteriorates an operational amplifier. An auto zero technique has been used for eliminating the problem, but the technique has a limitation due to clock feedthrough. Meanwhile, the offset adjustment method with the floating gate MOSFET is not only accurate, but also requires a very small area.

The input offset control is achieved by trimming the floating gate voltage of the transistor M5 shown in Figure 11. To initiate the trimming cycle, charging onto the floating gate is performed first no matter what the initial state of the floating gate voltage is. It will simplify the programming algorithm. After the cycle, a negative programming voltage will be applied to the control gate of the floating gate in a two-cycle process. During $\varnothing_1$ clock phase, the input of the op amp is shorted, and at the end of $\varnothing_1$, the sign of the op-amp output is latched by the comparator. During $\varnothing_2$, the electron stored in the floating gate will be removed through an injector according to the state of the sign. The above two phase nonoverlapping trimming operation continues to be performed until a precise analog voltage is established on the floating gate.

Sachinger[Sackinger, 1988] also suggested the offset control circuit with a ultra thin gate oxide floating gate MOSFET. But, due to its inherent characteristics, the charge storing capability of the floating gate MOSFET is relatively worse than that by Carley. In addition, the floating gate MOSFET in Carley's circuit does not claim any special fabrication process.

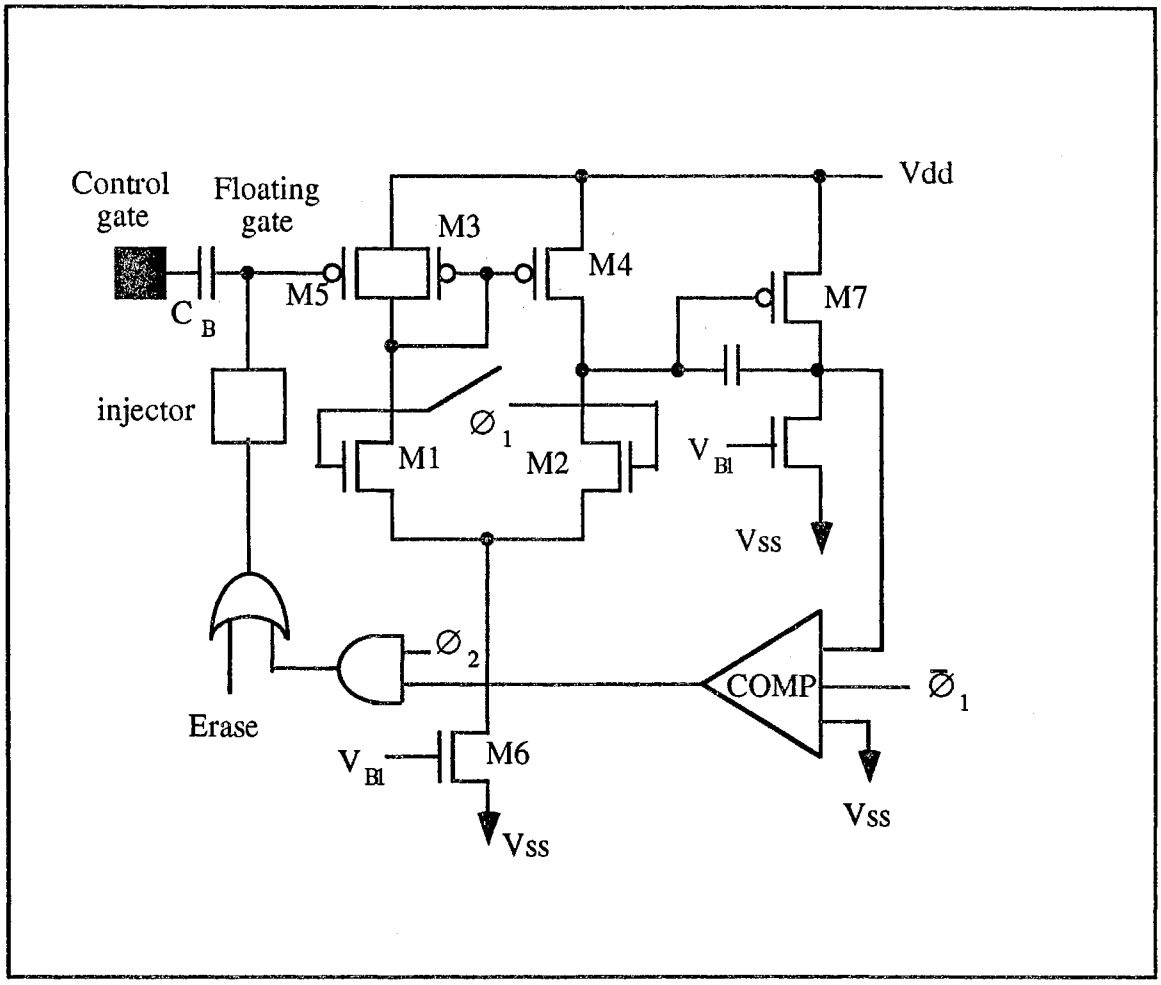


Figure 11. Schematic Diagram of a Trimming Circuit eliminating Input Offset Voltage of an Differential Amplifier

Finally, an insignificant amount of injection during programming will improve the trimming operation, leading to excellent resolution. But, unfortunately it must deteriorate programming time.

In 1989, Ong, T. C. et al. [Ong, 1989] provided a potential analog memory device. He demonstrated that the analog memory system could store various shapes of the signals properly with an insignificant number of errors. The way to store the analog signals into an array is first to sample the signals and program those into successive bits of EEPROM memory array. During the programming,

the programming voltage pulse duration is constant, but the magnitude of the voltage is varied according to the sampled signal. The adaptive operation reduces the programming time.

In 1991, Blyth, T. et al.[Blyth, 1991] invented an analog storage array device using an EEPROM technology. The analog device is capable of reproducing telephone quality voice.

As a programming mode, Blyth used a closed loop; Once an analog signal is inserted into the system, the signal is sampled and held until the voltage at an interesting intersection reaches the input analog voltage within a previously defined range. A feedback circuit which is comprised of a comparator and a sample-and-hold circuit will be used to achieve the purpose. The feedback circuit will contribute to compensating various individual cell characteristics, and the sample-and-hold circuit makes it possible to follow the continuous input signals without interruption from the EEPROM programming operation. The drawback of this analog array is that the system begin the programming operation only after an erasing operation ends. While the algorithm simplifies a programming controller, the unnecessary operation accelerates the window closing effect of an EEPROM device. Additionally, it consumes a programming elapsed time.

Basic Analysis of a Floating Gate MOSFET

The simplified capacitive equivalent circuit of EEPROM device, illustrated in Figure 8, utilizing the Fowler-Nordheim tunneling injection for both writing and erasing is shown in Figure 12.

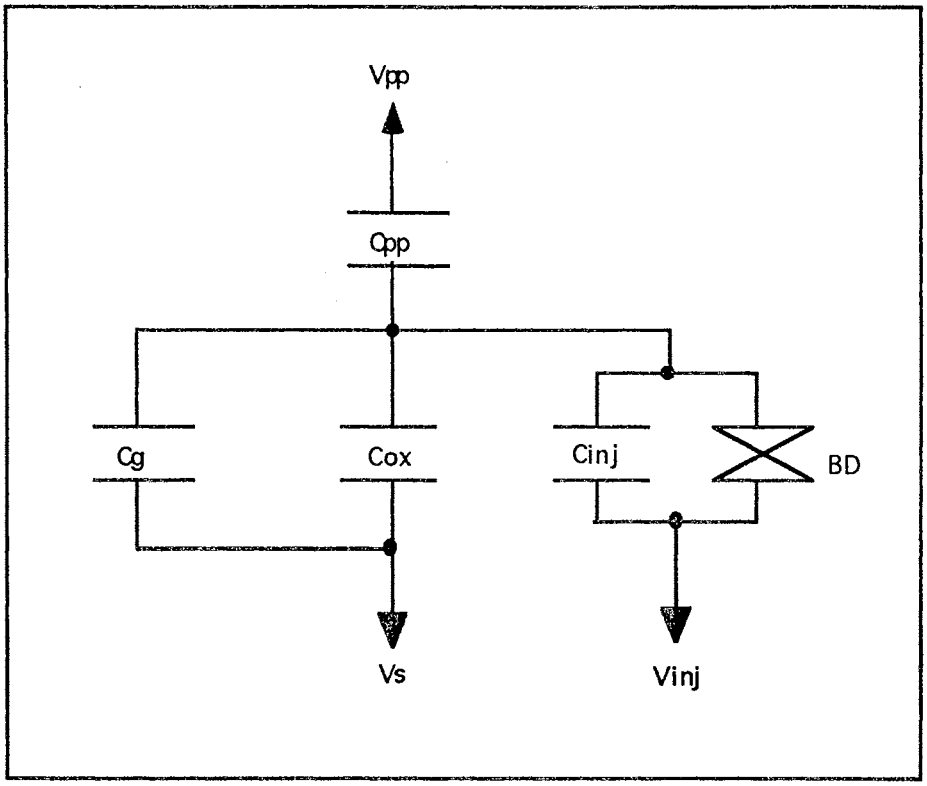


Figure 12. Simplified Model of EEPROM Device

Here, C_{pp} is the interpoly capacitance between the control gate and the floating gate, C_{inj} is the different interpoly capacitance between the injector and the floating gate, C_g is the gate capacitance under the channel of a measuring transistor, and C_{ox} is the capacitance between the floating gate and the substrate. The BD shown in Figure 12 indicates a bi-directional diode representing the Fowler-Nordheim tunneling mechanism. The current density through this diode can be expressed as

$$J = \alpha E^2 \exp(-\beta/E) \qquad (2\text{-}1)$$

where α and β are characteristic constants which can be determined from experimental data, and E is the electric field.

This equation implies that charge injection is exponentially dependent on the electric field. The electric field E is related to the absolute voltage across the tunnel oxide as

$$E = \frac{V_{tun}}{t_{ox}} \tag{2-2}$$

Here, t_{ox} and V_{tun} indicate the gate oxide thickness and the voltage across the tunnel oxide. The equation, (2-1) is valid when V_{tun} is greater than V_{γ}. Here, V_{γ} is defined as the minimum required voltage initiating the charge injection.

Consequently, the tunneling occurrence through BD depends on the voltage across gate oxide between the floating gate and the injector on the simplified capacitance model. Note that there are two different operations in the programming: writing and erasing. The writing is defined an operation, which tunnels electrons into a floating gate through the thin oxide. On the other hand, the erasing is defined an operation, which emits the electrons into the injector. The writing operation is accomplished by applying high programming voltage on the control gate and connecting the injector to the ground. For the erasing operation, the programming voltage is applied to the injector with the control gate being grounded. Thus by using the above two operations, it is possible to alter the weight stored in the cell. In the remainder of this paper, the verb "to program" will be used to indicate both the erase and the write operation.

Now the voltage across the gate oxide, V_{tun}, is calculated as a function of applied voltage on a control gate and an injector [Kolodny, 1983]. For given

external voltages, the programming voltage for the writing operation is determined by the voltage coupling ratio($K_w = C_{pp}/C_t$) and internal charges

$$|V_{tun}| = K_w V_{pp} + \frac{Q_{float}}{C_t} \tag{2-3}$$

Here, Q_{float} and C_t indicates the stored charge on the floating gate and the total capacitance($C_t = C_{pp} + C_{inj} + C_{ox} + C_g$) of the floating gate, respectively. Meanwhile, the programming voltage for the erasing operation is

$$|V_{tun}| = K_e V_{inj} - \frac{Q_{float}}{C_t} \tag{2-4}$$

Here, the programming operation for the erasing operation is decided by the voltage coupling ratio($K_e = (C_{pp} + C_{ox} + C_g)/C_t$) and internal charges. By using eq. (2-1) through (2-4), we can estimate the amount of tunneling current with the area of the injector. Note that the floating gate charge Q_{float} will vary during programming. As a result, the induced charges through the tunneling injector will become a nonlinear function of the programming voltage. This is the reason why a trimming circuit cannot adjust a weight of EEPROM device by a predetermined number of pulses to reach a desired value. In order to achieve the accurate programming, a feedback circuit implementation is required.

Now take a look at the relationship between a threshold voltage and an applied voltage. During programming, an amount of charge Q_{float} in the floating gate will be altered. Subsequently, it will change the threshold voltage which can be defined as [Kolodny, 1983]

$$\Delta V_t = -\frac{Q_{float}}{C_{pp}} \tag{2-5}$$

Then after the programming, the threshold voltage of the floating gate is

$$V_{tw} = V_{ti} + V_{pp}[1 - \frac{V'_{tun}}{K_w * V_{pp}}] \quad (2\text{-}6)$$

$$V_{te} = V_{ti} + V_{inj}[\frac{K_e}{K_w} - \frac{V'_{tun}}{K_w * V_{inj}}] \quad (2\text{-}7)$$

Here, V_{ti} is the threshold voltage with Q_{float}=0, V'_{tun} is the tunnel oxide voltage after the end of the programming pulse, and V_{tw} and V_{te} are the threshold voltages of the written and the erased floating gate, respectively.

The threshold voltage also can be represented by functions of programming time as follows[Kolodny, 1983];

$$V_{tw}(t) = V_{ti} + V_{pp} - \frac{1}{K_w} * \frac{B}{\ln(A*B*t + E_1)} \quad (2\text{-}8)$$

$$V_{te}(t) = V_{ti} - V_{inj}\frac{K_e}{K_w} + \frac{1}{K_w} * \frac{B}{\ln(A*B*t + E_2)} \quad (2\text{-}9)$$

where

$$A = \frac{S*\alpha}{t_{ox}*(C_{pp} + C_{ox} + C_{inj} + C_g)} \quad (2\text{-}10)$$

$$B = \beta * t_{ox} \quad (2\text{-}11)$$

$$E_1 = \exp[\frac{B}{K_w*(V_{pp} + V_{ti} - V_t(0))}] \quad (2\text{-}12)$$

$$E_2 = \exp\left[\frac{1}{V_{inj}*K_e + K_w*V_t(0) + K_w*V_{ti}}\right] \quad (2\text{-}13)$$

Here, S is the area that the injection occurs.

From the result, we can see that the threshold voltage is logarithmically dependent on time. Additionally, it is seen that when the time t is less than τ which is defined as follows;

$$\tau = \frac{1}{AB}\exp\left[\frac{B}{K_w(V_g + V_{ti} - V_t(0))}\right] \quad (2\text{-}14)$$

then, the threshold voltage remains almost constant. The threshold voltage will change greatly after the time τ.

We have seen that the magnitude and the duration of a programming pulse are critical factors determining the amount of charge injection during the programming. We can hopefully design an analog memory which will be capable of memorizing a desired value in a finite period. Unfortunately, we cannot achieve the two goals at the same time due to its contradictory characteristics. For example, a large magnitude/duration programming pulse will reduce the programming time considerably so that it can come up with real time system. However, it will deteriorate the quality of analog memory very much. On the other hand, a small magnitude/duration programming pulse will increase programming time, but will result in a high resolution. Thus, it is application dependent. Hence, it is not obvious which programming pulse is best for an analog memory implementation.

CHAPTER III

AN INNOVATIVE FLOATING GATE MOSFET

Introduction

A floating gate MOSFET has been designed to maximize the tunneling current density injected through the oxide between two polysilicon layers by using local field enhancement provided from a bump structure. Compared to the floating gate device introduced by Thomson[Thomsen, 1991], the modified floating gate which is fabricated in the same standard 2μ double-polysilicon CMOS technology reduces the programming voltage by 2.4V and the erasing voltage by 1.4V without any trade-off.

Local Enhancement Layout

An EEPROM device relies on the Fowler-Nordheim tunneling through SiO_2 for injecting and/or removing carriers on a floating gate. However, it requires very high oxide fields(3.2V/5nm) to induce electron tunneling to the floating gate through the oxide[Carley, 1989]. It can lead to problems such as junction or oxide breakdown, and elevation of trapup effects.

One way to reduce the programming voltage is to decrease the thickness of the gate oxide. From the above eq.(2-2), we see that the thinner the oxide, the

greater the electric field. But, a thin oxide not only implies a special fabrication process, but also causes a deterioration of the retention time of the cell[Nozawa, 1982]. The other way to reduce the programming voltage without reducing retention time is to take advantage of local field enhancement around objects with sharp corners[Thomsen, 1991][Carley, 1989][Lee, 1991]. It is noted that the field enhancement factor has been observed at roughly 4 due to the field line around a rectangular corner[Carley, 1989][Lee, 1991]. Thus the corner placed in the direction of the electron flow can reduce the programming voltage. It is also well known that poly to poly for injection has more capability to strengthen the electric field than poly to Si for injection due to the fact that the asperity on the polysilicon surface enhances the local electric field[Carley, 1989]. The corner of poly2 directing toward poly1 for writing or of poly1 to poly2 for erasing can further strengthen the electric field. Hence, we can anticipate that a greater charge movement will occur between two layers at the corners of the layout. The advantage of this layout technique is that the local field enhancement can be achieved without a special fabrication process.

Including the local enhancement due to the rectangular corner, we can imagine that the intrinsic stress generated at the corners of poly steps during fabrication causes a local inhibition of oxidation. The curvature makes the length of the oxidation between polys shorter illustrated in Figure 13. Thus, the thinning of the oxide on the corner is also supposed to contribute to the field enhancement.

We propose a unique way to reduce the programming voltage even further than the EEPROM designed by Thomsen without any special fabrication process or expansion of the capacitance coupling ratio[Chai, 1993]. Our EEPROM cell introduces a different shape of bump structure providing the local enhancement for writing as well as for erasing. Especially, the tip of the bump

has been designed to be as sharp as possible with a standard CMOS digital fabrication process.

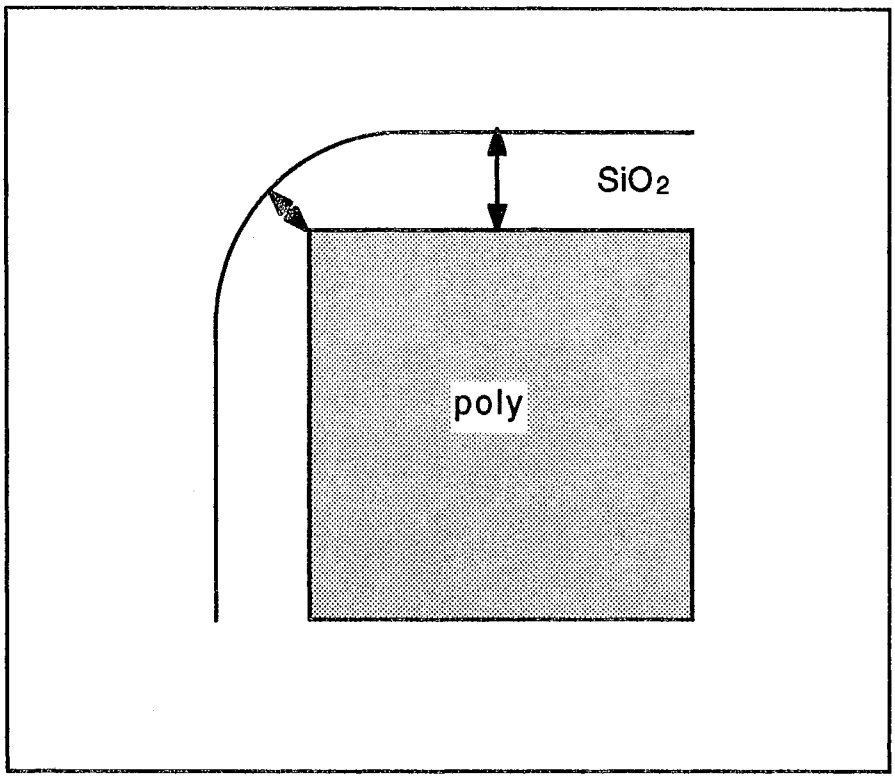


Figure 13. Thinning Oxide due to Curvature

Different Shape of Injector

In order to prove the local enhancement technique, we have tested 4 different chips containing 12 different types of injectors. The layouts are fabricated in a 2μ nwell(or pwell) double poly standard digital CMOS process available through in MOSIS. The oxide thickness between polysilicon is 75nm. The floating charge is measured indirectly from drain current of a MOS transistor attached to the floating gate. For each layer type, the physical shape of the current injector are characterized as follows;

Type 1: poly2 of the tunneling injector just crosses over the poly1 of the floating gate. This is exactly same as Thomsen's did[Thomsen, 1991].

Type 2-12: the geometrical shape of poly1 is changed to make number of bidirectional corners.

In type1, the primary injection for the writing is only due to asperity[Ellis, 1982], and the erasing injection is based on the asperity as well as the two edges on the top of poly1. The intense electric field from poly 1 to poly 2 in type 1 is shown in Figure 14, and the arrow indicates the relative strong intensity of electric field along the edge compared to the other sides during the erasing operation. In addition to the edges and the asperity, the intense electric field is provided by the number of corners on both directions in other types. Table I shows the parameters for the different floating gate structures. Figure 15 illustrates the several layouts from type 1 through type 12. Actually the shape of type 2 and 3 is exactly same as the type 11 and 12, respectively. But the size of the tip in type 2 and 3 is different from that in type 11 and 12. Thus Figure 17 contains totally 10 layouts, and Table I shows 12 different types.

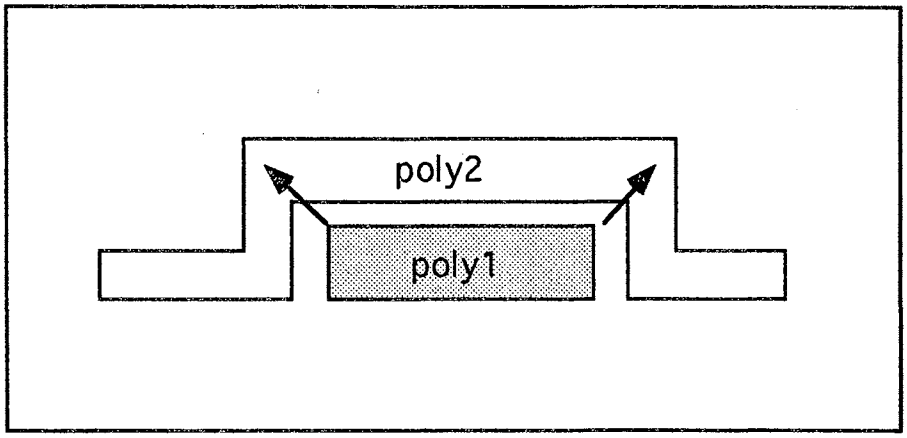
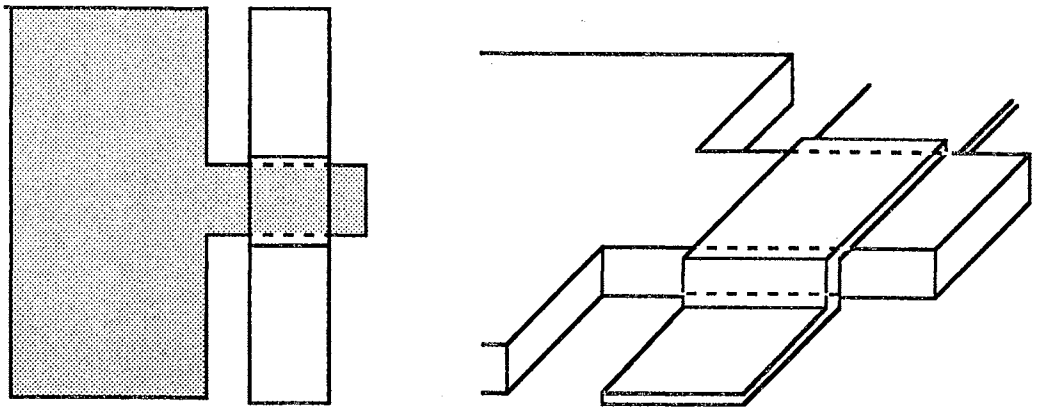


Figure 14. Side View of Type1 Floating Gate MOSFET

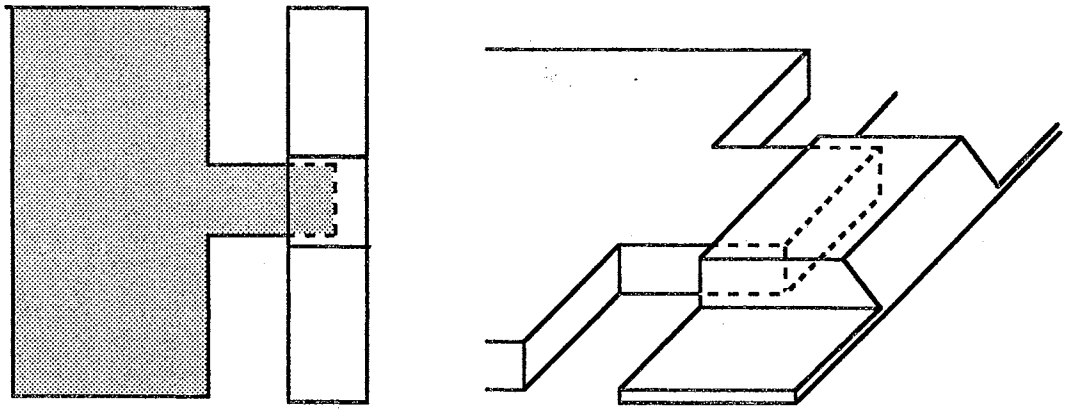
TABLE I

SEVERAL PARAMETERS RELATED TO
EACH TYPE FLOATING GATE MOSFET

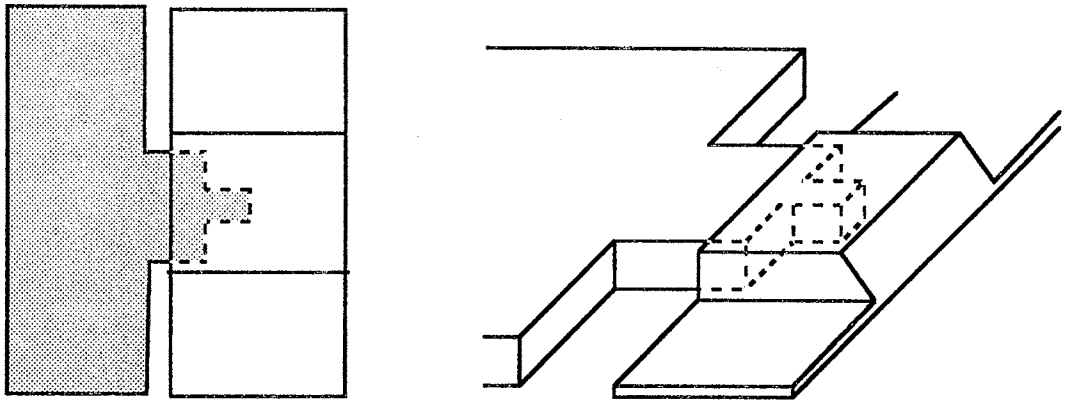
type	tip size (um)	edge length (um)	no of corner (write)	No. of corner (erase)	coupling ratio (write)
1	2	4	0	0	0.767
2	2	6	0	2	0.767
3	2	14	2	4	0.837
4	2	18	2	0	0.815
5	1	9	2	4	0.765
6	1	17	2	4	0.759
7	1	24	3	5	0.731
8	1	9	1	3	0.771
9	1	20	3	5	0.742
10	2	30	6	8	0.844
11	1	5	0	2	0.783
12	1	10	2	4	0.759



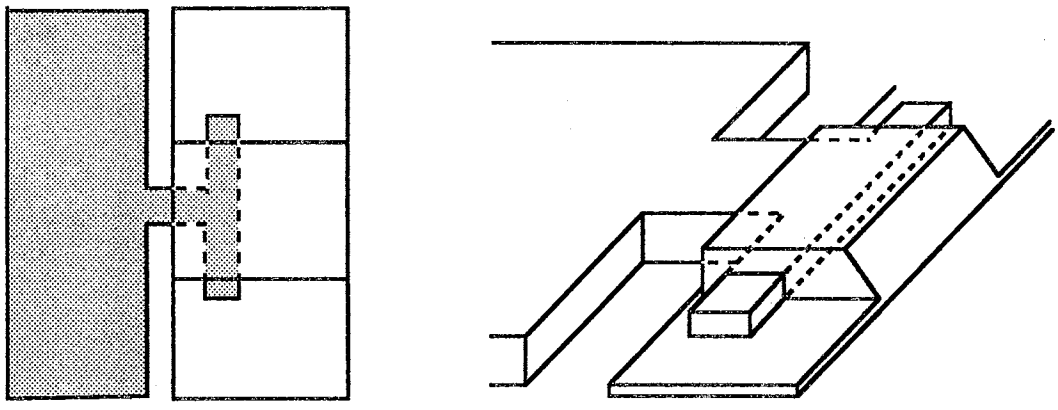
(a)



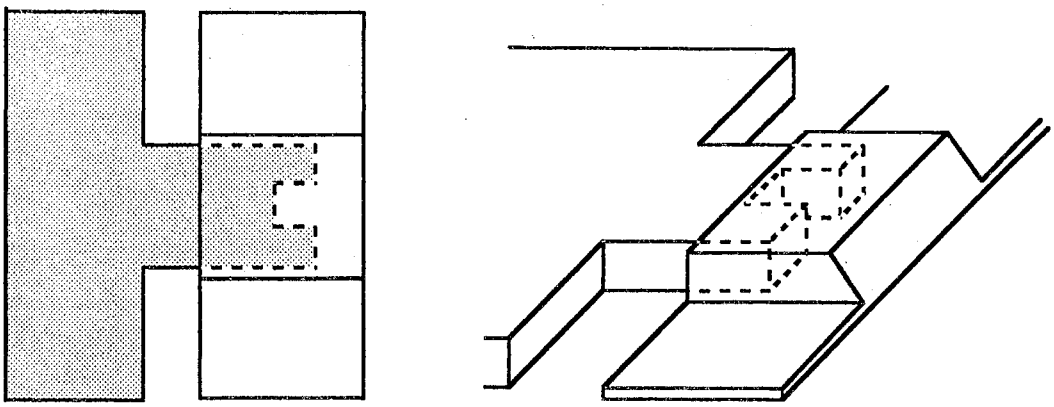
(b)



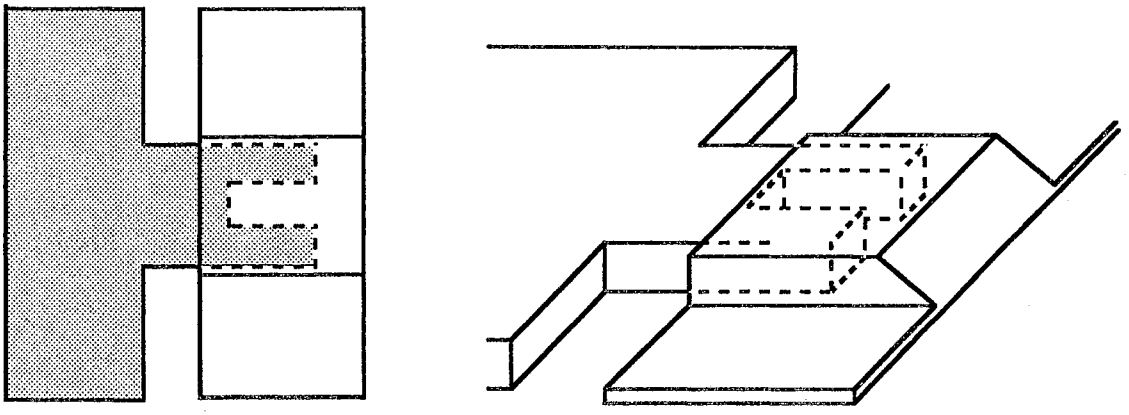
(c)



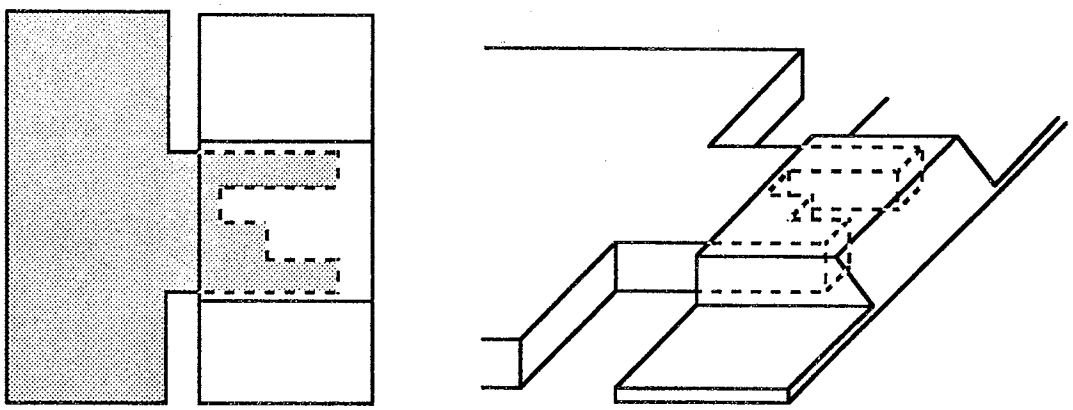
(d)



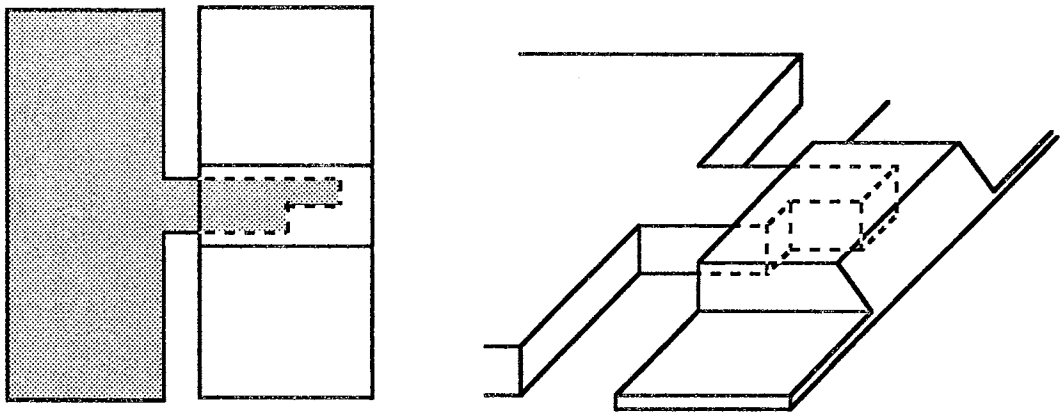
(e)



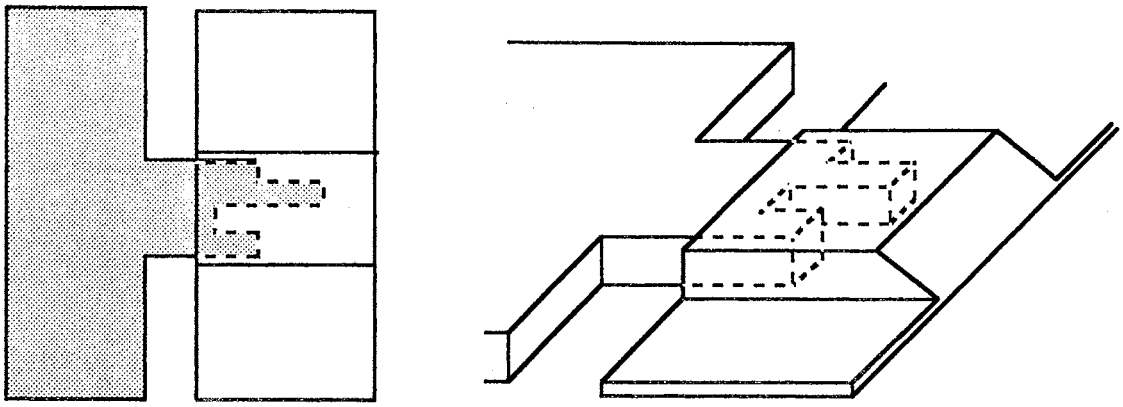
(f)



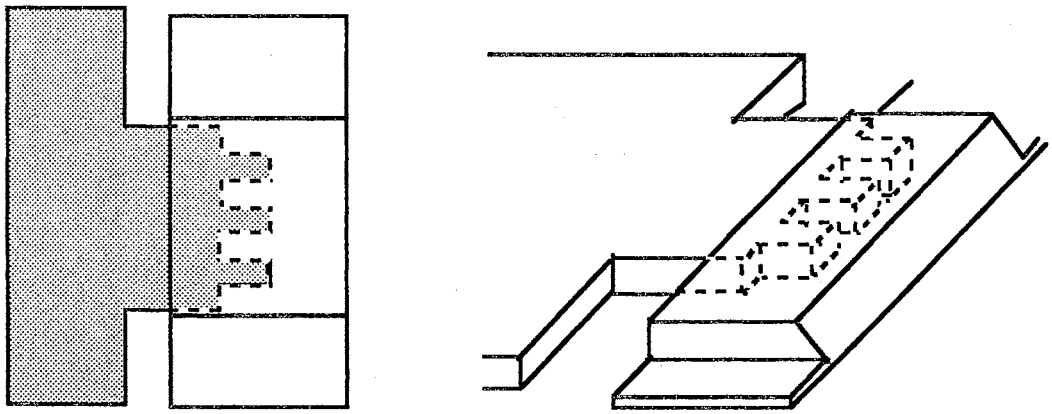
(g)



(h)



(h)



(i)

Figure 15. Top View and 3-D of Several Different Types
(a) Type1 (b) Type2 or Type11 (c) Type3 or Type12
(d) Type4 (e) Type5 (f) Type6 (g) Type7 (h) Type8
(i) Type9 (j) Type10

Experimental Method

It is very difficult to measure an injection current directly during programming since an amount of current is extremely small. Hence, an indirect measurement is preferred to measure the current. We plotted the absolute magnitude voltage across the gate oxide vs. its time derivative to evaluate the two types. The way to obtain the plot is to apply numbers of specific magnitude voltage pulses between the control gate and the injector, and to measure the change of the drain current. According to the change, we can extrapolate the voltage change across the tunneling oxide.

Figure 16(a) shows a block diagram for measuring the current injection. It is noted that the rising time of the applied pulse is long enough to avoid the charge injection induced from high frequency component. The charge movement is not from the Fowler-Nordheim injection, but from peak electric field across gate oxide. In order to create the required pulse shape, the pulse generator basically consists of RC delay circuit. The rising time we choose is approximately 5sec. Figure 16(b) shows the whole shape of the programming pulse.

The procedure of the testing is as follows;

First, looking at the testing procedure during write operation.

step 1: Apply one high programming pulse on the control gate, and connect an injector to ground. (While sw2 and 4 are on, sw 1 and 3 are off)

step 2: Connect both lines to ground. (While sw3 and 4 are on, sw1 and 2 are off)

Step 3: Measure the drain-source voltage Vds on the measuring transistor.

step 4: Go to step 1 unless there is no change in Vds.

Otherwise, stop the testing

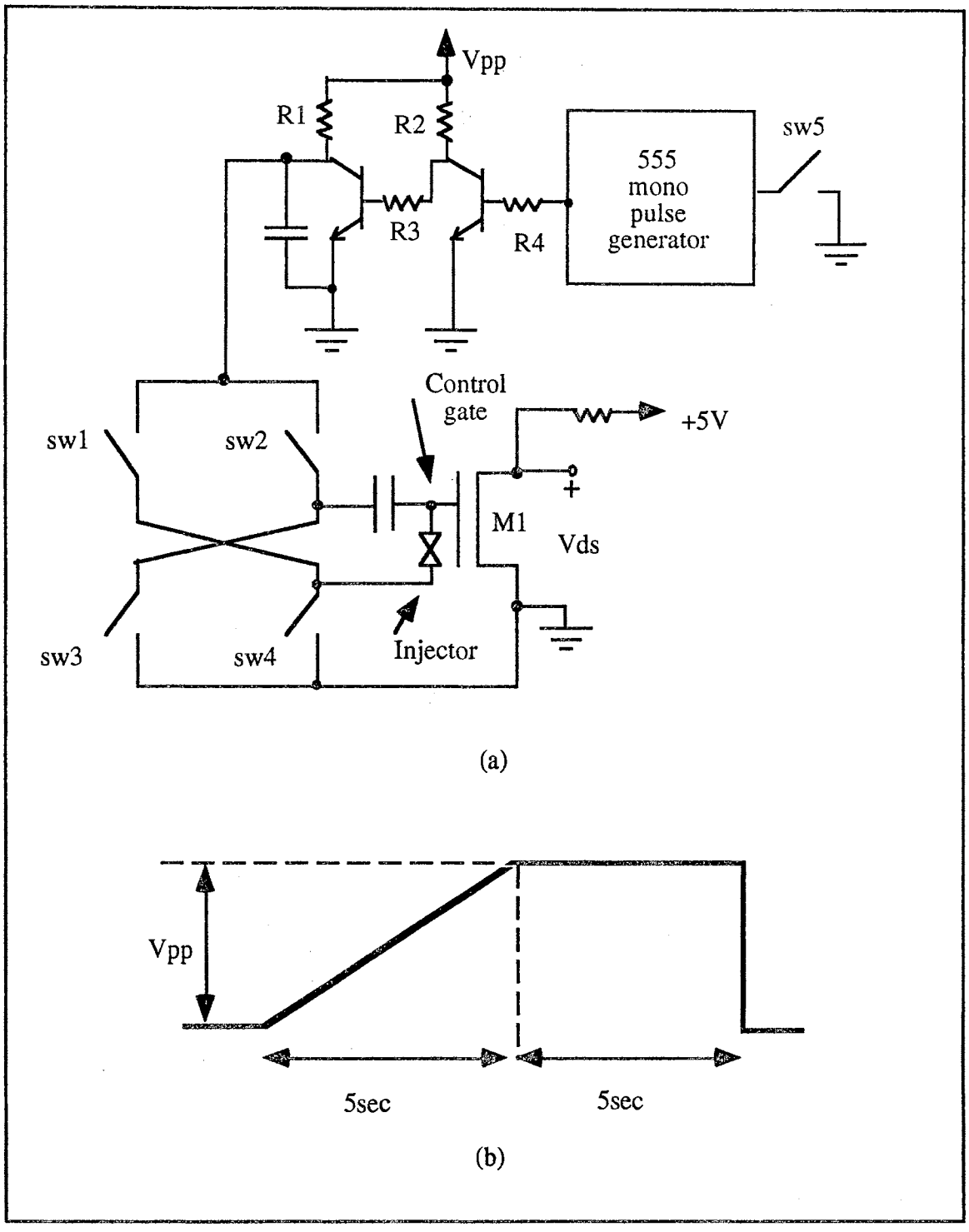


Figure 16. (a) Circuit Diagram of EEPROM Injection Current Testing Circuit (b) Shape of Programming Pulse

Second, looking at the testing procedure during erase operation.

step 1: Apply a high programming pulse to the injector, and connect control gate to the ground. (while sw1 and 3 are on, sw 2 and 4 are off)

step 2: Connect both lines to ground. (While sw3 and 4 are on, sw1 and 2 are off)

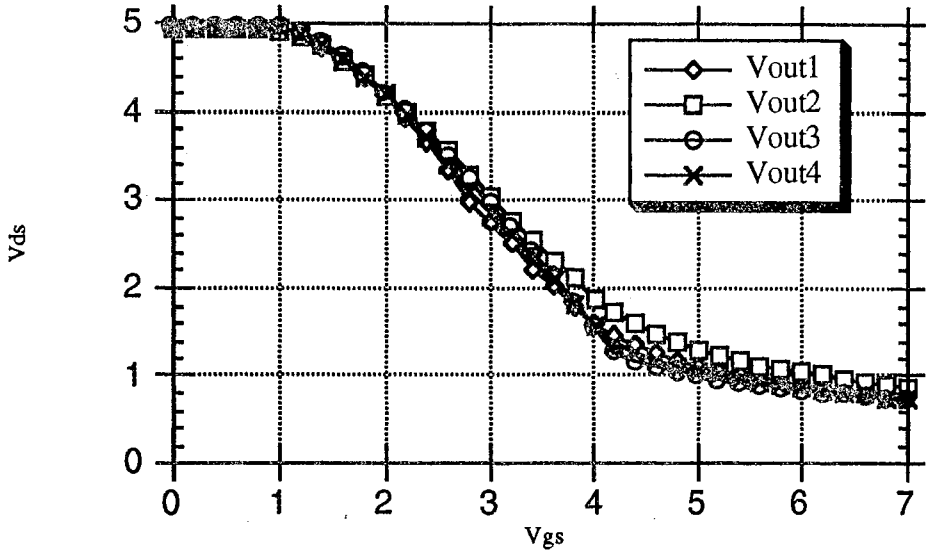
Step 3: Measure the drain-source voltage Vds on the measuring transistor.

step 4: Go to step 1 unless there is no change in Vds.

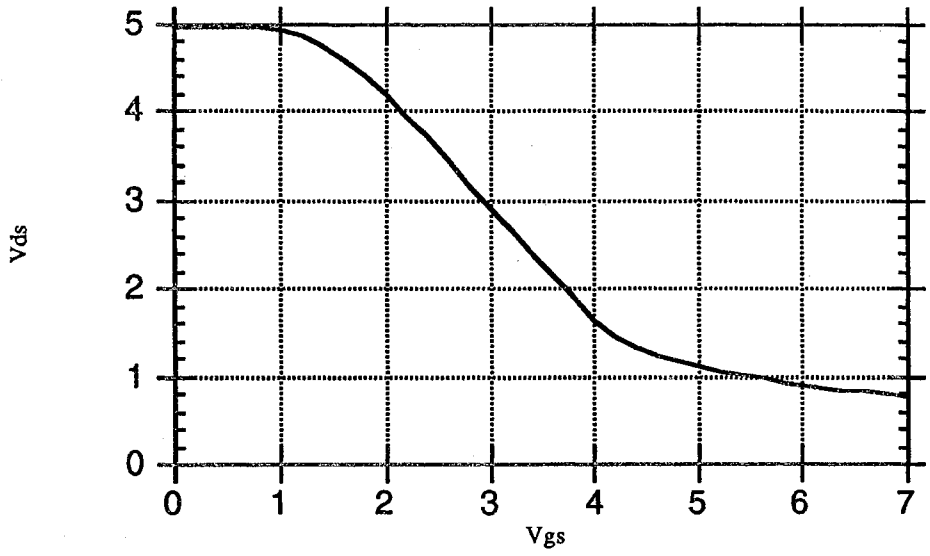
Otherwise, stop the testing.

The floating gate voltage can be presumed by observing a plot denoting the relationship, Vgs vs. Vds from a MOSFET, which has the same geometry of the measuring transistor. Figure 17(a) illustrates four plots indicating the relationship. Each plot is obtained in different die area. Figure 17(b) shows a plot averaging those four plots. With the figure, we can extrapolate the voltage across the gate oxide.

We plotted the voltage across the gate oxide vs. its derivative of each type. The graph will show the minimum requirement of the voltage initiating the emission of electrons. It will clarify the injection efficiency on each type.



(a)



(b)

Figure 17. (a) Four Transistor Characteristic Functions
(b) Plot Showing Average of Four Graphs

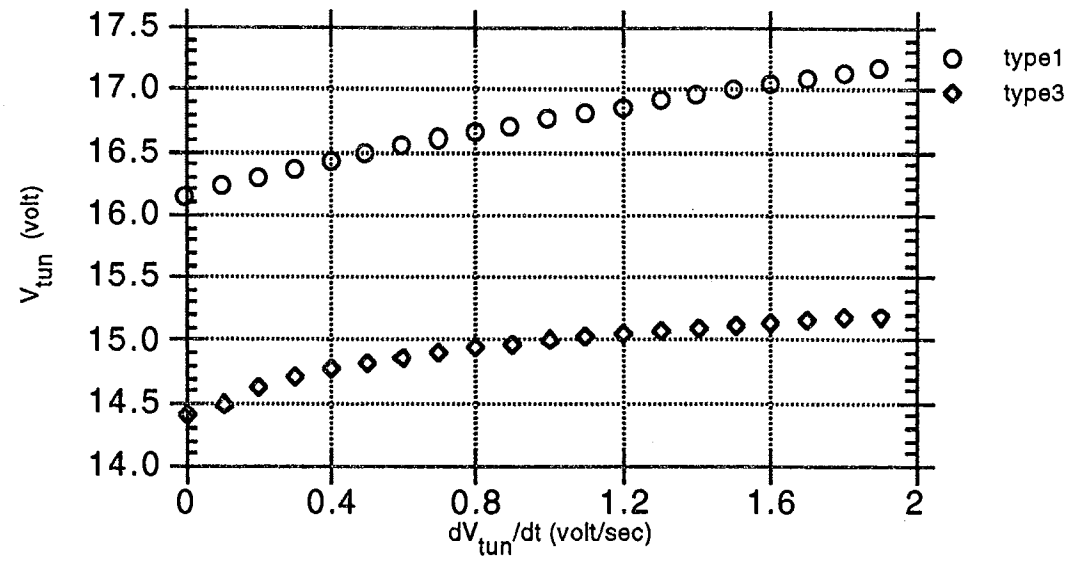
Experimental Results

As a matter of convenience, the following results and discussion will be presented on each chip unit.

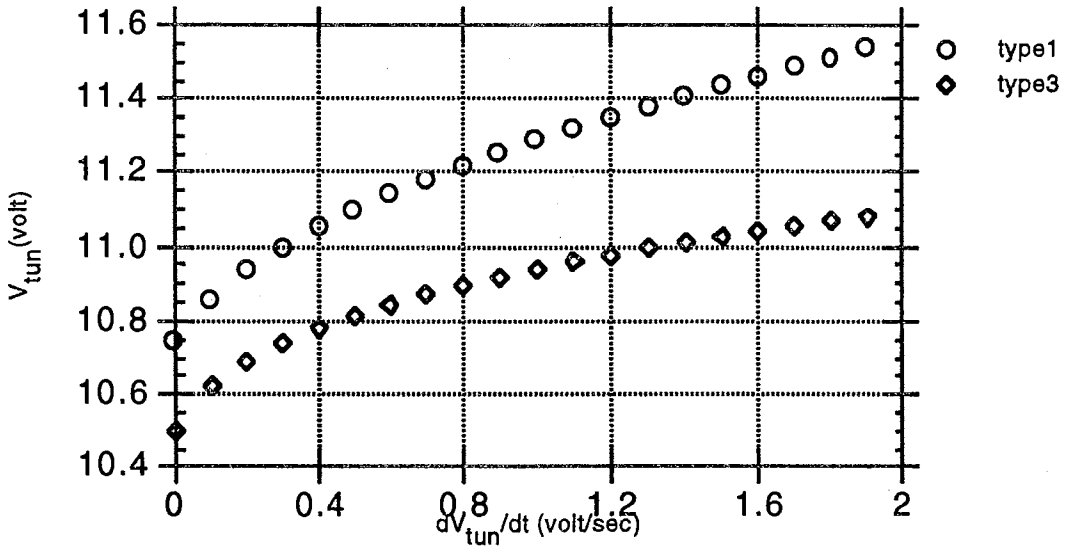
Chip1

We tested two different types 1 and 3 in chip 1. No rectangular corners from poly 2 to poly 1 exist in type 1. But in type 3, there are 2 corners to the direction. The corners in type 3 will provide an intense electric field during writing operation. Thus, we expect that the required programming voltage initiating the charge injection for the tunneling in type 3 is less than that in type 1. The result shown in Figure 18(a) is absolutely in good agreement with our expectation; the tunneling in type 1 and 3 occurs at 16.1V and 14.4V, respectively.

Meanwhile, in the erasing operation of type 1, some edges exist on the top of poly 1 instead of the rectangular corners. The edges will provide stimulus for electrons to move from poly 1 to poly 2. In type 3, both the edge and two corners toward the direction from poly 1 to poly 2 exist. The output of those two types is illustrated in Figure 18(b). From the results, it is observed that the operating voltages difference between type 1 and type 3 in erasing operation is negligible compared to that in the writing operation. It may arise from the fact that not only the corners, but also the edges are very important factors to determine the local enhancement. So the additional local enhancement of the corners in type 3 does not influence the erasing operation so deeply as the writing operation.



(a)



(b)

Figure 18. (a) Voltage Across Oxide vs. Its Derivative for Chip1 during Writing Operation (b) Voltage Across Oxide vs. Its Derivative for Chip1 during Erasing Operation

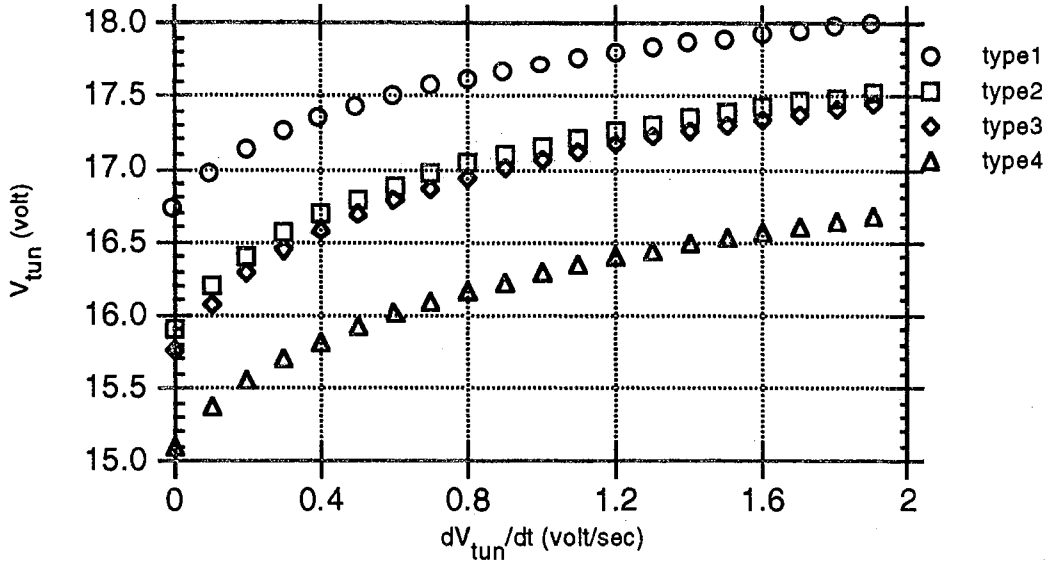
Chip2

We investigated four different types, 1,2,3 and 4 in chip2. It is observed in Figure 19(a) that the injector with rectangular corners to the direction from poly 2 to poly 1 is over the alternatives in terms of the writing voltage level; the writing voltage in type 3 and 4 is less than the voltage in type 1 and 2. Hence, it matches our expectation.

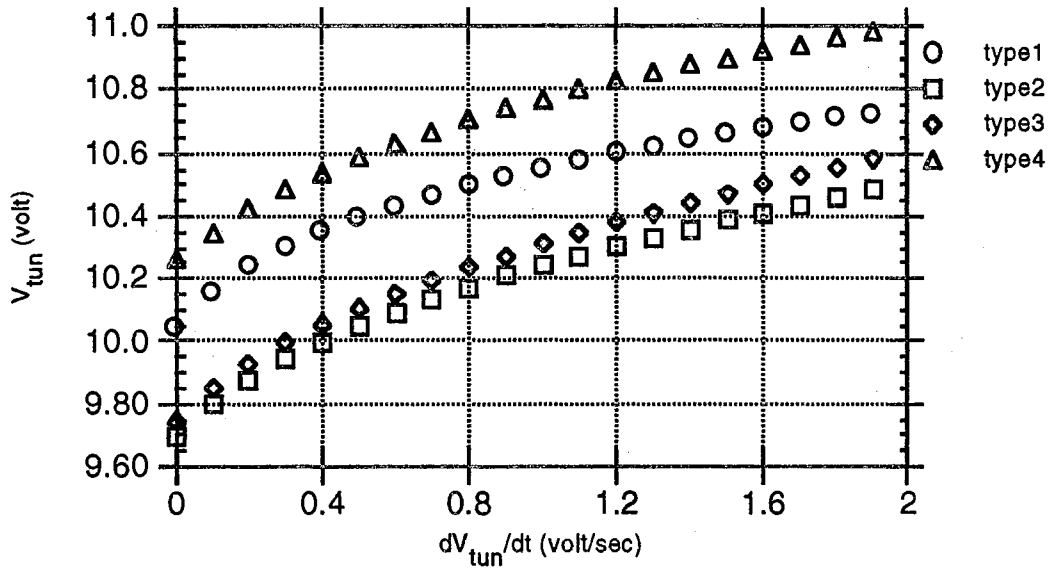
Note that the voltage level of type 1 and 3 is shifted up compared to that found in chip 1. In addition, type 4 is better than type 3, and type 2 is better than type 1 in terms of the low programming voltage level without any special reasons(the number of corners of both are exactly same). Those might arise from the fact that the thickness of oxide is not flat on the chip, or the number of tests(four times on each type) are not enough to demonstrate the unique characteristics of each type. Therefore, it is very hard to match type 1 with type 2, and type 3 with type 4. Some differences were detected even in the same type. Hence, instead of exact value of V_{tun} initiating the tunneling, we have to observe the trend inherently involved in each type.

Like chip1, the difference between each type in the erasing operation shown in Figure 19(b) is also insignificant compared to that in the writing operation. It is also seen that type 2 and 3, in which the corners exist, more easily accelerated the erasing operation than type 1 and 4, in which no corners exist.

At the beginning of the test, we expected that the amount of tunneling charges are linearly proportional to the length of the edge on the top in erasing operation. But the result shown in Figure 19(b) does not correspond with our expectation that the erasing voltage level in type 4 is lower than that in type 1 due to the length of the edge.



(a)



(b)

Figure 19. (a) Voltage Across Oxide vs. Its Derivative for Chip2 during Writing Operation (b) Voltage Across Oxide vs. Its Derivative for Chip1 during Erasing Operation

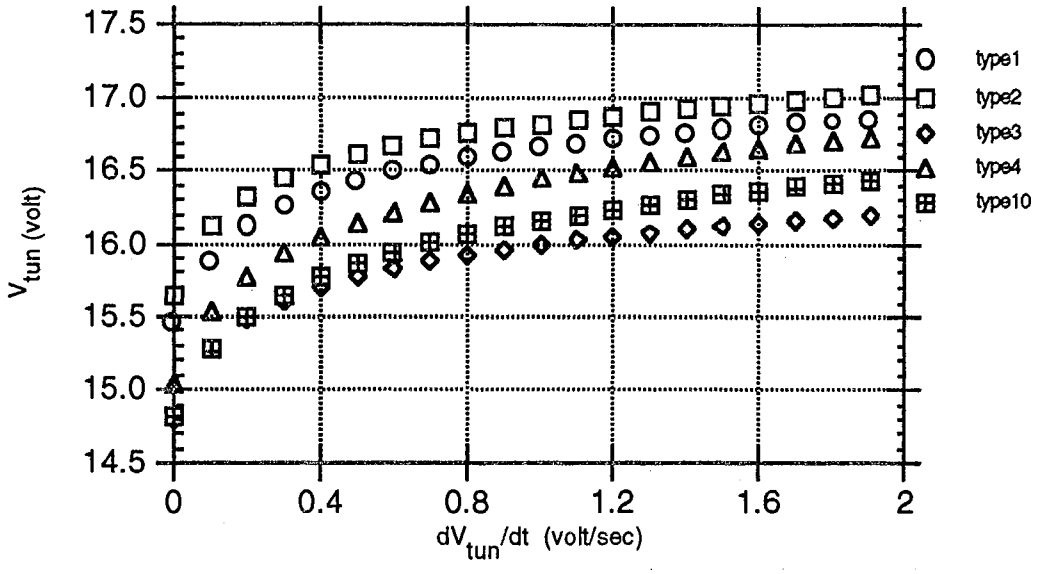
It means that the length of the edge is not critical factor influencing the electric intensity.

Chip3

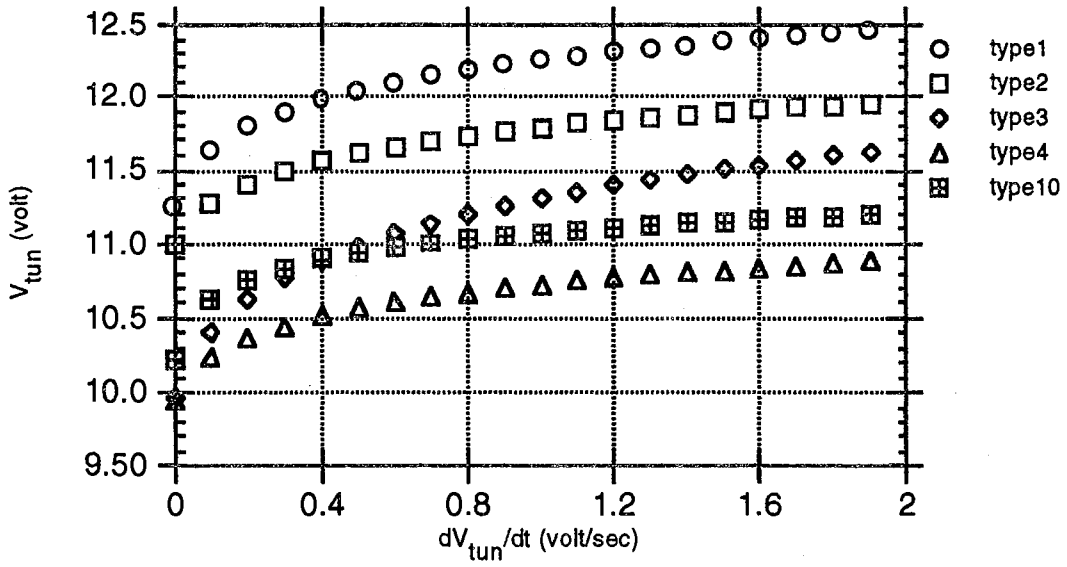
Unlike in previous cases (p-well process), chip3 and 4 were fabricated in n-well process. The trend that the writing voltage on the injector with bumps is lower than the voltage on the injector without bumps continues to be valid in n-well process, verified by Figure 20(a).

In addition to the well type, we investigated the effect of the number of corners for the programming voltage. Thus, we designed type 10 in which the structure is the same as type 3 except that several identical bumpers are concatenated. Hence, by comparing type 10 with type 3, we can speculate the difference which might have existed. Conclusively speaking, we cannot find any distinctive difference between type 3 and 10. The Fowler-Nordheim tunneling equation shows that the build-up charges on the floating gate are exponentially proportional to the programming voltage. Meanwhile, the charges are linearly proportional to the number of bumps. Thus, the additional bumps to the floating gate with bumps don't influence the build-up charge on the floating gate so much as lots of other factors, such as the oxide thickness etc. Presumably the same thing happens to the length of edge seen in chip 2. Hence, we can imagine that the programming voltage difference between type 3 and 10 is negligible. But, it is somehow curious that the voltage in type 3 is less than that in type 10 over around 0.3 volt./sec.

The result shown in Figure 20(b) is in good agreement with our expectation except that the erasing voltage in type 4 is lowest of all in spite of being no rectangular corners in type 4.



(a)



(b)

Figure 20. (a) Voltage Across Oxide vs. Its Derivative for Chip3 during Writing Operation (b) Voltage Across Oxide vs. Its Derivative for Chip1 during Erasing Operation

The irregular oxide thickness and the small numbers of testing might generate the situation. The relative small difference on each type in the erasing operation will magnify those effects. Hence, it is not sufficient to conclude that type 4 is preferred to type 3 in terms of the erasing voltage level. If we increase the number of experiments, the result may be reversed.

Chip4

Up to now, the tip size of injectors we have described have used 2 μm. It is obvious that we have observed the design rule provided from MOSIS; otherwise the connection of layout cannot be guaranteed. In chip4, we tested some specific forms of layouts in which the size of the tip on injector is just 1 μm shown in TABLE II and Figure 21

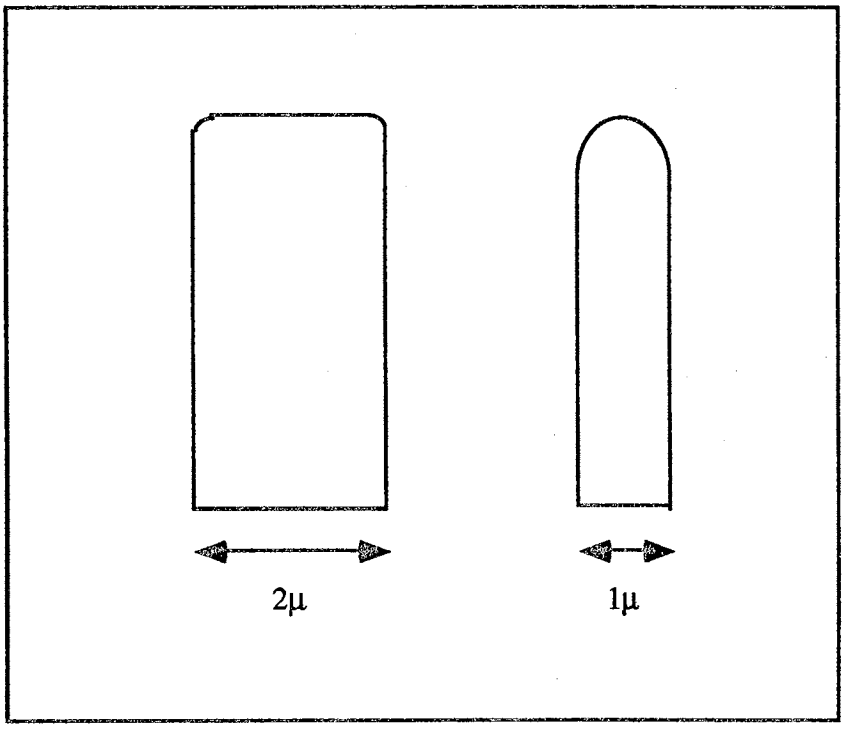
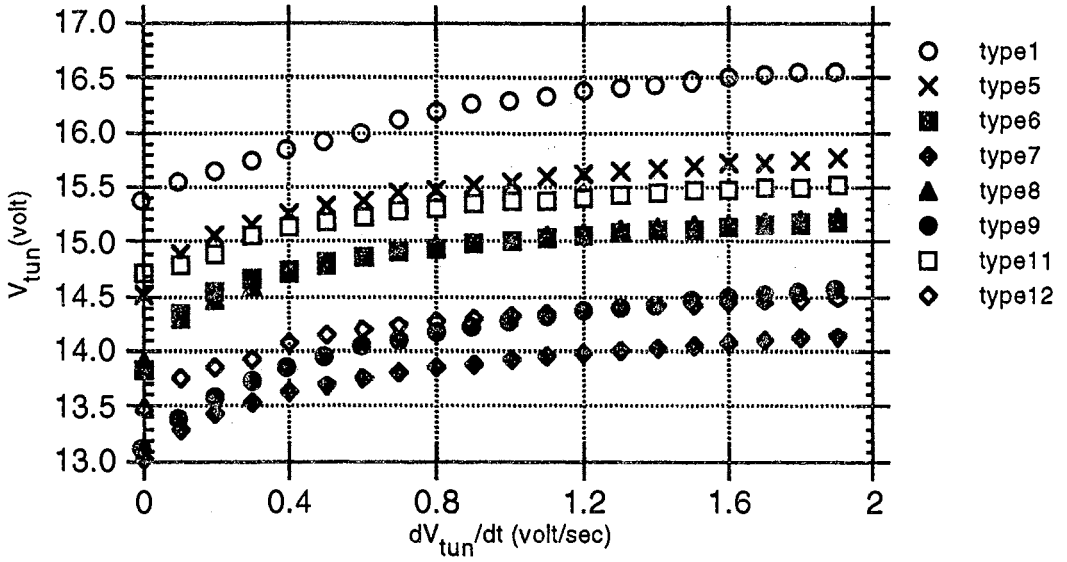


Figure 21. Expected Tip Shape on 1um and 2um

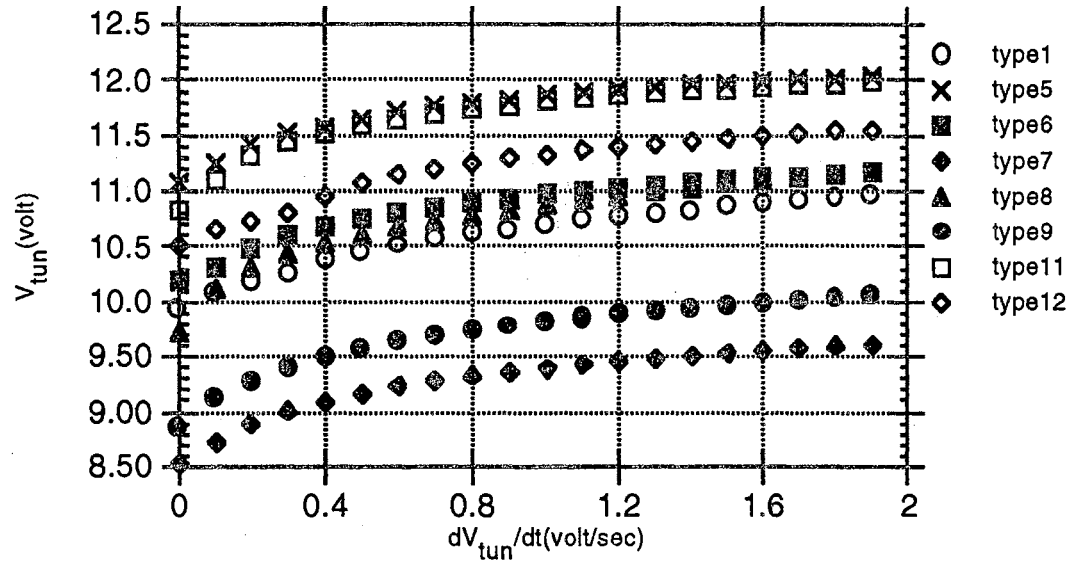
Of course, it will violate the design rule. However, we could guess that the tip under consideration is sharp compared to the previous ones. So instead of the rectangular corners in chip 1 through 3, we used several sharp angle tips in chip 4. For example, type 3 and type 12 are same in the layout, but different in the size of the tips. While the bump of type 3 has two corners from poly 1 to poly 2, the bump of type 12 has just one sharp tip to the same direction (the number of corners in TABLE II are based on the standard layout). It is already known that the angle of the tip influences the strength of the electric field[Spindt,1976]. Figure 22(a) shows that objects with sharp tip are much better than type 1(in chip4, only type 1 is supposed to have rectangular corners) in terms of programming voltage level. It means that the tip angle is very important factor intensifying the electric field. This is especially true when the length of the tip is long, then the result is more desirable due to the fact that the longer the length of the tips, the sharper the angle of the tips. Figure 22(a) showing that type 7 and 9 are superior to the others supports our assumption.

The tiny tips not only contribute to the enhancement of the electric field, but also reduce the size of the floating gate. For example, the area of the tip in type 2 and type 11 is $4\mu m^2$ and $2\mu m^2$, respectively. Then, the coupling ratio can be kept in constant with the half size of the control gate. Consequently, the overall size of the floating gate can be reduced to half.

Figure 22(b) shows that type 7 and 9 have better performance than type 1 in the erasing operation. However, the situation for the others is different. Maybe this is the same reason we confronted in chip 2 through 3: the non-flat oxide thickness and the insufficient number of tests.



(a)



(b)

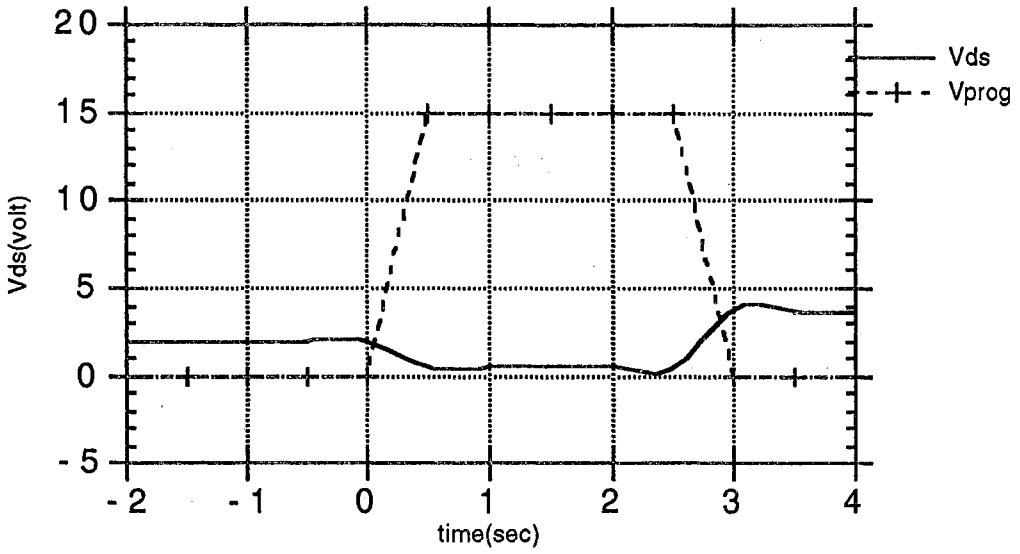
Figure 22. (a) Voltage Across Oxide vs. Its Derivative for Chip4 during Writing Operation (b) Voltage Across Oxide vs. Its Derivative for Chip1 during Erasing Operation

EEPROM Instability

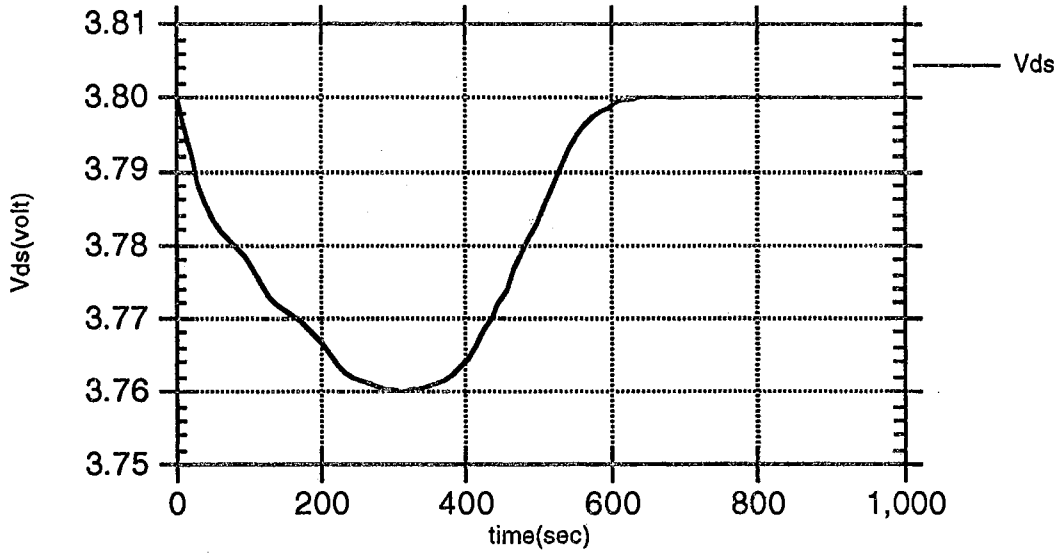
The retention time of a floating gate is supposed to be permanent. However, in reality, there is a leakage in the floating gate. The retention time is classified as two different sorts: short-term and long-term instability.

Figure 23 (a) and (b) shows the change of Vds on the measuring transistor during and after writing operation, respectively. The Vds fluctuates depending on the shape of the programming pulse due to the fact that the programming voltage is added to the floating gate voltage during programming. After programming, a few charges are trapped up in oxide due to the oxide defect[Mielke, 1987]. The buildup charge in oxide will induce the short-term instability since the charge being under very unstable state is settled to a new equilibrium state near SiO_2 interface[Carley, 1989]. Figure 23 (b) shows the short-term instability after a programming pulse. However, the magnitude of the instability is relatively small compared to the magnitude of the floating gate voltage. Hence, the unstable state may be insignificant. But sometimes the instability cannot be allowed in a specific system requesting very high precision. Then it will delay the overall system operation since the system has to wait until the floating gate voltage goes to a stable state.

In addition to the short-term instability, there is a long term instability since oxide defect allows electrons to leak off charges from floating gate[Mielke, 1987]. Note that the oxide thickness influences the long-term instability; the thicker the oxide, the longer the retention time[Wang, 1980]. Therefore, in terms of the retention time, the cells with bumps, type2 through type12, may deteriorate the performance of the non-volatile memory device since the thinning of the oxide adjacent to each corner is expected in those structures.



(a)



(b)

Figure 23. Instability of Vds (a) During Programming (b) After Programming(Short-Term Instability due to Oxide Trapup)

Bakes of devices could be used effectively as a long term retention test, since a floating gate leaks off the charge significantly during a day at a high temperature. Using the amount of charges from the floating gate, we can predict how long the relevant cell can retain sufficient charges at a room temperature[Carley, 1989][Nozawa, 1982]. Assuming that the retention time characteristic at elevated temperatures is described by a thermionic emission model:

$$\frac{Q(t)}{Q(0)} = \exp\left(-t\nu\exp\left(-\frac{\Phi_B}{kT}\right)\right) \tag{2-14}$$

where Φ_B is the energy barrier at the polysilicon-SiO_2 interface, and ν is the dielectric relaxation frequency of electrons in the polysilicon, k is Boltzman's constant, and T represents the temperature in Kelvin. Here, Φ_B and ν on each cell should depend on device structure[Nozawa, 1982].

In this experiment, we tested type7. The type7 has the best performance with respect to the programming voltage level, but is supposed to be the worst in terms of the retention time. In order to obtain those Φ_Bs, we measured Q(t)/Q(0) of each type at the temperature $130^{\circ}C$, $160^{\circ}C$, and $200^{\circ}C$.
Eq. (2-14) can be rearranged as follows;

$$\ln\left(\frac{\ln\left(\frac{Q(0)}{Q(t)}\right)}{t\nu}\right) = -\Phi_B\frac{1}{KT} \tag{2-15}$$

Where the unknown values are ν and Φ_B.

Then, we can extrapolate a straight line based on the three different points. From the slope and the intersection of the plot, we can obtain Φ_B and ν.

The way to measure the variation of the charges on the floating gate is shown below,

$$\frac{Q(t)}{Q(0)} = \frac{V_T(t) - V_{T0}}{V_T(0) - V_{T0}} \qquad (2\text{-}16)$$

where $V_T(t)$ is the threshold voltage at time t after programming, $V_T(0)$ is the initial threshold voltage, and V_{T0} is the threshold voltage before any charge has been transferred into a floating gate.

In order to obtain V_T experimentally, the $\sqrt{I_{ds}}$ of the measuring transistor was measured at several V_{gs} points shown in Figure 24. The estimation is based on the theory that the square root of the current I_{ds} is a linear function of gate-source voltage with intercept of V_T in a strong inversion region[Tsividis, 1988]. From Figure 24, the value V_{T0} =0.6V was obtained.

The charge reduction obtained from the threshold voltage shift is used to measure Φ_B and ν experimentally at three different temperatures as shown in Figure 25. From the result, Φ_B=1.149eV and ν=9.9E6/sec are obtained. Based on the result, the retention time of type 7 shows 0.1% leakage of charge around 61 yeas at a room temperature. Consequently, the thinning of the oxide thickness is assumed not to be serious in the floating gate MOSFET.

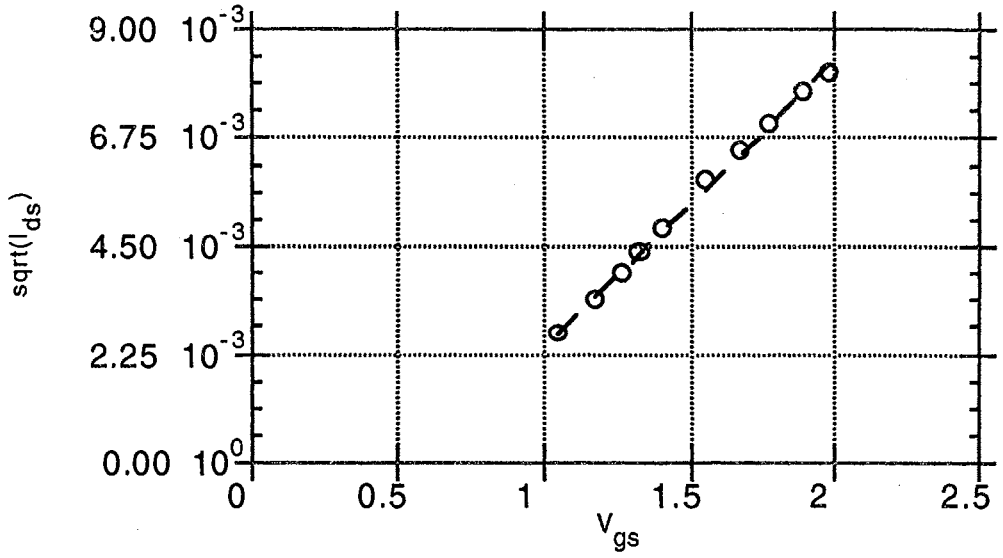


Figure 24. Graphical method for determining Vth

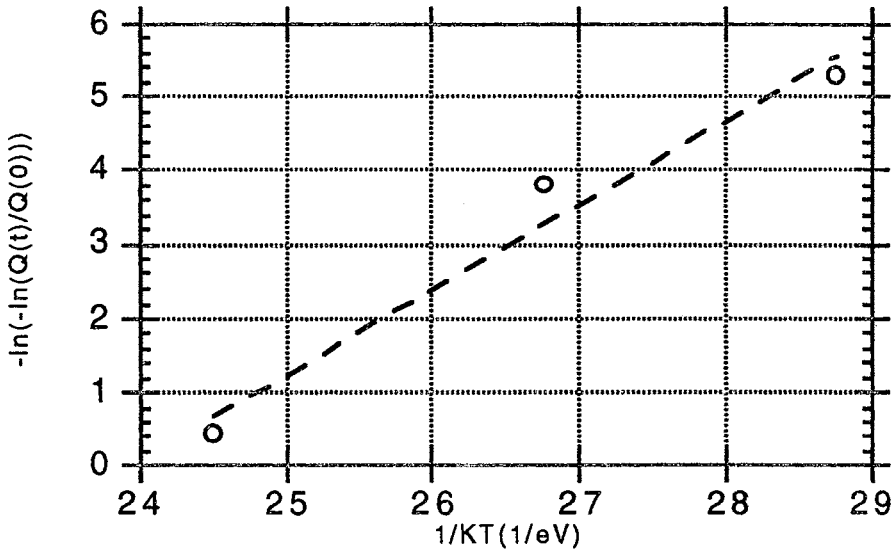


Figure 25. A plot of the retention characteristics

CHAPTER IV

A 2X2 ANALOG ARRAY IMPLEMENTATION

Introduction

This chapter provides the description of a 2 by 2 analog array implemented with an innovative floating gate. The analog signal storing device which is expandable to any size of memories has been designed to reduce a programming voltage, enhance a resolution of storing signal, and simplify control circuits compared to conventional memories.

Design Requirement

The analog memory can be classified as two groups depending on the style of the programming: open loop and closed loop. In the open loop memory system, predetermined programming pulses will be applied to an EEPROM device according to a desired threshold voltage in a cell of interest. This is feasible because there is an approximate linear relationship between the threshold voltage and the programming voltage[Ong, 1989]. While the method may simplify an overall control system of the memory, we cannot expect an exact matching between the real and the expected threshold voltage due to the imperfect linear function caused by the variation of the floating gate charges. The eq. (2-3) and (2-4) in chapter II explain why it is difficult to store analog data

linearly by applying sequential constant voltage pulses. On the other hand, the memory system implemented with a feedback circuit will keep observing and modifying the threshold voltage in the cell of interest until the threshold voltage reaches a certain predetermined resolution. If a low-frequency tracking accuracy is more desirable than speed, then this is accomplished by the closed loop strategy.

Our system uses a closed loop structure to monitor the current state of the cell of interest in an array. It will not only delay the whole programming operation, but also lead to a relatively complicated controller. However, a fine resolution will be expected if a reasonable magnitude and duration programming pulse is provided. The accuracy will drastically increase the performance of the analog memory.

Our design of the analog memory of the closed loop strategy includes some of unique characteristics. The following features of our analog memory are key advantages over the others. First, the programming can occur at any place no matter what the current floating gate voltage is. The unconstrained beginning of the programming not only reduces the total programming time, but also prolongs the lifetime of the memory device due to the reduced number of stressing. The storage device made by Blyth begins programming after the erasing operation on a selected cell. The programming algorithm seems to be very simple, but it will drastically impair overall programming time due to the unnecessary erasing operation. The problem becomes severe when the analog memory is applied to a neural network, where weight change of each neuron is negligible during the programming. Hence, the sudden fluctuation caused by the programming after the erasing operation is not appropriate in the neural network application.

Second, we have designed a sensor with a folded cascode amplifier which has been known to contain high gain, wideband, high PSRR and CMRR, and etc. over an simple differential amplifier. The open loop gain of the amplifier is constructed to be 50dB. Note that the analog memory has been designed to sample an input voltage in the range 1Volt to 3.6Volt with 10mvolt resolution. The goal of the project is to replace a digital word with a single analog memory cell. The high gain amplifier makes it feasible if one programming pulse increases a threshold voltage within 10mV. In addition to the high gain with just a single stage, the gain bandwidth and phase margin are around 5MHz and 80 degrees with a 2pF capacitive load, respectively. As a result, the circuit design is extremely stable without an additional compensation circuit.

Third, the offset voltage of the comparator no longer influences our whole circuit operation due to the fact that the trajectory of the reading and the programming of a cell in the array is identical. An input offset voltage is expected in a typical MOS differential pairs, which is used for comparing a reference with a selected node voltage in our design. For example, let us assume that there exists a geometry mismatching in the current mirror of the operational amplifier. Then after programming, there should be some errors between the voltage of the reference node attached to an incoming signal and the voltage of the selected cell in the array. Assuming that the same amount of voltage difference occurs during the following read operation, then the measured value corresponds to the original one due to the cancellation of the same amount of the error.

Fourth, the unique shape of the injector structure described in the previous chapter is adopted as a cell of the analog array. It drastically reduces the programming voltage so that the internal circuitry of the whole chip is out of

a breakdown voltage. In addition, it reduces the size of the programming pulse generator which occupies a dominant area of the memory.

Overview of 2x2 Analog Array

A simplified block diagram illustrating a 2 by 2 analog array is given in Figure 26. More detail is shown in Figure 27. Our storage system consists of a 2 by 2 analog array, a PEDC (programming enable/disable determination circuit), a PTDC(programming type determination circuit), four HVPGs(high voltage pulse generator), and a comparator. The PEDC should determine whether the programming operation will continue or not. The role of the PTDC is to determine the type of the programming, writing and erasing depending on the current state of the comparator. The HVPG generates three different level voltages according to the location of the selected cell. Finally, the comparator is used for observing the programming state. The detailed circuit description of each block will be provided later.

There are two different modes in the memory operation: programming and reading. During the programming, the switches, sw1 and sw2 shown in Figure 26, are open and closed, respectively. The interconnection allows the system to compare a reference with a chosen cell voltage in the array. On the other hand, during the reading operation, the sw1 and the sw2 are on and off, respectively. Then the amplifier becomes to be a buffer. The value of the selected floating gate voltage will be sent to the off-chip through this amplifier. It is seen that the trajectory of the signal routing of both the programming and the reading is exactly the same since just one comparator supports both operations. It helps to eliminate the input offset voltage problem, which is known to be a very critical factor deteriorating a comparator.

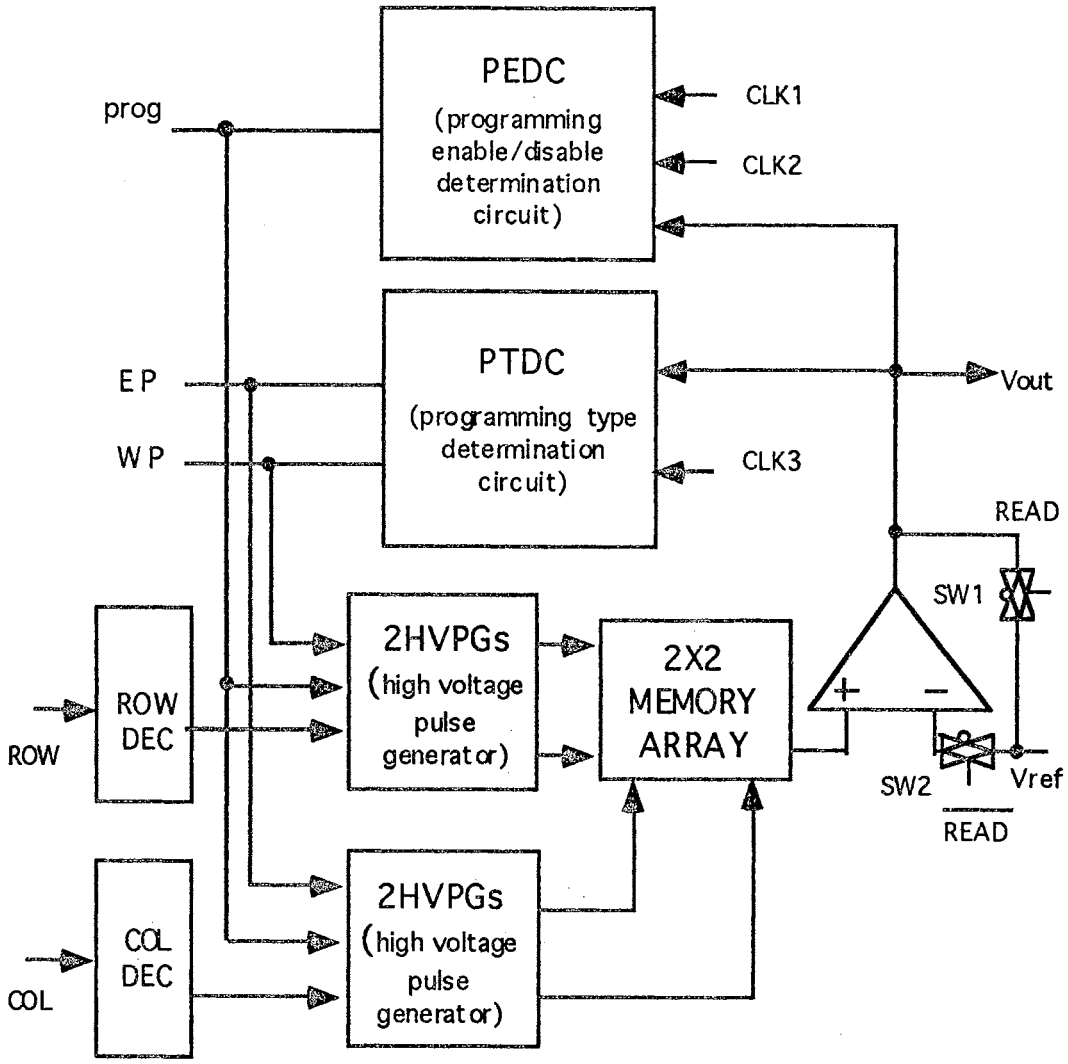


Figure 26. A 2x2 Analog Array Block Diagram

The circuit diagram of the 2 by 2 analog array is shown in detail in Figure 27. In this work, between every intersection of the 2 by 2 matrix, a floating gate is inserted with its control gate connected to the positive programming pulse generator for writing and its injector tied to the other pulse generator for erasing. The transistors, M1A, M1B, M1C, and M1D attached to each floating gate are used to measure the amount of floating gate voltages by the knowledge of the drain current. The drains of these transistors are connected to the other transistors, M3A through M3D, being used for selecting a row of this array. The drain of the row selecting transistor also is tied to another transistor M4A and M4B used for choosing a column of the array. Once both selecting transistors are chosen by setting the enable signals, "Row" and "Col", then the nodeX shown in Figure 27 will be charged or discharged depending on the amount of charges in the selected floating gate. It is important to note that the voltage of the nodeX will be greater than 2.5V with the condition that the Vin, floating gate voltage, is greater than the Vref, reference voltage, during the programming operation. Otherwise, it will be reversed. The fact that voltage at nodeX is 2.5V means that the floating gate voltage is exactly equal to the reference voltage, assuming that there is no input offset voltage. During the reading operation, the voltage of the nodeX indicates the selected floating gate voltage.

Now, we will show the procedures required for both reading and programming operations step by step. For convenience, we assume that the M1A is the cell of interest. Adjacent memory cells, M1B, M1C, and M1D, should not be disturbed during the programming of a particular cell.

First, look at the read operation. During the read operation of M1A, the following sequence of operations must take place;

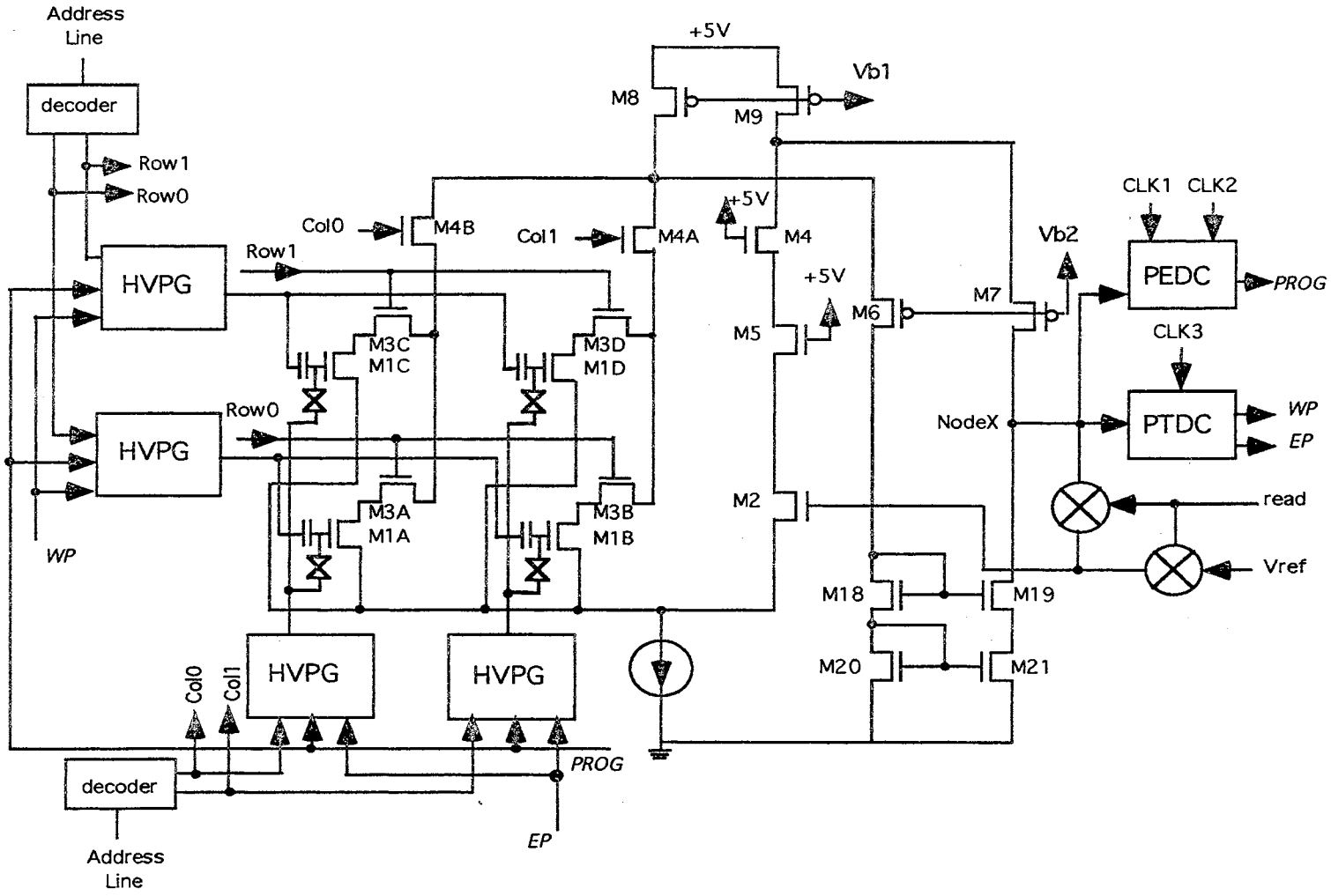


Figure 27. A 2x2 Analog Memory Circuit Diagram

step 1: The content of the memory location indicated by the memory address register is interpreted in the decoders.

step 2: The decoders set Row0 and Col0 lines to be on. M4B and M3A are enabled to pass current through M1A.

step 3: Record the voltage of the nodeX. It corresponds to the floating gate voltage of the selected MOSFET.

The programming operation(writing/erasing) is a process adjusting the floating gate voltage to the same as the reference voltage. To program information into the floating gate from an external device, the following sequence of operations must take place,

step 1-2: same as the read operation

step 3: According to the voltage on the nodeX, the PTDC circuit attached to the comparator determines which operation, writing or erasing, will perform on the next cycle.

step 4: If the voltage on the nodeX is 2.5V, the signal "prog" from the PEDC will be disabled, and it will terminate the operation. Otherwise, go to the next step.

step 5: According to the result from the PTDC, a specific HVPG will apply one programming pulse to the control gate(injector) of M1A, and tie the injector(control gate) of M1A to ground. At the same time, Vmid from the other HVPGs will be applied to the remaining row and column, Row1 and Col1, in order to avoid unnecessary programming.

step 6: Go to step 4

Comparator Design

General Design Requirement

In digital memories, the high programming voltages are applied in a binary fashion to cause a large amount of tunneling current to write one or zero. However, in the analog memory, due to the unlimited number of output states, the comparator design which is used for tracing the programming operation, is somewhat deliberate. The type of the comparator considered includes the low-speed, but the high gain differential amplifier. In addition, the wide range CMR, high CMRR and PSRR are preferred for not deteriorating the performance of the amplifier. Figure 28 shows the implementation of the self compensated folded cascode amplifier, which is suited for satisfying those conditions very well. The amplifier has been designed with an analog CMOS 2μ double poly nwell technology in ORBIT process. In the circuit diagram, the floating gates are removed in order to perform the circuit analysis easily. Just one of the array is chosen at a time to compare with a reference voltage by enabling one of the switches(M3A through M3D) for selecting a row and one of the others(M4A and M4B) for selecting a column. The input to the reference node comes from the off chip and is designated Vref. Including the switches, two more complementary switches exist in the comparator. Those switches are used for distinguishing the reading from the programming mode. That is, the enabled "read" signal activates the transistor, M14 and M15. It leads the comparator to be a closed loop connection so that the amplifier should act like just a buffer. It means that the system is capable of recognizing the voltage of the floating gate MOSFET selected by the decoders.

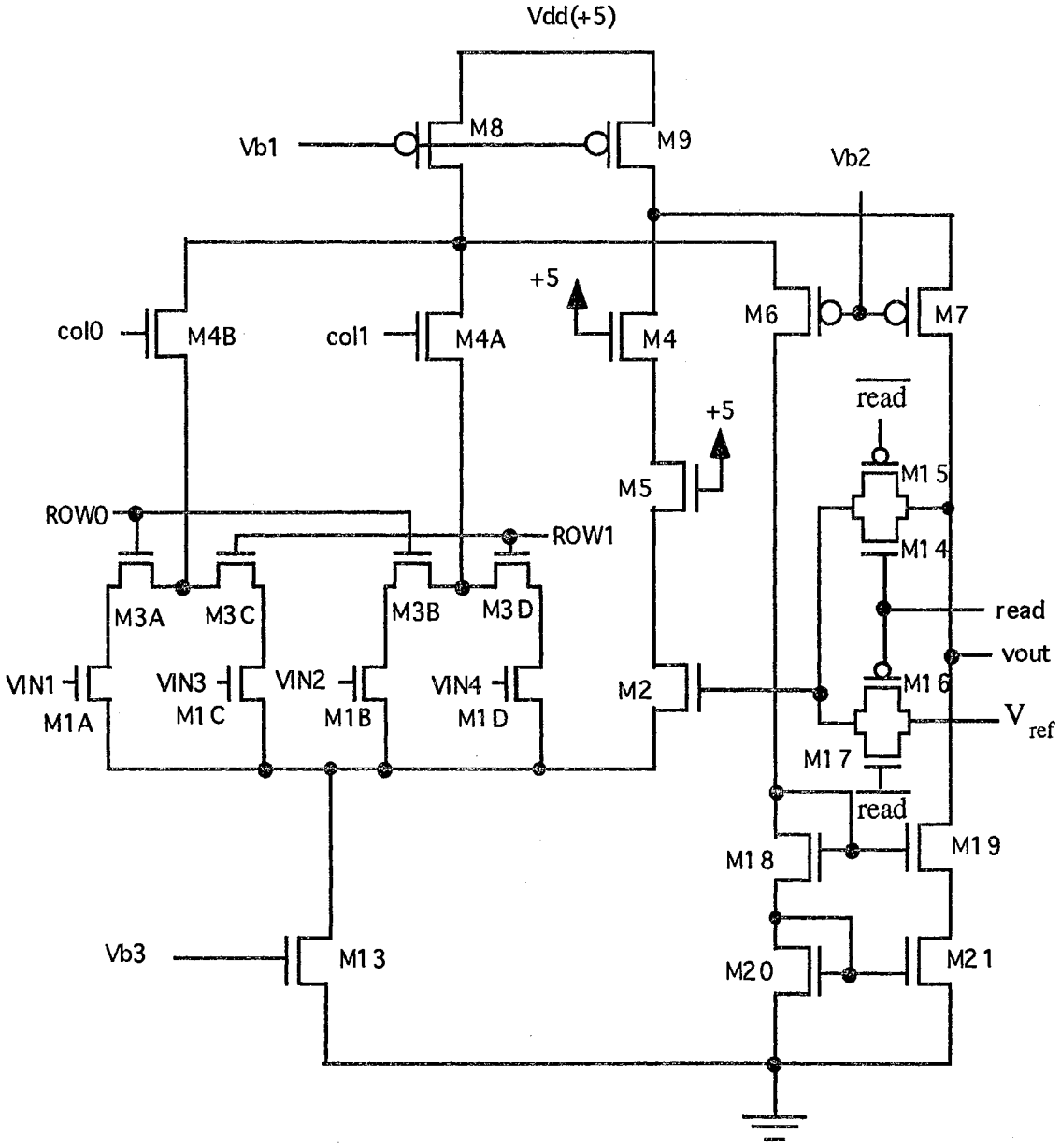


Figure 28. A Folded Cascode Amplifier Circuit Diagram

Meanwhile, the incoming signal of the transistor M2 will be the reference voltage with the disabled "read" signal. Then the arrangement for executing the programming operation is completed.

The layout of the comparator utilizes a common centroid geometry to reduce the input offset voltage which can be caused by threshold voltage, geometry, and temperature mismatch[Allen, 1987]. An example of this layout is shown in Figure 29.

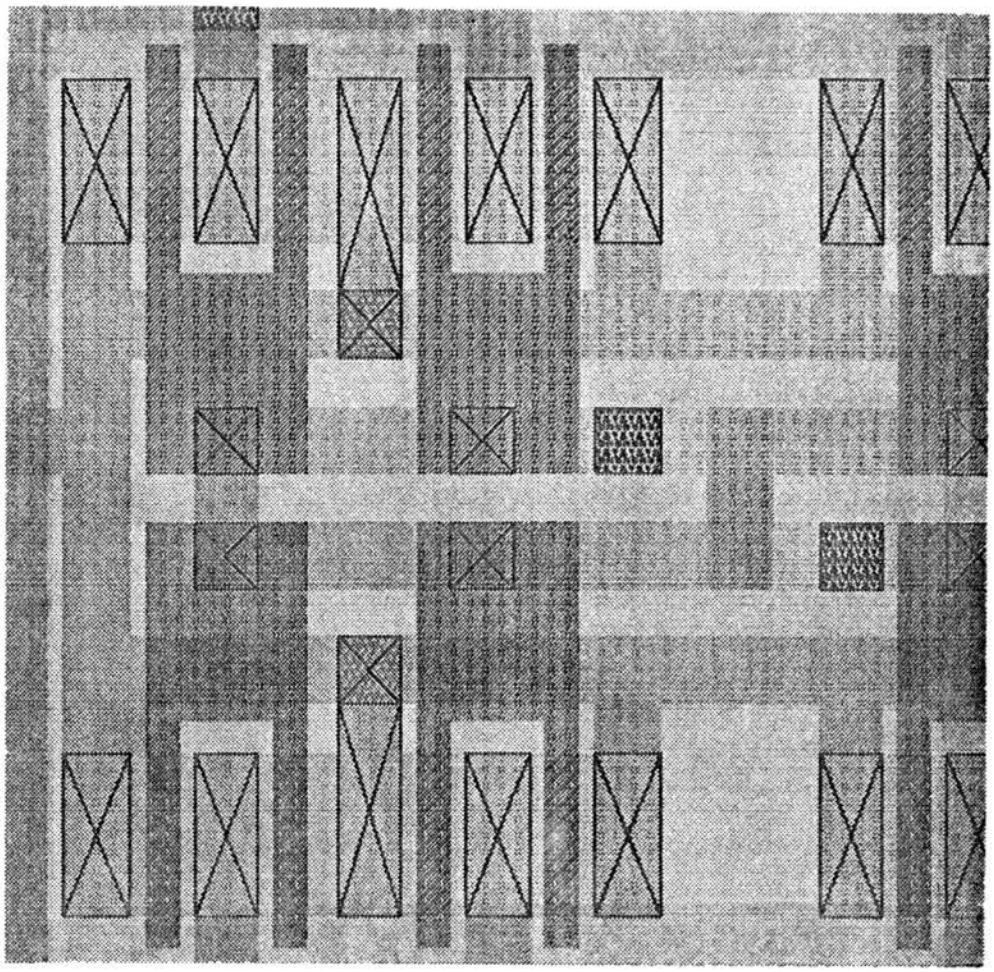


Figure 29. Layout of a Common Centroid Geometry

In our comparator design, the offset voltage caused by two pairs, M8 and M9, and M20 and M21, are partially compensated by the common centroid. But the technique is not applicable to the remnant transistors, M1A through M1D due to complexity. Especially, it is impossible to develop a large array implemented with the common centroid technique. Thus, some offset voltage still exist. Due to the mismatch of a transconductance of the input transistors, the floating gate of interest is not identical to the reference voltage even after programming is completed. But, fortunately, during the reading operation of the same intersection node, the mismatch will induce to the same amount of error. Theoretically, no matter how much geometrical mismatch between those nodes exists, the value read from the analog memory will correspond to the initially applied reference voltage. The design will eliminate even the error created by the irregularity of each floating gate MOSFET.

The length of M1A through M1D is chosen to satisfy the minimum requirement, 2um. All the other transistor lengths have the same length to avoid inconsistency. The width of the transistors, M1A through M1D, is chosen to be 4um. The increment of the transistor width will contribute to increasing a DC gain. However, assuming that the size of the memory expands from small to large, then it should deteriorate information storage density.

The TABLE II shows the geometry of each transistor in our design. Each parameter follows the specification of Figure 28 and Figure 30.

The differential amplifier is biased with the cascode current amplifier shown in Figure 30. It is useful in maximizing the bias voltage range[Allen, 1987]. Additionally, the type of current source was chosen for its high output resistance. Consequently, the bias current will be tolerable to the variation of DC voltage.

TABLE II

GEOMETRY OF EACH TRANSISTOR
SHOWN IN FIGURE 28 AND FIGURE 30

transistor	(width/length)	transistor	(width/length)
M1NB	40/2	M4	4/2
M2NB	10/2	M4A	4/2
M3NB	40/2	M4B	4/2
M4NB	40/2	M5	4/2
M5NB	10/2	M6	40/2
M6NB	10/2	M7	40/2
M1PB	80/2	M8	80/2
M2PB	20/2	M9	80/2
M3PB	80/2	M13	40/2
M4PB	80/2	M14	3/2
M1A	4/2	M15	6/2
M1B	4/2	M16	6/2
M3B	4/2	M17	3/2
M4B	4/2	M18	5/2
M2	4/2	M19	5/2
M3A	4/2	M20	40/2
M3B	4/2	M21	40/2
M3C	4/2		

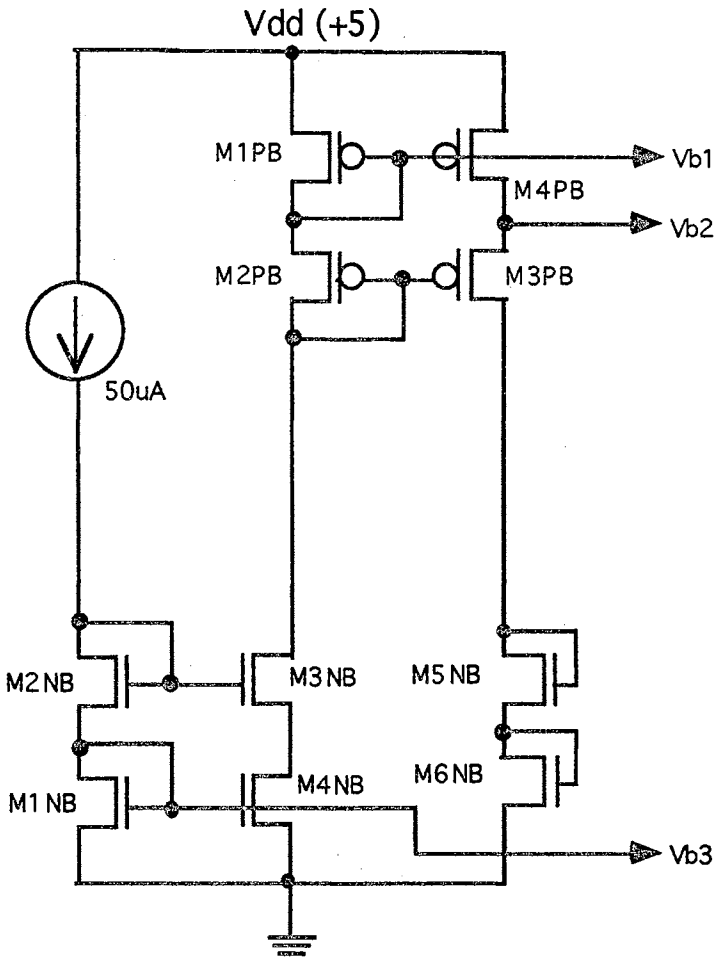


Figure 30. Cascode Bias Circuit

Design Analysis

For clarity, all the switches in Figure 27 are not shown in the simplified schematic diagram of Figure 31. Also, only one cell among four, M1A through M1D, is chosen to represent an input in the differential amplifier.

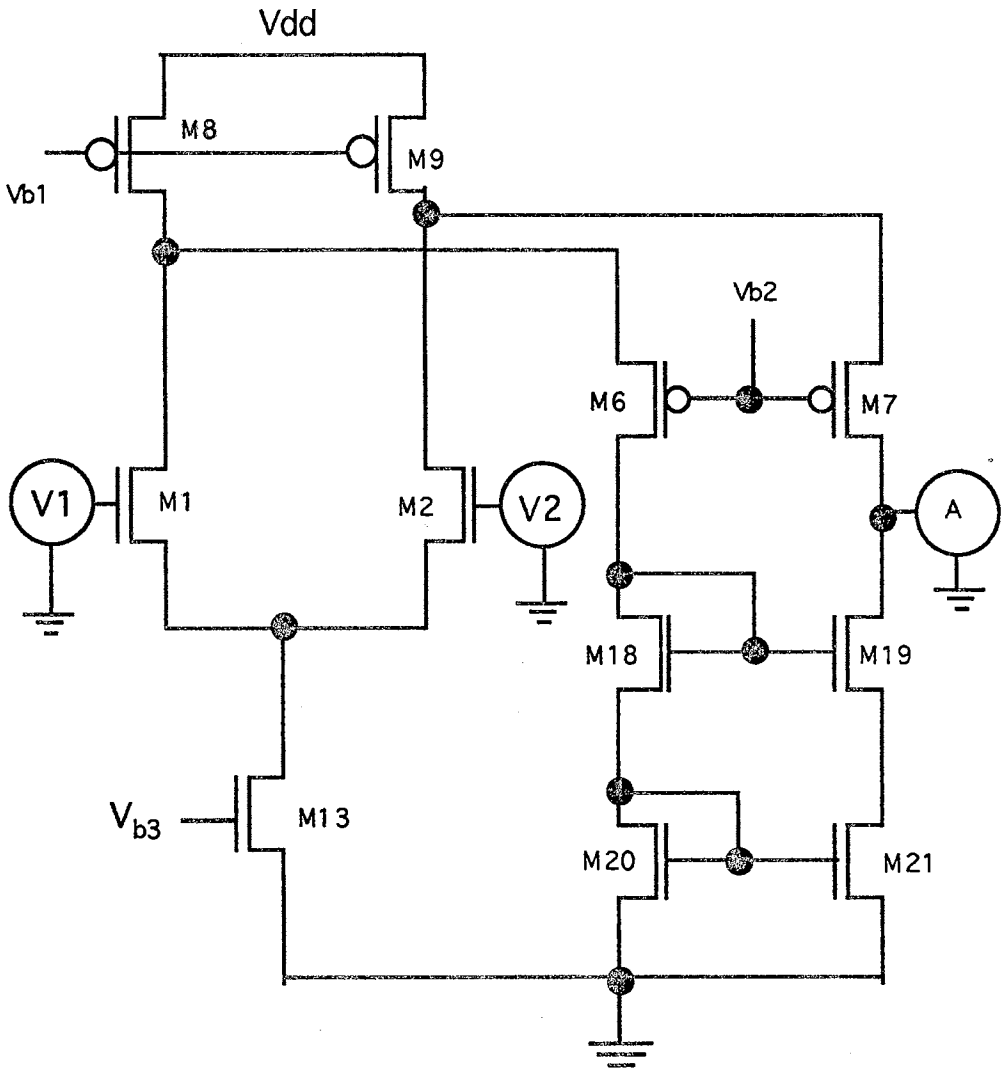


Figure 31. Simplified Folded Cascode Schematic Diagram

First, let us analyze CMR (common mode range) which is the input signal range for which both transistor M1 and M2 are in the saturation region. The positive CMR can be expressed as

$$CMR\ + = Vdd - \Delta V_8 + V_t \tag{3-1}$$

where ΔV is overdrive voltage, V_t is threshold voltage, and Vdd is the power supply.

Similarly, the negative CMR is

$$CMR\ - = \Delta V_{13} + \Delta V_1 + V_t \tag{3-2}$$

From (3-1) and (3-2), we see that the input CMR range is wide.

One disadvantage of this cascode amplifier is that the voltage swing is not very good due to the configuration of the cascode output.

The low and high limitation of V_{out} is given as

$$V_{out}(\min) = \Delta V_{19} + \Delta V_{21} \tag{3-3}$$

$$V_{out}(\max) = V_{dd} - \Delta V_9 - \Delta V_7 \tag{3-4}$$

Figure 32 shows the dc characteristics of the amplifier. As is observed, the range of the signal swing is not rail to rail. The digital circuitry, however, follows the amplifier so that the aforementioned problem is not critical in our design.

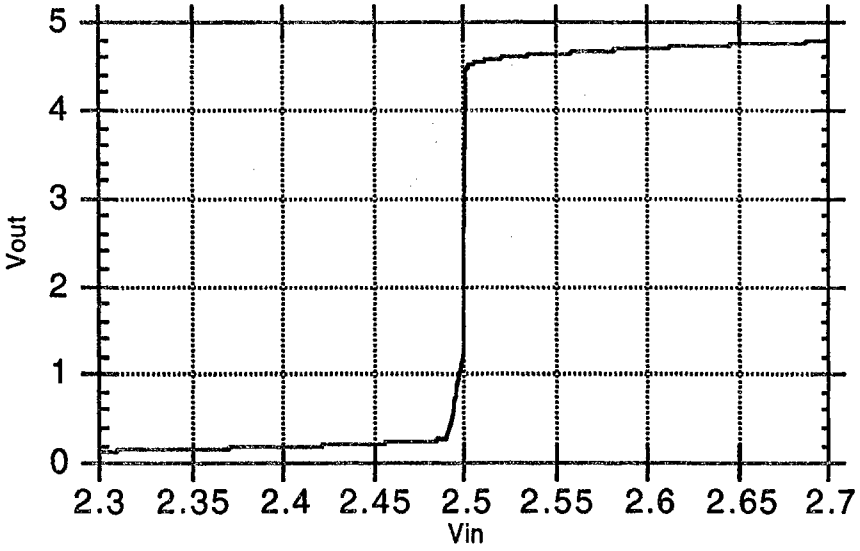


Figure 32. Open-Loop DC Transfer Characteristics

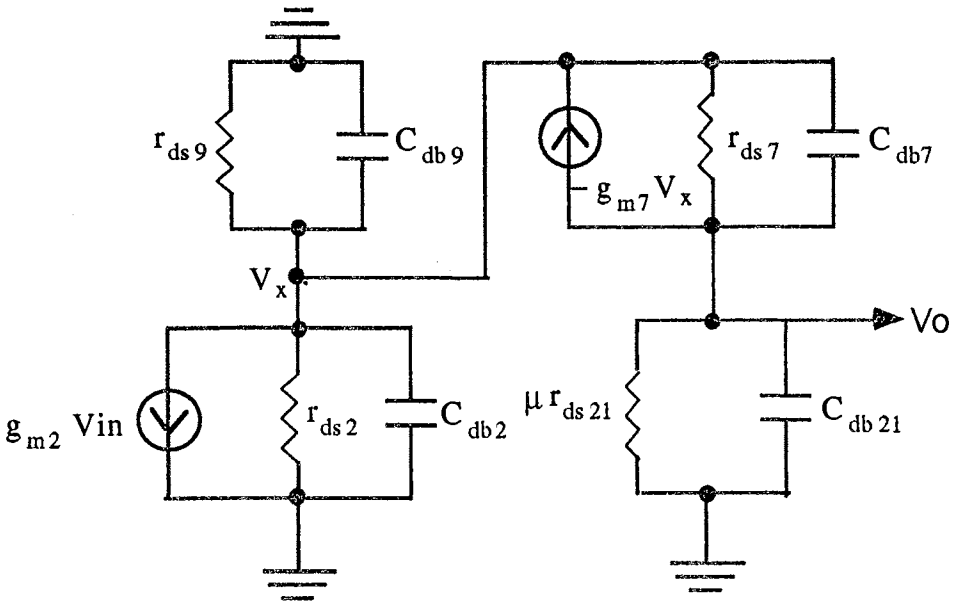
Next, the small signal performance of the cascode amplifier can be analyzed using the small-signal model of Figure 33(a) which has been simplified in Figure 33(b). In this analysis, only half the circuit is considered due to the symmetrical characteristics of the amplifier. It is assumed in the analysis that both sides of the amplifier are perfectly matched.

$r_{dst\,1} = r_{ds\,9} \| r_{ds\,2}$, $r_{dst\,2} = r_{ds\,7}$, $C_{dbt\,1} = C_{db\,2} + C_{db\,9}$, and $C_{dbt\,2} = C_{db\,7}$.

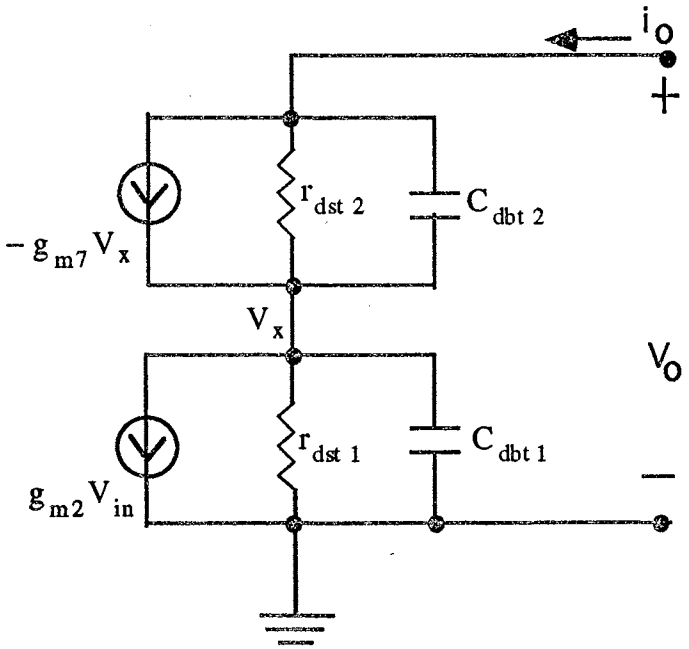
The node equations are

$$g_{m2} V_{in} + (g_{dst\,1} + sC_{dbt\,1}) V_x = -g_{m\,7} V_x + (g_{dst\,2} + sC_{dbt\,2})(V_o - V_x) \quad (3\text{-}5)$$

$$i_o = -g_{m7} V_x + (g_{dst\,2} + sC_{dbt\,2})(V_o - V_x) \quad (3\text{-}6)$$



(a)



(b)

Figure 33. (a) Small Signal Equivalent Circuit for Folded Cascode Amplifier (b) Simplified Small Signal Circuit

To obtain g_{meff}, V_o is set to zero, and substituting for V_x in eq. (3-6) results in

$$g_{meff} = \frac{i_o}{V_{in}} = \frac{g_{m2}(g_{m7} + g_{dst2} + sC_{dbt2})}{g_{dst1} + sC_{dbt1} + g_{m7} + g_{dst2} + sC_{dbt2}} \tag{3-7}$$

The above equation can be simplified by the assumption.

$$g_{m7} >> g_{dst1} + g_{dst2} \tag{3-8}$$

Finally,

$$g_{meff} \approx \frac{g_{m2}(g_{m7} + C_{dbt2})}{g_{m7} + sC_{dbt1} + sC_{dbt2}} \tag{3-9}$$

Hence, the transconductance of the amplifier is

$$g_{meff} = g_{m2} \tag{3-10}$$

The output resistance is

$$r_0 = \mu_7 r_{ds9} || \mu_{19} r_{ds21} \tag{3-11}$$

where $\mu = g_m / g_{ds}$

Finally, the dc gain is

$$A_v = g_{m2}(\mu_7 r_{ds9} || \mu_{19} r_{ds21}) \tag{3-12}$$

Next, the AC response of the amplifier is analysed. There are two poles in the transfer function as $P_1 = -\frac{1}{\tau_1}$ and $P_2 = -\frac{1}{\tau_2}$.

Assuming that $C_L >> C_{db7} + C_{db19}$, then the dominant pole is calculated as

$$\tau_1 = (\mu_7 r_{ds9} || \mu_{19} r_{ds21}) C_L \tag{3-13}$$

The non-dominant pole is approximately as

$$\tau_2 = (r_{dst1} || r_{dst2})(C_{dbt1} + C_{dbt2}) \tag{3-14}$$

Thus, GBP is

$$GBP = \frac{g_{m2}}{C_L} \tag{3-15}$$

Figure 34 shows the AC response of the amplifier. The dc gain is 50DB, and the phase margin is 80°. Figures 35 through 38 show the closed-loop transfer characteristics, slew rate, common mode gain and phase response, and and power supply ripple gain.

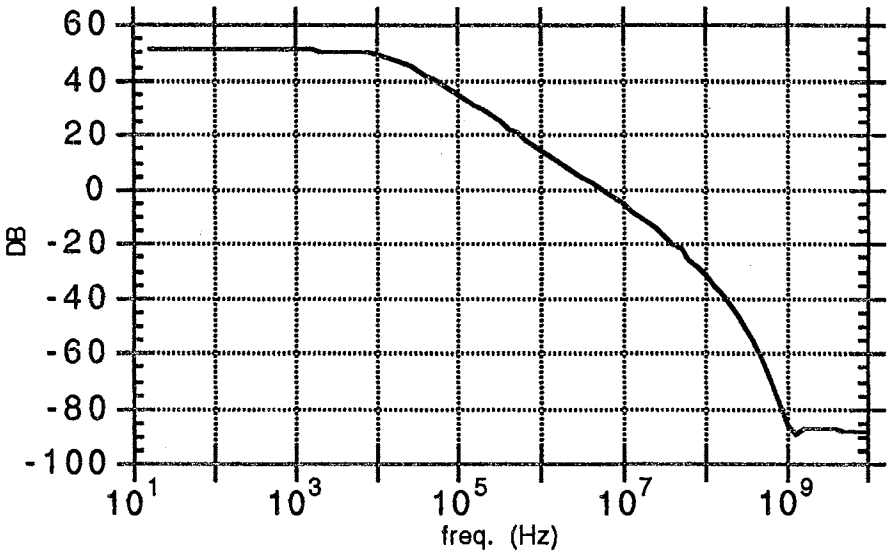


Figure 34. Open-Loop AC Transfer Magnitude Response

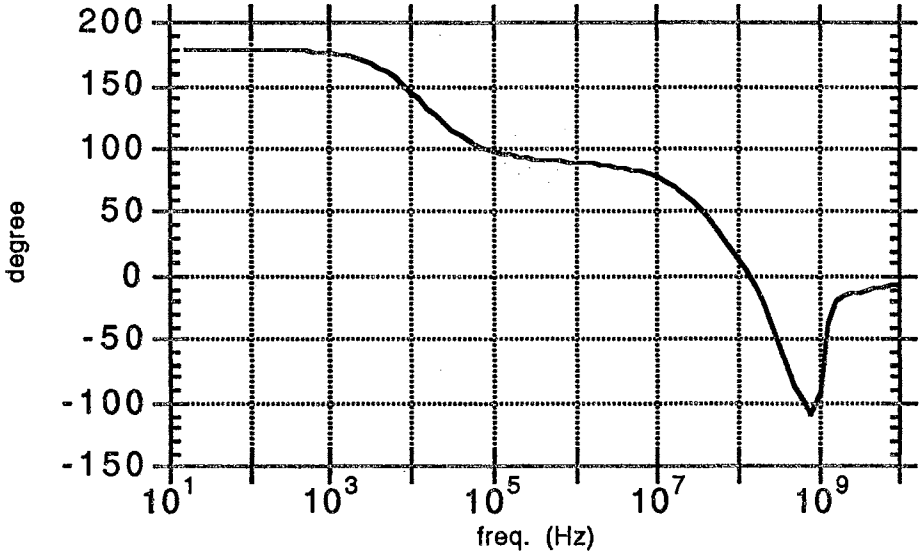


Figure 35. Open-Loop AC Transfer Phase Response

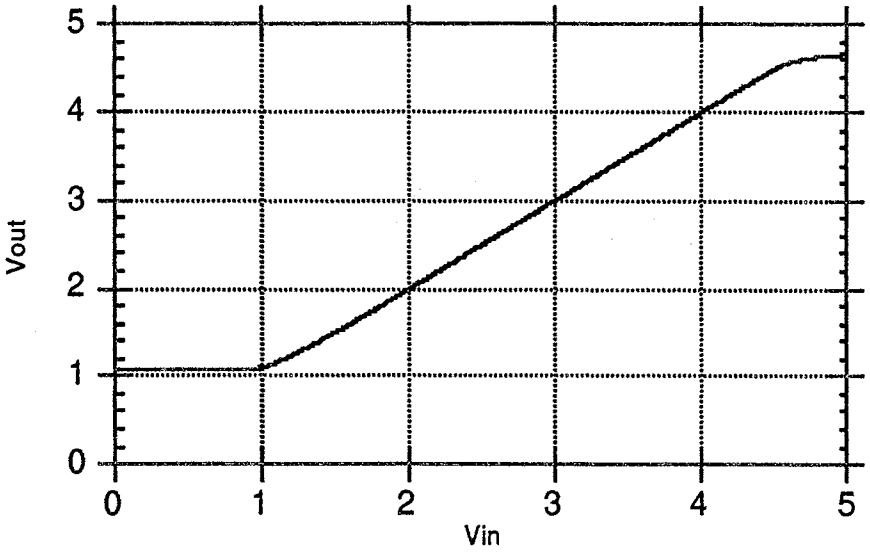


Figure 36. Closed Loop Characteristic Response

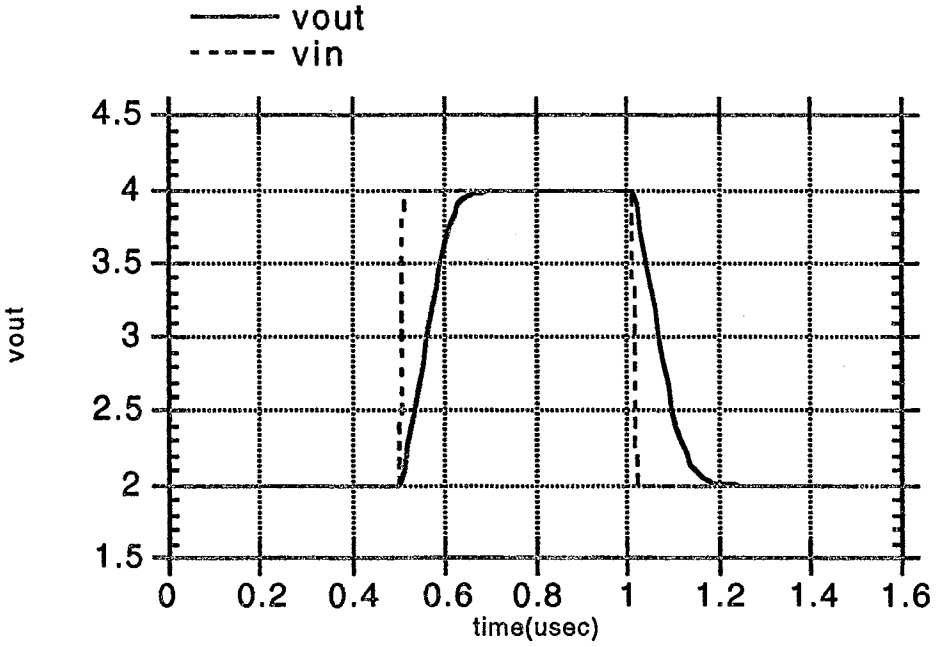


Figure 37. Slew Rate

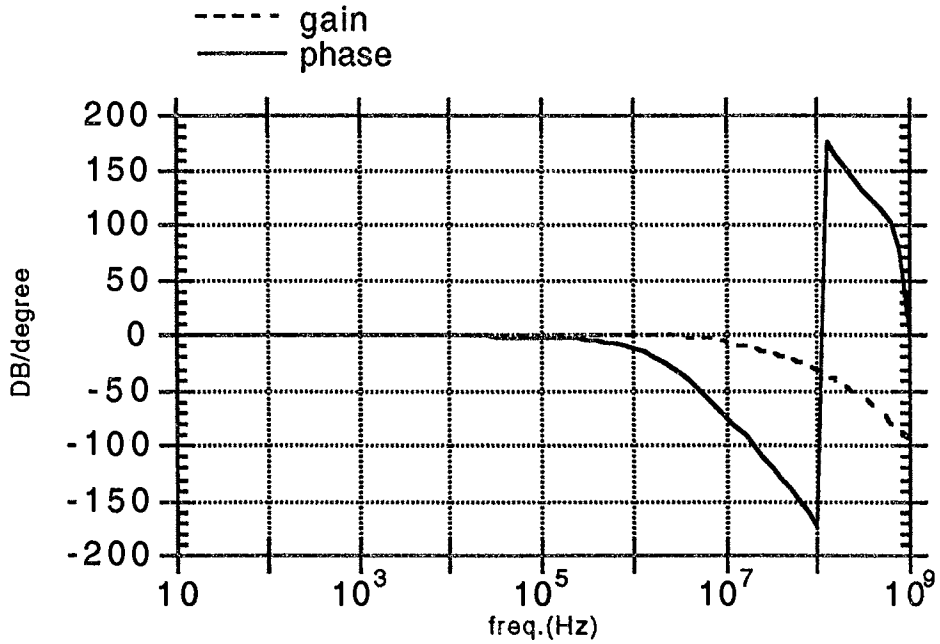


Figure 38. Common Mode Gain and Phase Response

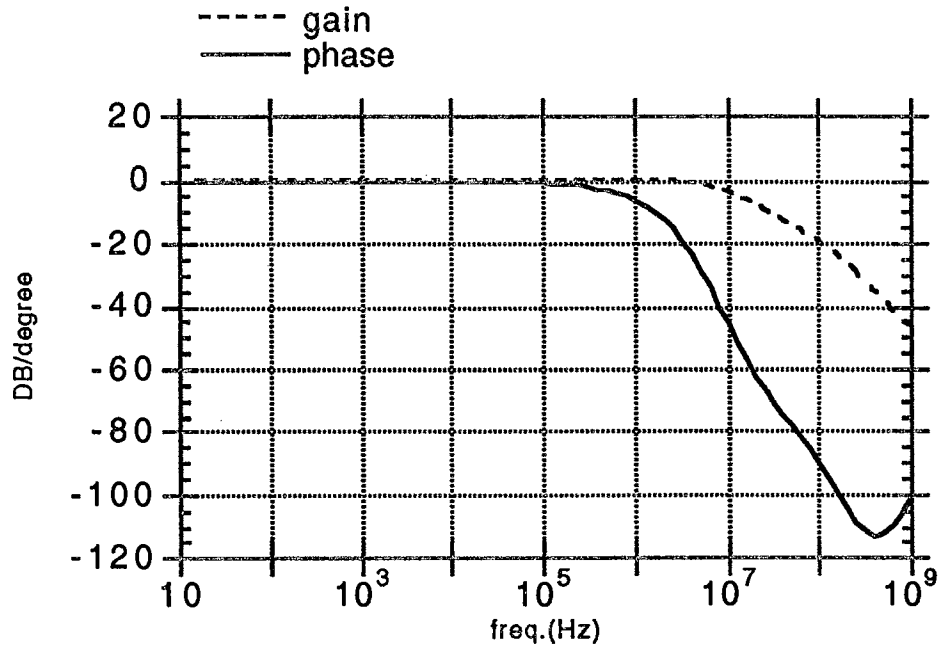


Figure 39. Power(Vdd) Gain and Phase Response

The following table summarizes the folded cascode amplifier characteristics.

TABLE III

SPECIFICATION OF DESIGNING THE DIFFERENTIAL AMPLIFIER

Specifications	value
Supply Voltage	0 to +5
Input Bias Current	50uA
Input CMR	+5 to 1.4V
DC Voltage Gain	50DB
Phase Margin	80^{o}
Output Voltage Swing	0.4 to 4.6Volt
CMRR	50DB
PSRR+	50DB
GBW Product	10^{7} Hz
Load Capacitance	2pF

PEDC Circuit Design

The PEDC circuit shown in Figure 40 has been employed for scanning, which is indispensable for the closed loop trimming operation of an analog memory. The objective of the PEDC is not only to activate the programming at the beginning of the operation, but also to terminate the programming when a node of interest is closed to a reference node within a predetermined range.

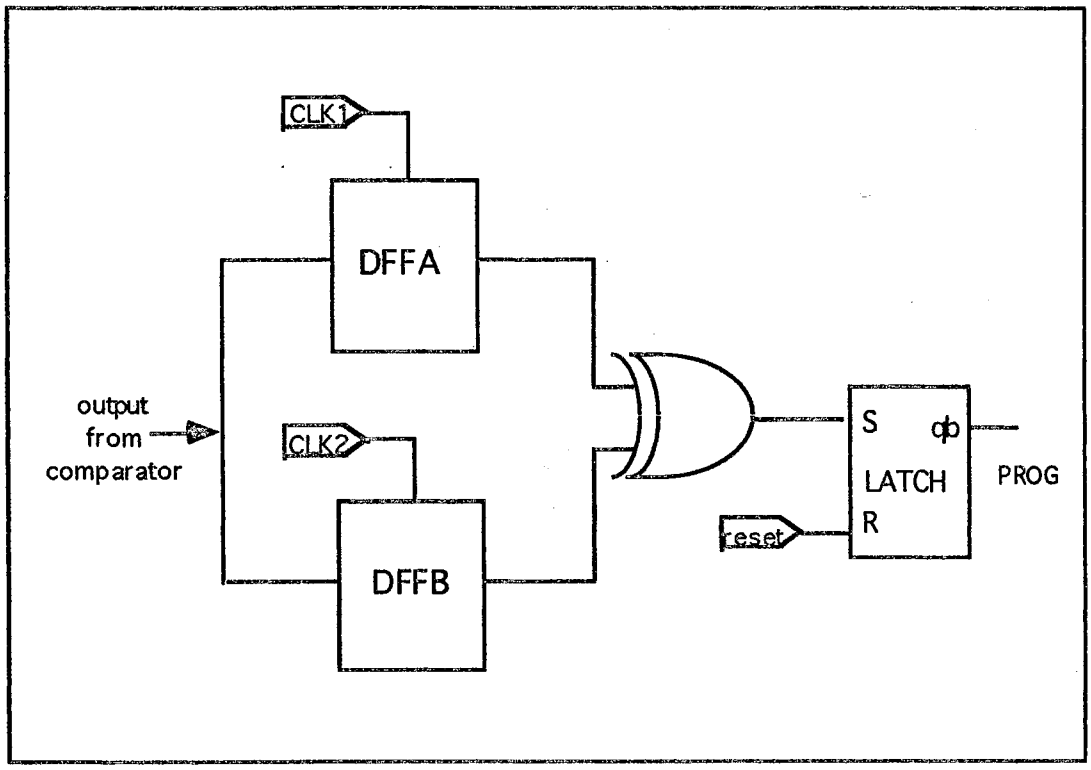


Figure 40. A PEDC Circuit Diagram

The PEDC consists of a XOR, two D-type flip flops, and another latch circuit. It is assumed that the DFFA, one of two D-type flip flops, latches the output state of the comparator shown in Figure 40 while the clk1 is on. The state will be kept constant until the next clk1 signal arrives. Similarly, the DFFB reads the output state of the comparator while the clk2 is on. As the DFFA does, it contains the value until the next clk2 arrives. The first and the second states read from DFFA and DFFB define the previous and current state, respectively. Assuming that the gain of the amplifier is high, two states will be +5volt or ground except during just short transition term.

During the interval between the clk1 and the clk2, the system provides one programming pulse into a control gate(an injector) with an injector(a control gate) to ground. The digital level of the current and previous state can be different since the programming pulse changes the output of the comparator. Then, the programming operation will be ended. If the states are identical, the programming operation should keep going until the state change is detected..

In order to recognize the state change behavior, the current and the previous state are written into the XOR circuit driven by the clk1 and clk2. According to the input states, the output from the XOR logic will be "1" or "0". The XOR gate provides a logic 1 output when any one, but only one, of its inputs is in the logic 1 state. Thus logic "1" indicates that the previous and the current state are different. Consequently, it immediately disables the "prog" signal shown in Figure 40 so that the programming operation will be terminated. Otherwise, the signal "prog" is kept high, and the programming operation proceeds without interruption. As we have seen so far, the initialization and the termination does not depend on the absolute, but the relative value of the current and the previous state. It will eliminate the unnecessary erasing operation during the programming. This is a very interesting feature applicable

to the weight control in a neural network because the weight change on each neuron is usually negligible during the weight changes.

The roles of the latch circuit following the XOR circuit is not only to initiate the "prog" signal being on but also to hold "prog" signal high until the XOR detects the state change. The function of the latch circuit is shown in following TABLE IV.

TABLE IV

TRUTH TABLE OF LATCH CIRCUIT

Set	Reset	Q	Qbar
0	0	Q	Qbar
0	1	0	1
1	0	1	0
1	1	0	0

For proper activation of the enabling signal "prog", we apply one short term pulse into the reset terminal of the latch circuit. Provided that the "reset" signal is high, the initial setting of the latch is to begin Q with "0" and Qbar with "1" under the assumption that the output of the XOR gate is low. The assumption

is reasonable because the two inputs of the XOR gate are the same before the programming operation. Next convert the "reset" signal state from high to low. It mandatorily keeps the "prog" signal to be high. Then the system begins to apply a consecutive programming pulse into the specific element of the array.

The on state of the "prog" will be abruptly changed as soon as one of the two inputs of the XOR is changed. It means that the floating gate voltage of the node of interest is closed to the voltage of the reference node within a predetermined voltage range. Then, the system will terminate the programming operation. Figure 41, a IRSIM result, shows the situation.

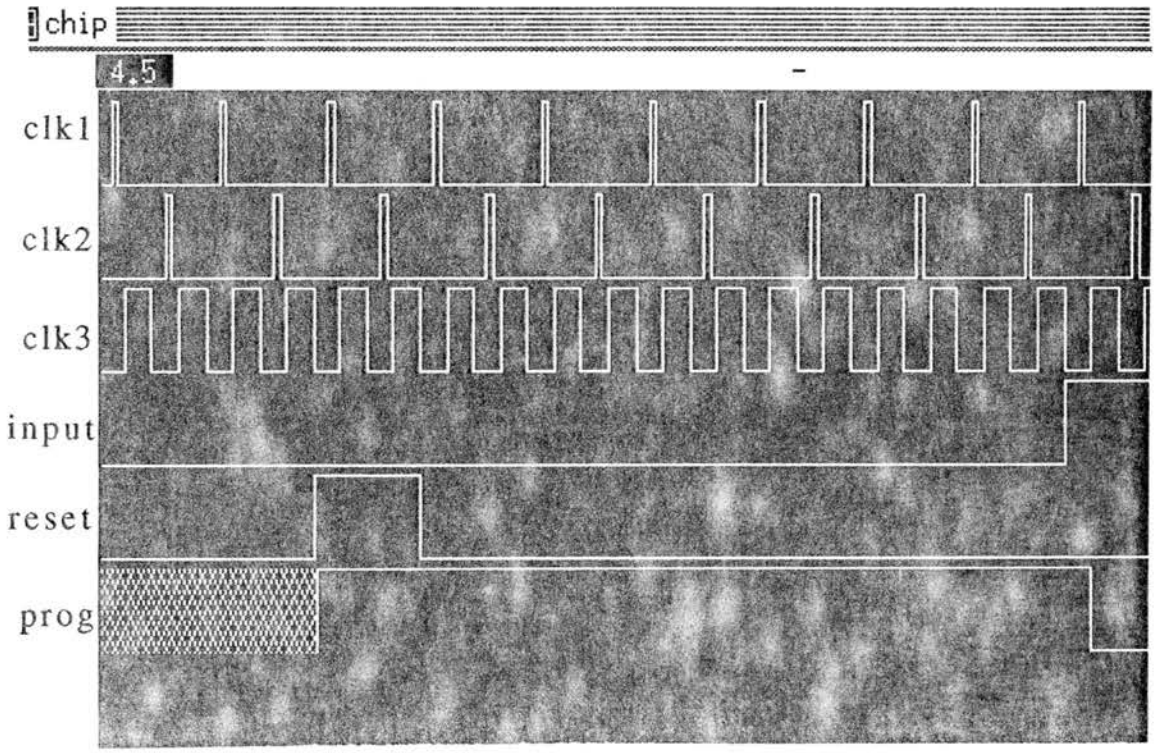


Figure 41. PEDC Simulation

For proper working of these operations, the clocking scheme is a very important feature. The system has been constructed with three different clocks; clk1 through clk3. While "clk1" and "clk2" are used for transferring the comparator output states into the XOR circuit, "clk3" is for stimulating HVPG. The clk3 basically occurs during the interval of clk1 and clk2 in order to achieve the proper operation of the XOR. Note that the clk1, the clk2, and the clk3 are the non-overlapping clocks. The clocking scheme must be observed in order to prevent the system from the misinterpretation of the floating gate voltage. Otherwise, a fraction of the programming voltage across the gate oxide will be added to the real voltage induced by the charge on the floating gate, and the obtained result will be far from the actual value of the floating gate voltage. Consequently, the system terminates the programming improperly. The other fault also can be imagined during the settling of electrons in a floating gate after a programming pulse is applied. This is shown in Figure 23 in chapter III. Hence, the time interval, t1 plus t2, is required to avoid having the system measure the floating gate voltages during the convergence. We have preferred the short term pulse so that efficient injection takes place in the periods of the programming. Because it has been observed that the convergence time is proportional to the amount of charges flowing through the gate oxide[Carley, 1989]. Figure 42 illustrates three clocks, and the change of the floating gate voltage during the programming operation.

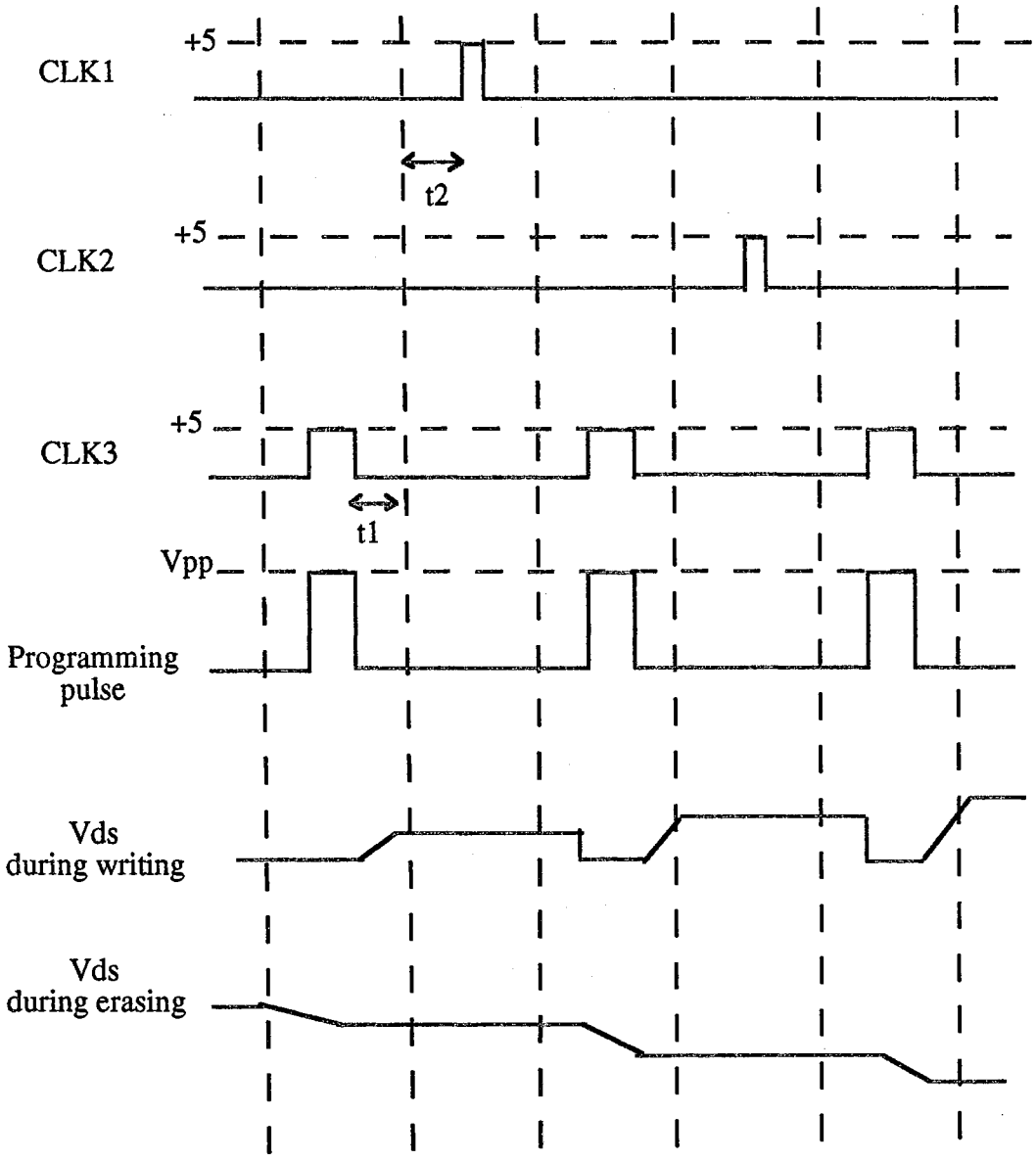


Figure 42. Overall System Clock Diagram

PTDC Circuit Design

As mentioned earlier, there are two different types of programming operations: writing and erasing. The writing operation is required when a floating gate voltage of interest is higher than a reference voltage. During the writing operation, a positive high programming pulse will be applied to a control gate of the relevant cell. On the other hand, the erasing operation is necessary when the selected floating gate voltage is less than the reference voltage. During the erasing operation, a positive high programming pulse will be applied to an injector.

The PTDC shown in Figure 43 has been provided to differentiate the writing from the erasing operation during the programming operation.

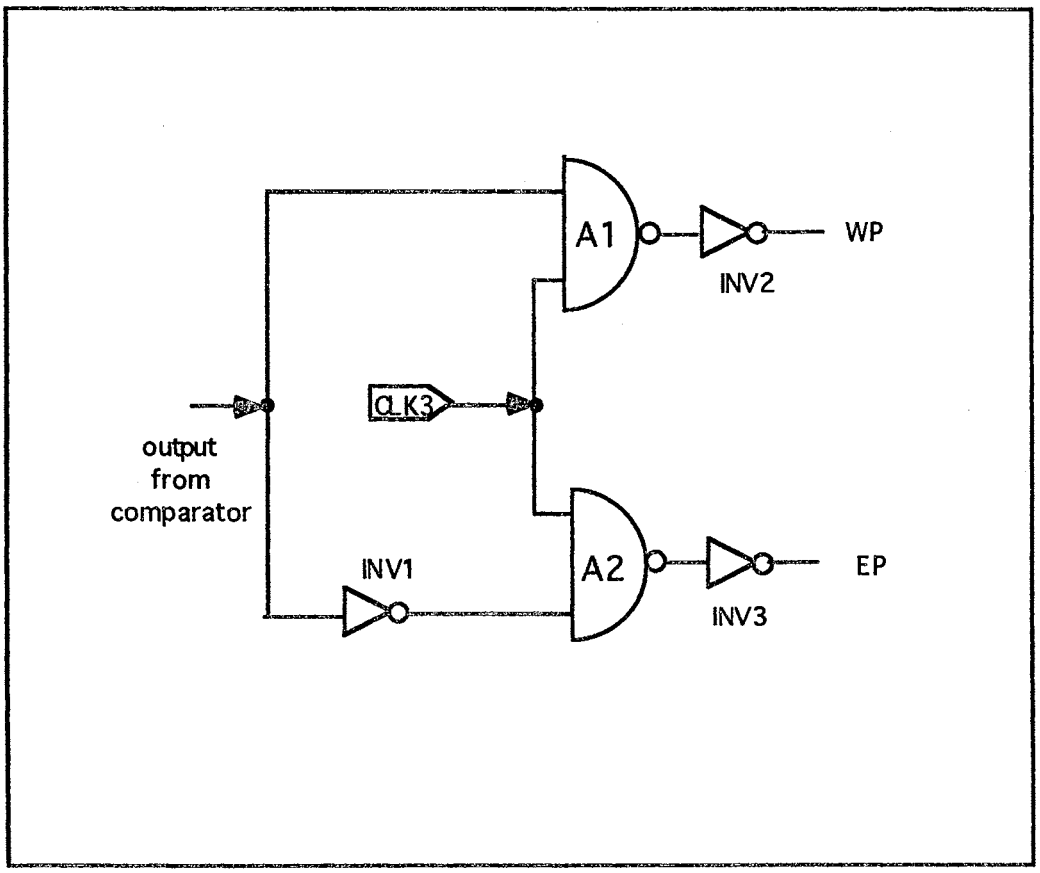


Figure 43. A PTDC Circuit Diagram

The strategy to decide the type of the operation is very simple. For example, let us assume that a floating gate voltage is less than a reference voltage. At this time the output of the high gain comparator becomes zero. Once the situation is detected by the inverter, inv1, shown in Figure 43, "EP" signal will be enabled, and "WP" is disabled. The "EP" and "WP" are the system level control signals which have been used for activating the erasing and the writing operation, respectively. It will assign the output of A1 shown in Figure 43 to be ground. On the other hand, the A2 passes the "clk3" with digital voltage level. The pulse and ground are sent into the HVPG placed in the column and the row of the array to activate the high programming pulse and ground, respectively. Finally, the high programming pulse from the HVPG located in the relevant column is applied to the injector, and the control gate continues to be tied to ground. To the contrary, when the floating gate voltage is larger than the reference voltage, then the pulse will pass through A1. Then the "WP" stimulates the HVPG located in the row. Then, the programming pulse will be applied to the control gate with the injector grounded. By this way, the bi-directional programming is feasible in the memory.

HVPG Circuit Design

A HVPG has been developed to selectively route high-programming voltage to a floating gate structure. The crosscoupled circuit shown in Figure 44 is a key circuit for the routing. The "Vin" designated on the input terminal is a conventional digital output level signal driven by a typical digital circuitry. Then, the maximum voltage level of the crosscoupled circuit is Vpp(around +15). Figure 45 shows the inverted output voltage, Vout, according to a particular Vin.

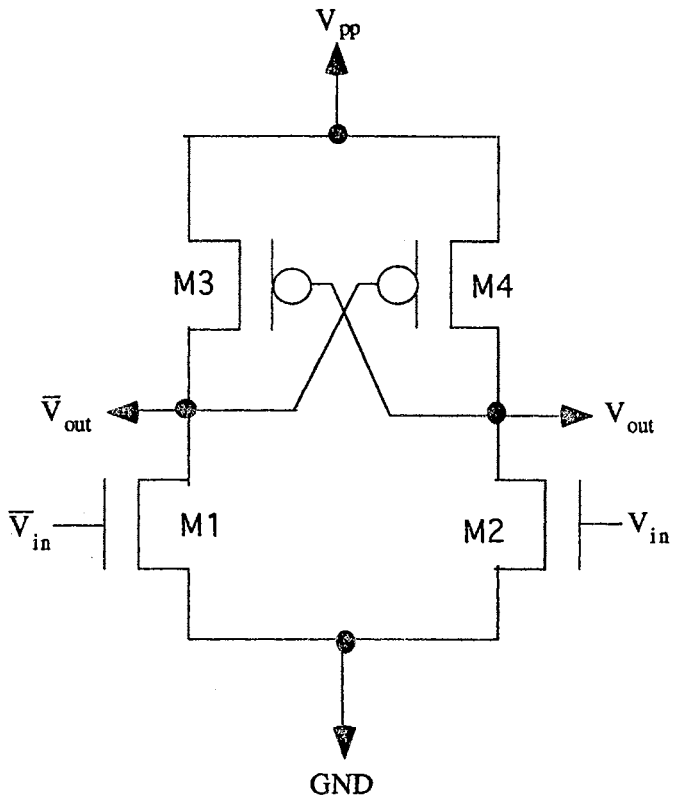


Figure 44. A Crosscoupled Circuit

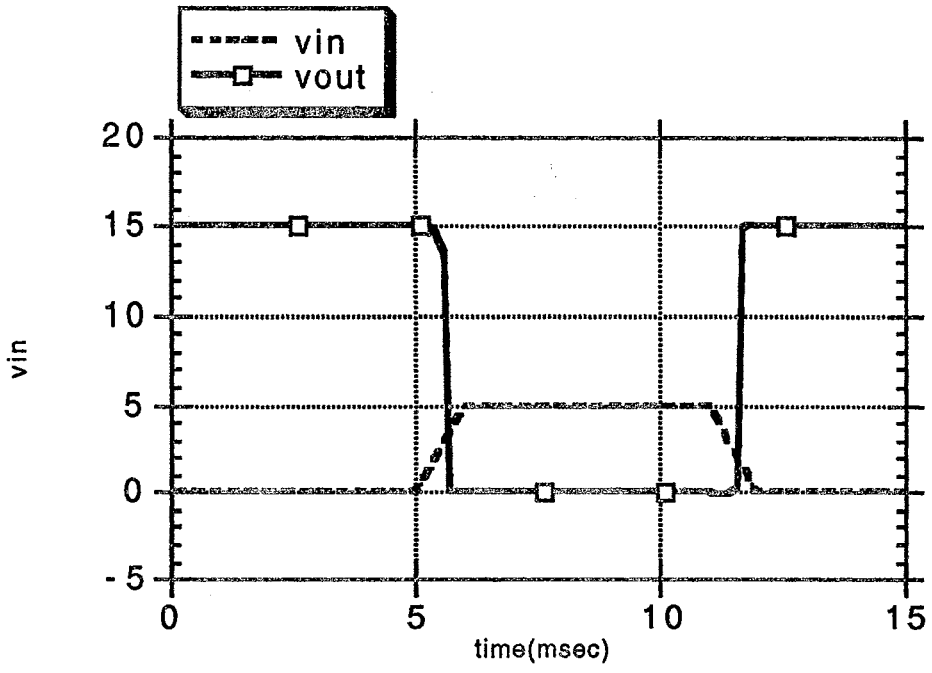


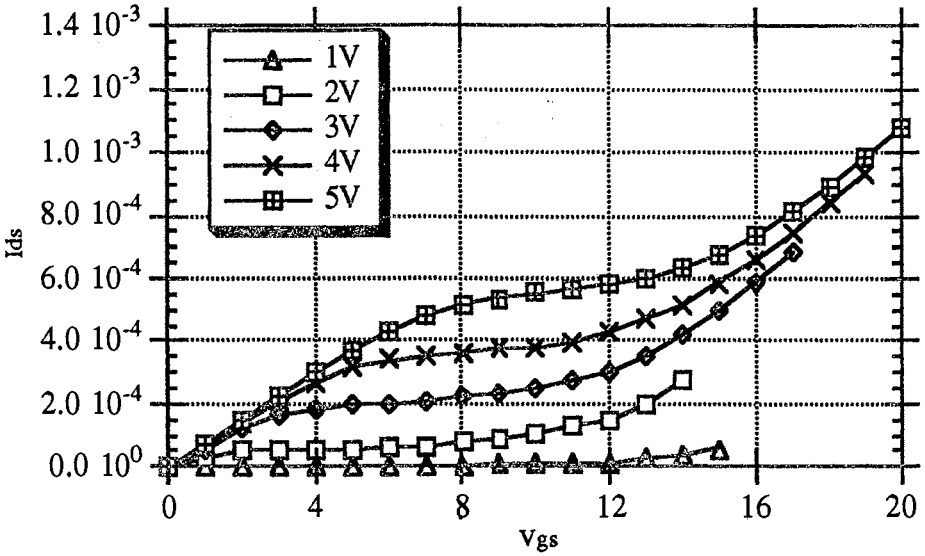
Figure 45. Vout vs. Vin for Crosscoupled Inverter

It should be noted that the magnitude of the voltage output from the crosscoupled circuit is high enough to induce a junction breakdown voltage, which usually disrupts the normal operation of the circuitry fabricated by a CMOS standard fabrication process. The high voltage level pushes the transistors, M1 through M4 shown in Figure 44 to be in the adjacent region of the junction breakdown voltage. Many techniques have been introduced so far for reducing the problem. One way is to utilize a lower doping concentration of a drain in a transistor because the junction breakdown voltage is inversely proportional to the impurity concentrations of the P and N regions[Mann, 1990]. The technique seems to be working very well. However, it not only requires a special fabrication process, but also occupies a very large area with a donut shaped structure. McConell[McConnell, 1991] uses a translation circuit, which looks like the above crosscoupled circuit. In the McConell design, the translation circuit, however, contains four transistors and two midway voltages between the upper and lower high-voltage rails. The principle of McConell's idea is that the high programming voltage is distributed into four transistors so that the voltage level to each transistor should be diminished. The circuit designed by us becomes to be significantly simplified compared to the McConell's circuit due to the removal of those four transistors and midway bias voltages from the crosscoupled circuit. The local enhancement layout technique mentioned in chapter III contributes to the reduction of the programming voltage level. So the translation circuit does not require the additional transistors, which release some burden of load. However, the voltage level still seems to be very dangerous to the whole circuitry. Thus, instead of the extra transistors and two more DC bias voltages, we adopted a long channel transistor which is supposed to be a high junction breakdown boundary. It arises from the fact that

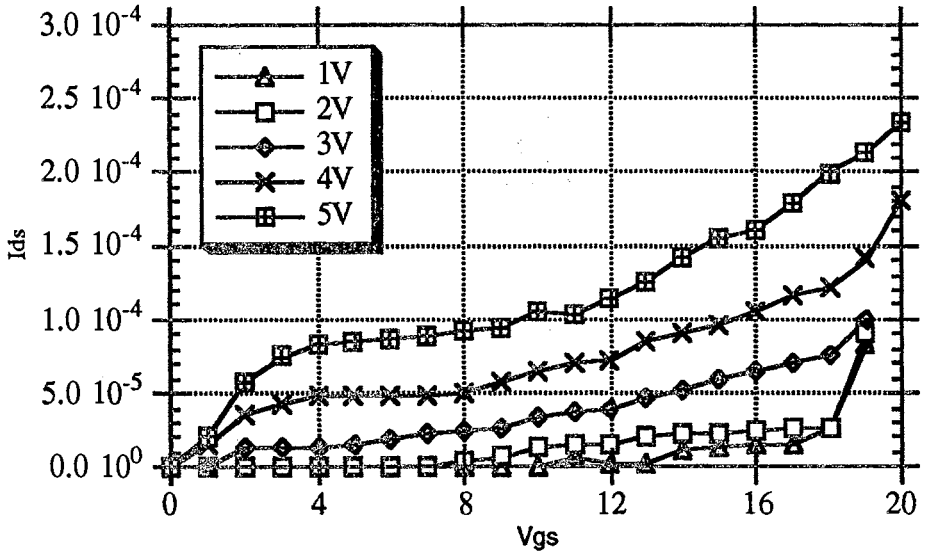
the drain-depletion region in a short channel causes a drain-to-source current to rise by increasing a drain-to-source voltage[Gray, 1988]. In order to prove the statement, we designed and tested two types of transistors in several die areas; one transistor with the width and length ratio is 4/2, and the other is 4/8. The graph shown in Figure 46 illustrates that the longer the channel is, the higher the junction breakdown boundary. By using the characteristic, the HVPG circuit implemented with the long channel device can drive Vpp without any perturbation induced by the breakdown voltage.

In addition to the Vpp, the HVPG also generates Vmid(around +8V) which is used for preventing an unselected cell from being programmed by the high programming voltage. Without the intermediate voltage between the rails, some cells will be disturbed during the programming operation. The 2 by 2 array shown in Figure 47(a) demonstrates the situation clearly. The symbol representing a floating gate transistor is depicted in Figure 47(b).

Let us assume that the cell (0, 0) is chosen for the writing operation in the array. The programming pulse generated by the HVPG No.1 located in the first row and the first column will be applied to the cell (0,0). The second assumption is that all the other HVPGs except HVPG No.1 ties to ground. Then there should be charge injections in the cell (0,0) as well as the cell (0,1) since the voltage difference between the control gate and the injector is Vpp for both floating gates. In order to prevent the disturbance, the conventional memory designer has developed two pass transistors attached to each floating gate MOSFET[Yaron, 1982]. The structure of the memory seems to be working well. However, it definitely occupies a huge area, which has been a serious obstacle of the memory design with the EEPROM cell. In our array, there is no pass transistor isolating a cell of interest from the adjacent cells in the 2 by 2 array.

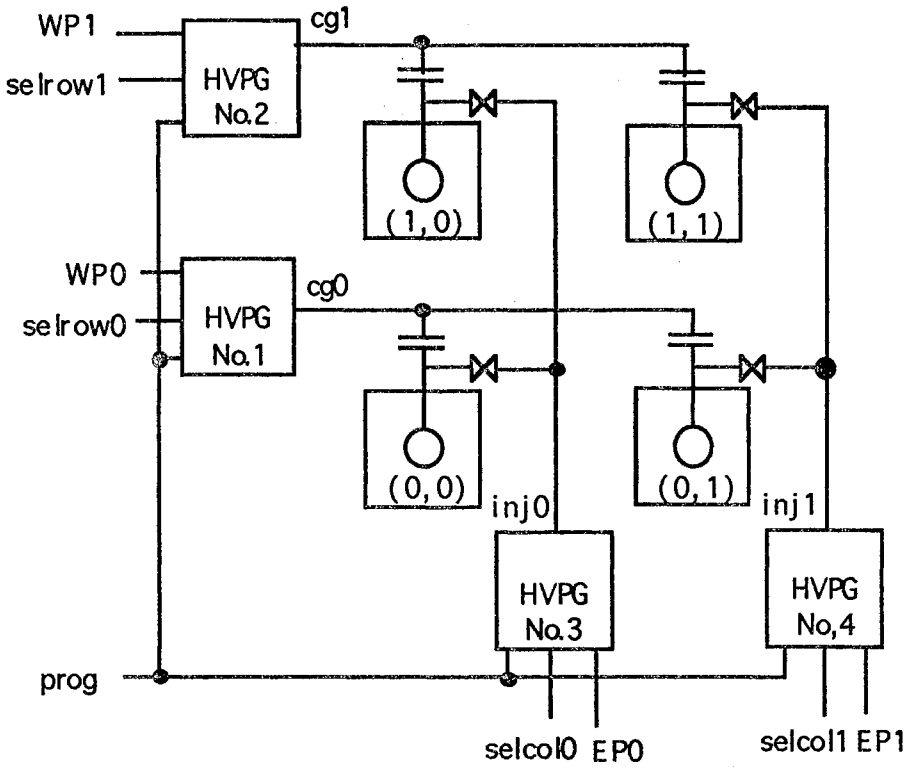


(a)

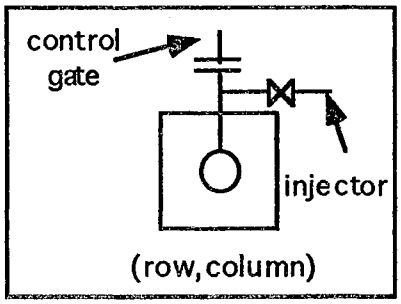


(b)

Figure 46. (a) Width-Length Ratio 4/2 (b) 4/12 Transistor Characteristic Function



(a)



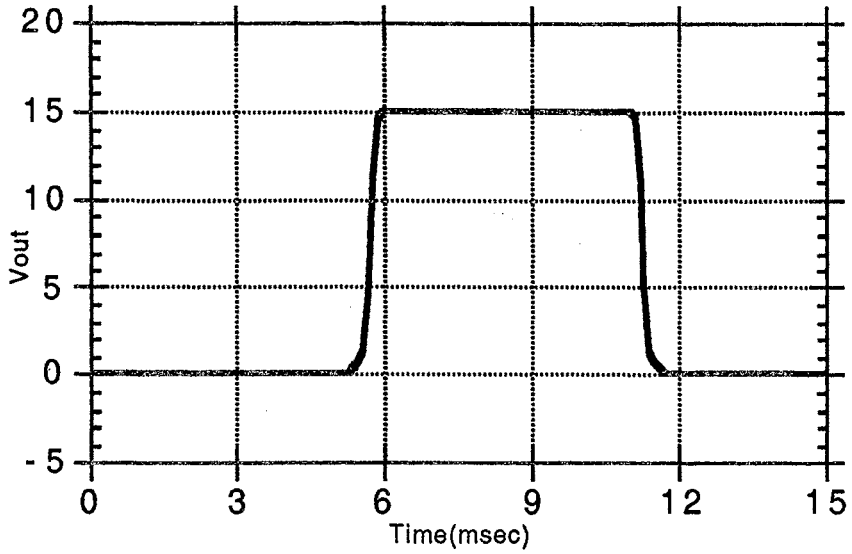
(b)

Figure 47. (a) Simplified Block Diagram of 2 by 2 Array
(b) Symbol Representing a EEPROM Cell

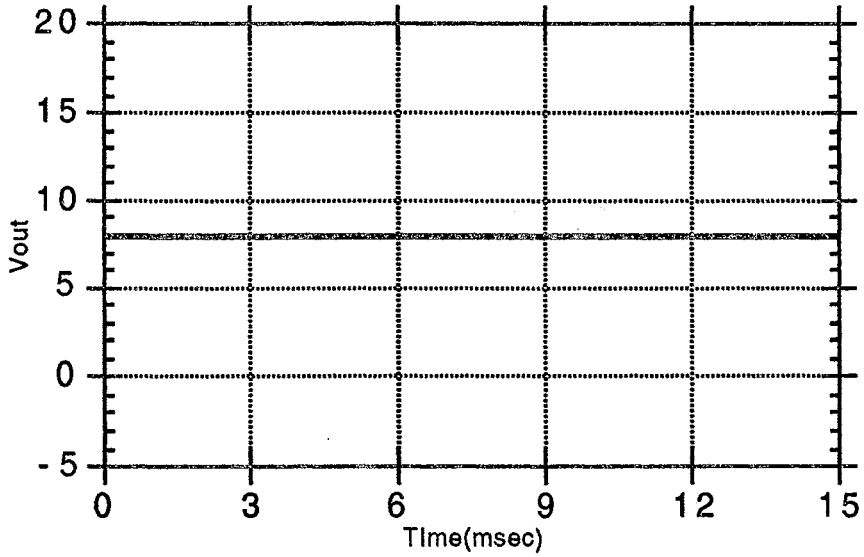
Instead of the extra transistors, we employ one extra bias voltage, Vmid. The voltage is the middle of the voltage between Vpp and ground. In the previous example, if the voltage Vmid is generated from the HVPG No.2 and No.4, then the voltage differences between the control gate and the injector of the cell (0,0), cell (0,1), cell (1,0), and the cell (1,1) in the array are 15V, 8V, 8V, and 0V, respectively. Consequently, the charge injection occurs only in the cell (0, 0). No other cells except the cell (0, 0) have a electric field high enough to initiate the Fowler-Nordheim tunneling injection. It will definitely enhance the scaling of the memory. As a trade-off, it requires an additional DC bias voltage. The four Pspice simulated outputs shown in Figure 48 correspond to the output from each HVPG in the above example.

One more interesting logical feature of the programming strategy is that the every column and row is connected to ground except during the programming operation to avoid the unnecessary power consumption. The following table V indicates the voltage of each HVPG during four different situations based on the above example. As you see, the only period in which the HVPG is activated is during the programming operation.

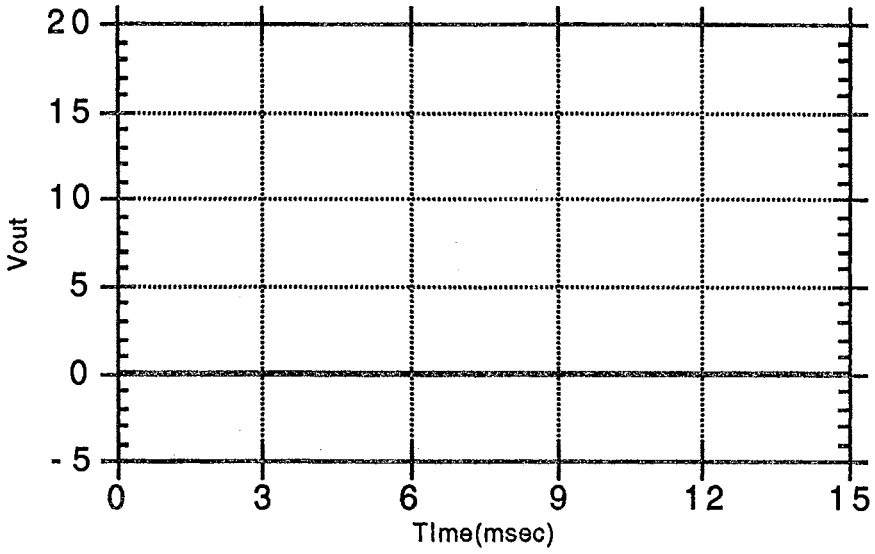
In order to satisfy the above algorithm with a hardware implementation, the HVPG basically demands a special circuitry shown in Figure 49. The circuit consist of two NMOS and PMOS transistors with three power lines, Vdd, Vmid, and ground. The p and the n-channel transistor with its drain connected to Vpp and to Vmid, is driven by input signal VinA. It will prevent both power lines(Vdd and Vmid) from being alive at the same time. When the digital voltage level "sel" signal from decoder is high during programming, then high programming voltage Vpp is delivered to the output of the circuit. Otherwise, Vmid will be passed to the output.


20
15
10
5
0
- 5
Vout
0
3
6
9
12
15
Time(msec)

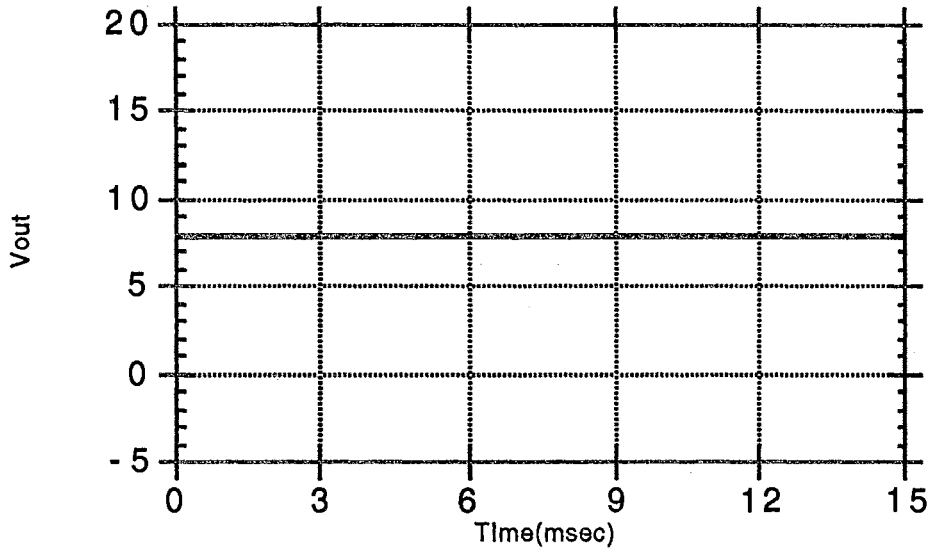
(a)


20
15
10
5
0
- 5
Vout
0
3
6
9
12
15
TIme(msec)

(b)



(c)



(d)

Figure 48. Programming Voltage Output Emitted from (a) HVPG No. 1 (b) HVPG No. 2 (c) HVPG No. 3 (d) HVPG No. 4 Based on the Example

Table V

OUTPUT OF FOUR HVPGS ON DIFFERENT OPERATION MODES OF ANALOG MEMORY BASED ON THE ABOVE EXAMPLE

	Idle	Read	Writing	Erasing
Row0	GND	GND	A pulse (Vpp)	GND
Row1	GND	GND	Vmid	Vmid
Col0	GND	GND	GND	A pulse(Vpp)
Col1	GND	GND	Vmid	Vmid

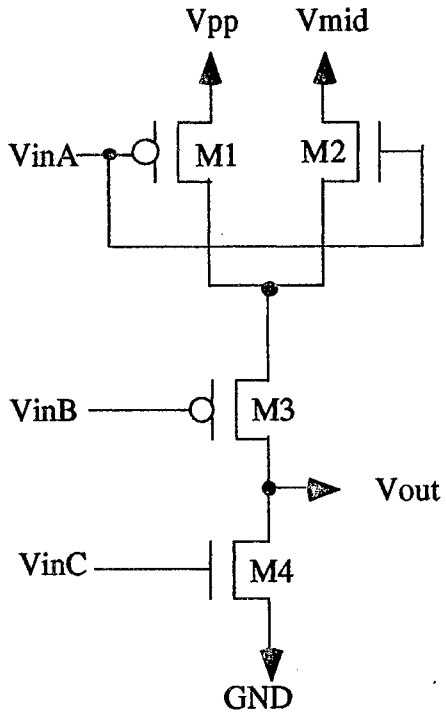


Figure 49. Circuit Diagram being Capable of Selectively Routing Three Different Voltages, Vpp, Vmid, and Ground

The scheme enables the HVPG to provide Vpp to only the selected row(column).

The purpose of the p-channel pass transistor, M3, which is placed in series between the above two transistor and the n-channel transistor, is to isolate Vpp and Vmid from ground while the output is tied to ground. The signal VinB is low during programming. To the contrary, the gate signal, VinC, associated with the transistor M4, should be high when the system is in an idle state. The disabled "prog" signal will make the signal to be high. The output of the circuit is also connected to ground during the period when the programming pulse is grounded.

In order to drive the all the transistors except M4 in the circuit, the crosscoupled circuit is indispensable since the voltage level is out of digital voltage boundary. Thus the two crosscoupled circuits are employed to convert the digital logic level into Vpp level. Additionally, some digital circuitry is also required to implement the HVPG for generating the gate signals of the crosscoupled circuits and the VinC shown in Figure 49. The following TABLE VI shows the digital logic scheme producing appropriate control signals.

TABLE VI

A TRUTH TABLE OF DIGITAL LEVEL LOGIC PRODUCING THREE INPUT SIGNALS VinA,VinB, AND VinC

	sel	WP(EP)	prog
Vpp	1	1	1
Vmid	0	1	1
GND	x	1	0

where the "x" is "don't care" mark.

Figure 50 shows the whole circuit diagram of the HVPG. The digital level logic gates are developed to satisfy the above requirements. The NAND gate enclosed by the dotted line contains two input signals, "WP(EP)" and "prog", which are common to every HVPGs. Thus, no matter what the size of the array is, the NAND gate will be placed just one time in a row and a column.

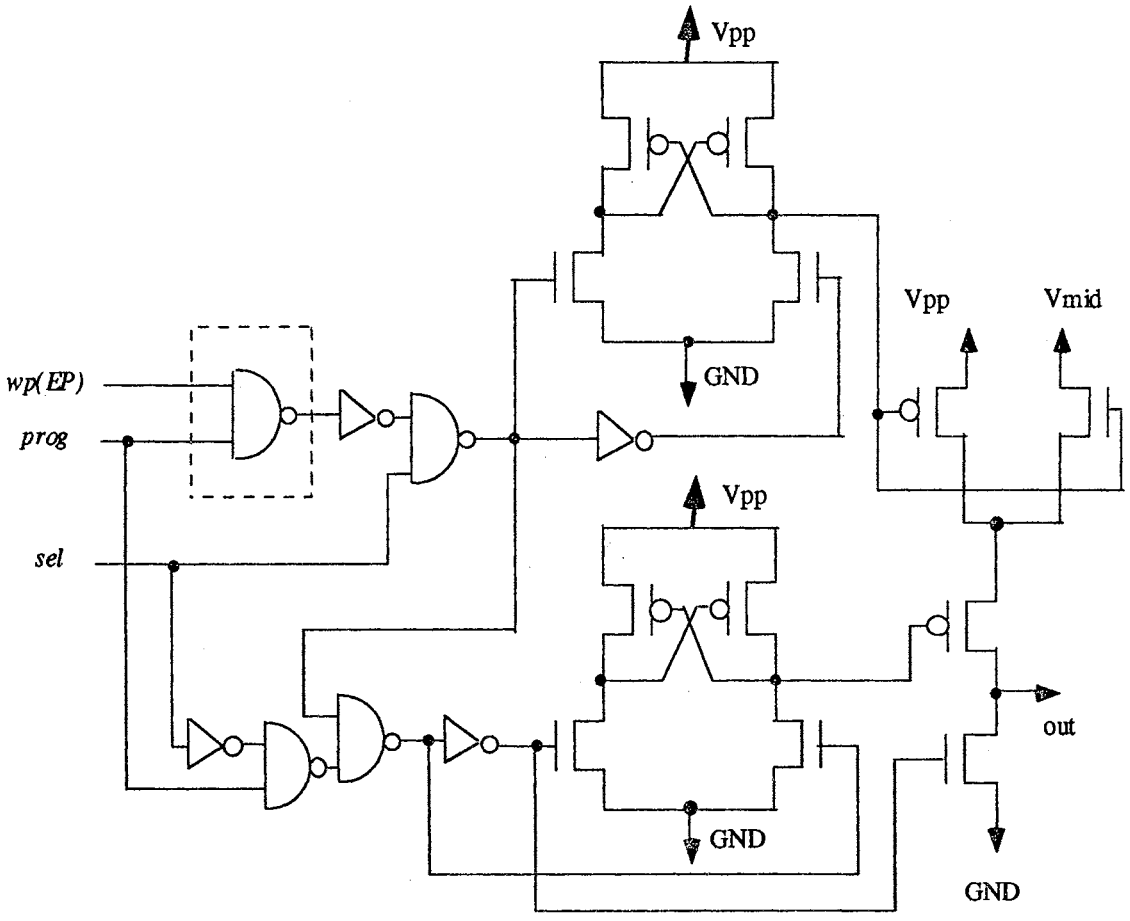


Figure 50. Overall Circuit Diagram of HVPG

Results and Discussion

Figure 51 shows the analog memory response to each reference input. We can see that the approximate linear relationship exists. The analog memory response does not follow the linear relationship for the range less than 1volt and greater than 4volt due to the negative feedback result shown in Figure 35. The data were obtained from averaging 25 different cases.

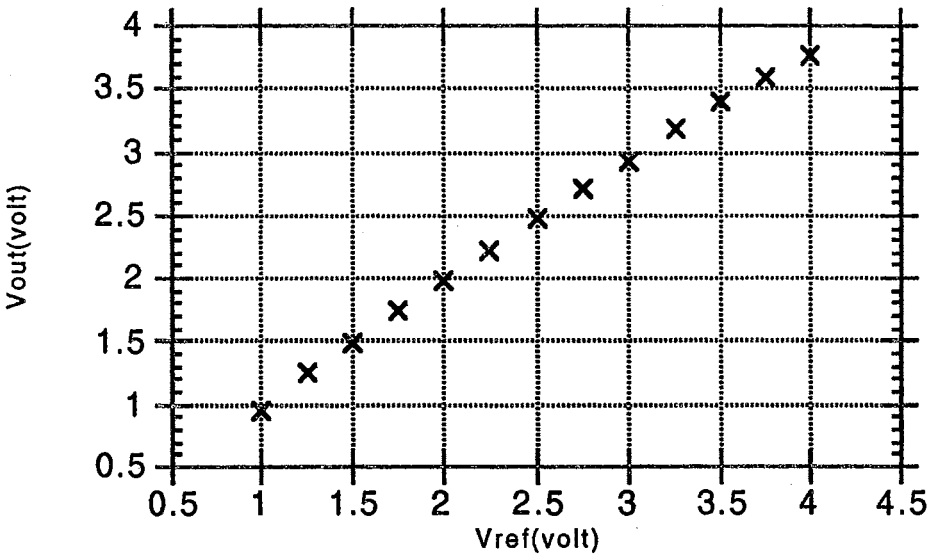


Figure 51. Analog Memory Response to Reference Input

Figure 52 illustrates the error occurred in each point. The Vdiff is defined as

$$V_{diff} = V_{ref} - V_{out} \qquad (3\text{-}17)$$

The data was obtained during the writing operation, during which output goes from high to low. It is noted that the errors are minimum when the reference voltage becomes 1.5volts, and gradually tends to increase in both directions. Between 1.25 volts and 2 volts range, the error of the memory is lower than the other range. It implies that the high resolution, of say 10mvolt, is valid only in the range 1.25volts to 2volts. It is equivalent to about 75 distinct levels, and it will substitute for 6 bit digital cells.

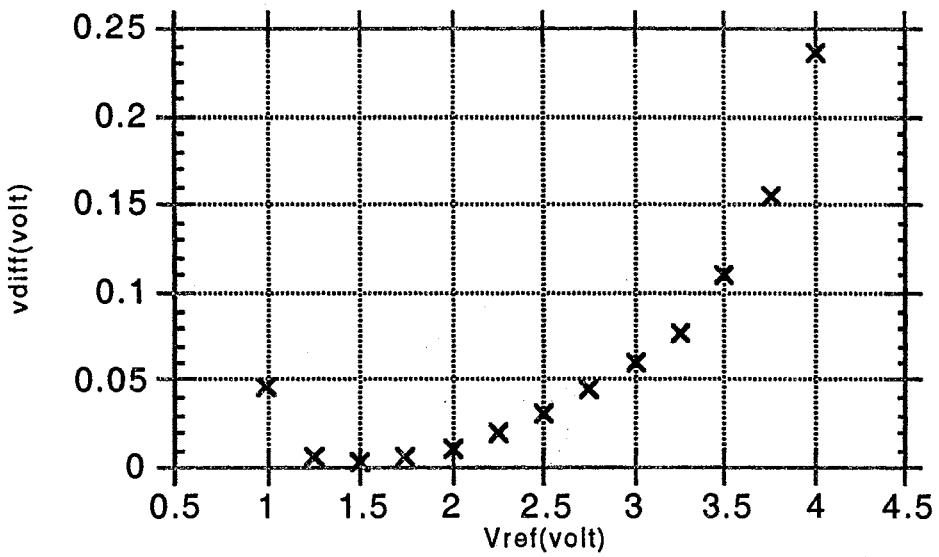


Figure 52. Magnitude of Mismatch between Stored and Reference Voltage

Figure 53 indicates the variance of the mismatch estimated in each reference point. From Figure 52 and 53, we can see that with the increase in reference voltage, the absolute error and the variance increase. One of the main parameters responsible for this is the common mode gain of the amplifier. The simulated common mode gain of the amplifier is already shown in Figure

37. According to the figure, the gain is almost zero. However, in the real amplifier, the common mode gain is found to increase as the reference voltage goes up. In order to decrease the common mode gain and increase the resolution of the memory, a cascode type of tail current source is required. The other reason for the increment of the error is due to the charge injected into the floating gate during programming. This is due to the fact that the electric field generated from the charges is in the opposite direction to the applied field. This will contribute to a high resolution due to an insignificant charge injection in the low reference voltage. However, the resolution on the high reference voltage is relatively bad. The influence of the varying electric field, somehow, is trivial compared to that by the gain. The reason will be explained later in this section.

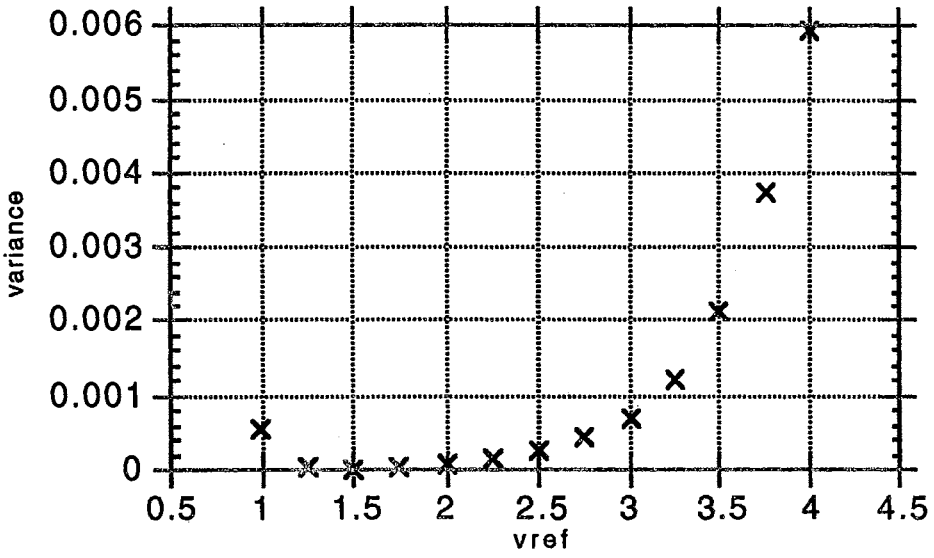


Figure 53. Variance of the Mismatch between Stored and Reference Voltage

The effect of the resolution dependence on programming magnitude is shown in Figure 54. The average error vs. pulse duration is plotted for a fixed programming voltage magnitude.

Avg. Error = (sum of each reference node error) / (number of nodes) (3-18)

In chapter II, it has been proven that the programming is logarithmically dependent to a pulse duration. The fact that the resolution of the memory with the longest pulse duration is the worst, matches our expectation. As seen, however, the resolution doesn't improve with shorter pulse duration than 1msec. The resolution becomes even worse as the duration increases beyond 1msec. It implies that there is another factor affecting charge injection during the programming. It can be concluded that there is a limitation to improve the resolution by decreasing the duration of programming pulse.

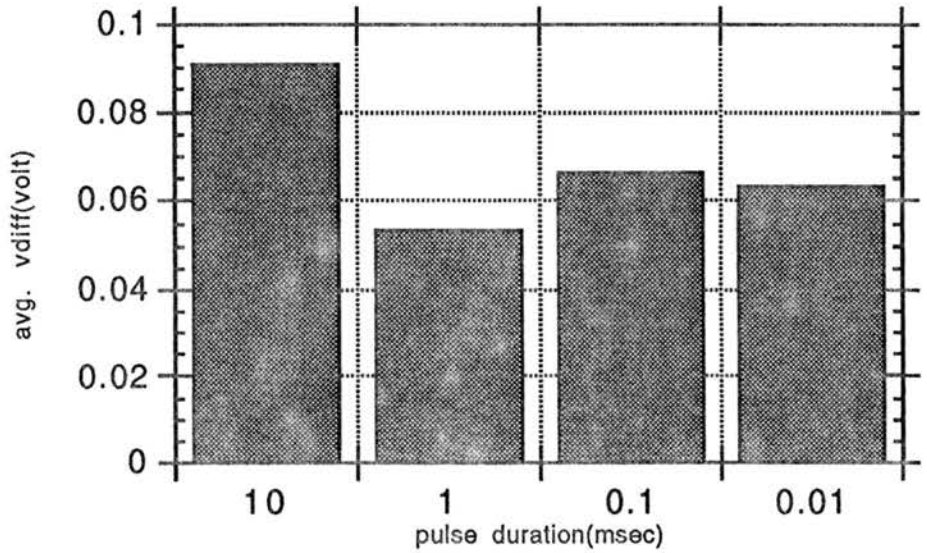


Figure 54. Average Error vs. Programming Pulse Duration

Figure 55 shows the total elapsed time in the same condition as the above. The time is estimated as follows:

Elapsed Time = total number of pulses * pulse duration during high (3-19)

In terms of Fowler-Nordheim tunneling mechanism, all the elapsed time should be equal since the charge injection occurs only during high programming pulse. The result corresponds to our expectation, except the fact that programming with the pulse duration of 0.01msec takes longer time than others. It may be due to the fact the time duration is too small to inject charge properly at this given programming voltage level. This was explained in an equation (2-14) in chapter II.

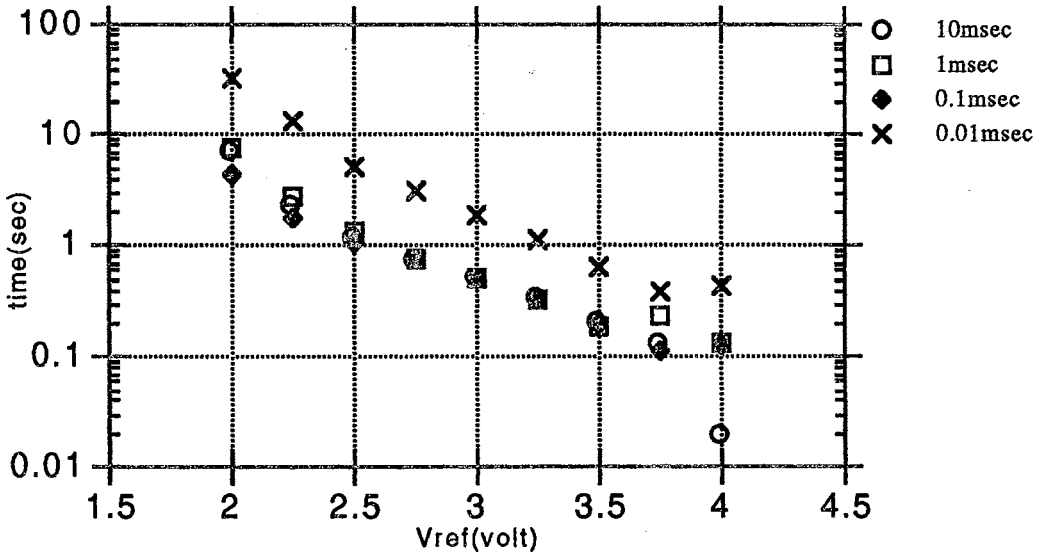


Figure 55. Elapsed Time vs. Programming Pulse Duration

Plot of the average error vs. magnitude of programming voltage is shown in Figure 56. The data are obtained with a fixed programming duration. As seen, the resolution decreases as the magnitude of the programming pulse for a given constant pulse duration increases. The result exactly matches our expectation. However, the absolute error between each magnitude is trivial in the interval from 14 to 16volts. It means that the magnitude of the programming pulse doesn't influence the resolution of the memory a lot except when the programming voltage exceeds 16.5volts. Meanwhile, the effect influencing the total elapsed time by the programming voltage magnitude is distinct. This is shown in Figure 57. Based on those properties, we can design a high speed analog memory since a high programming voltage magnitude reduces the programming time drastically without causing serious errors. But recall that the programming voltage level should not exceed the breakdown voltage of the internal circuitry.

Finally, it is shown in Figure 55 and 57 that the programming time increases as the reference voltage decreases during the writing operation. This is due to the increase of charges in the floating gate emitting the electric field, which is in the opposite direction to the applied field. It will lead to an increment of the programming time. To avoid this circumstance, it is recommended to implement an adaptive programming generator where the programming voltage magnitude is adjusted depending on the difference between a floating gate voltage and a reference voltage.

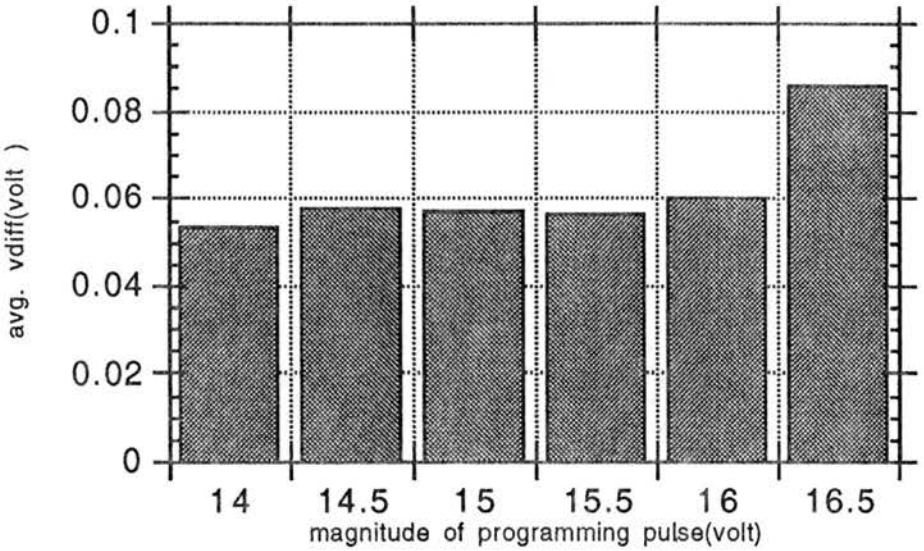


Figure 56. Average Error vs. Programming Pulse Magnitude

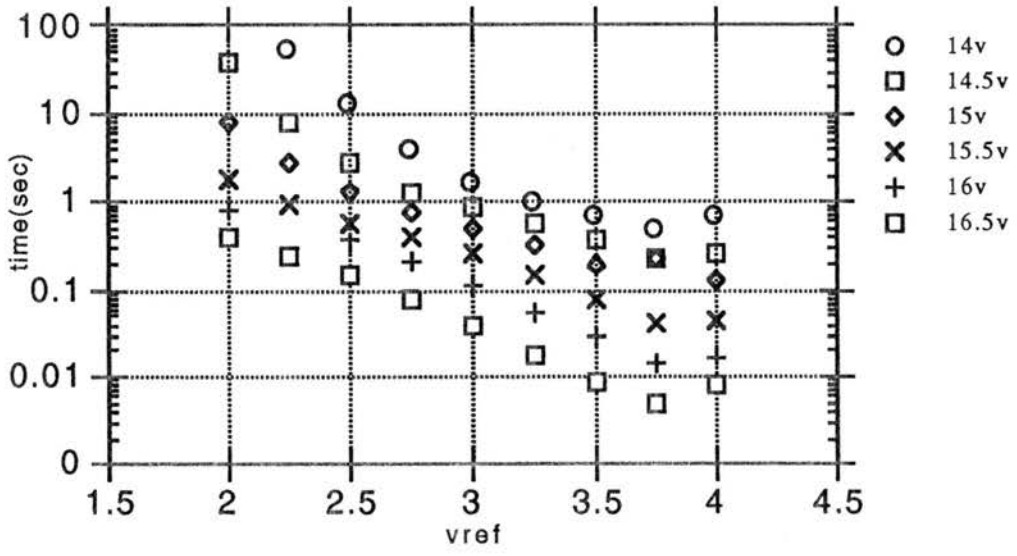


Figure 57. Elapsed Time vs. Programming Pulse Magnitude

CHAPTER V

A PROGRAMMABLE 8 BIT SERIAL A/D CONVERTER

Introduction

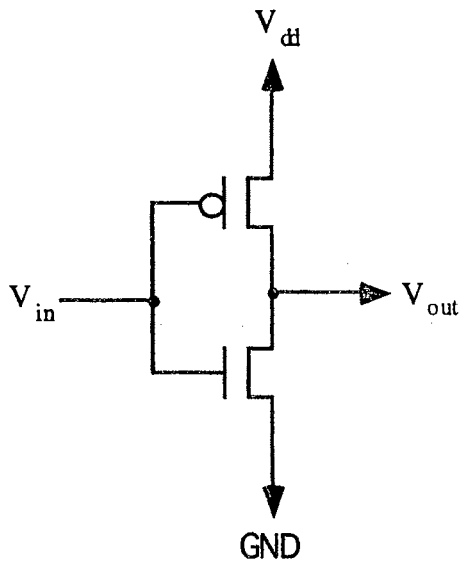
An A/D converter determining an output digital word corresponding to an analog input signal has been exploited due to its usefulness. The criteria determining the quality of the converter depends on a running speed and a resolution. It is necessary to have a smaller conversion time for a high performance converter. One method of achieving the small conversion time is to use a parallel architecture. Meanwhile, the increment of the number of bits will increase the resolution. However, there is a limitation obtaining the fine resolution in the A/D converter without settling down some factors such as an offset error, a scale factor error, and a linearity[Allen, 1987]. Especially, the offset error has primarily influenced the resolution of the converter. An auto-zero technique has been used to eliminate the offset error. But the technique is not only very space consuming technology, but also is inaccurate due to a feedthrough effect[Allen, 1987]. A different method implemented with a floating gate MOSFET has developed to eliminated the above problem. In addition to the resolution improvement with less space, the converter has been designed to consume low power by utilizing a CMOS technology.

Principle of A Programmable A/D Converter

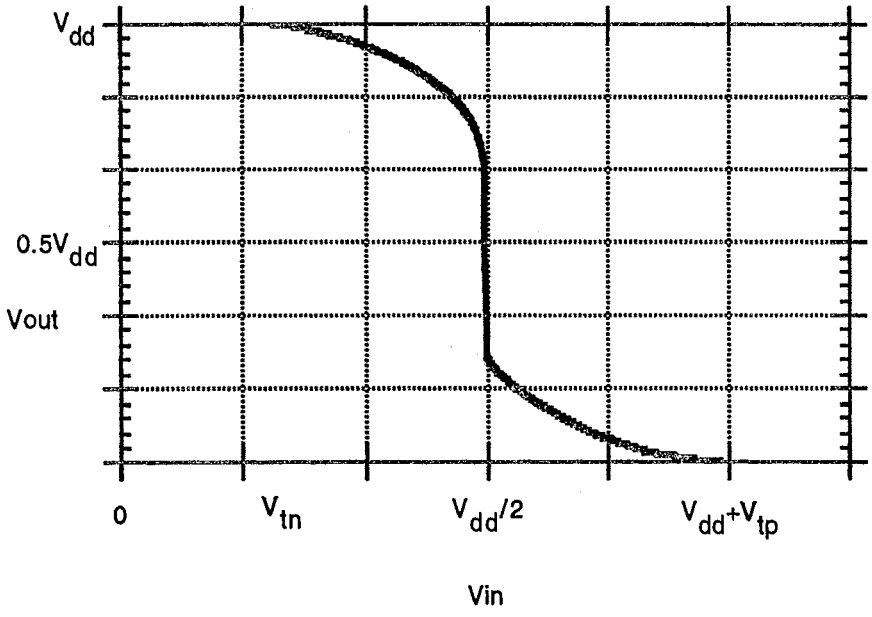
Figure 58 (a) and (b) shows a CMOS inverter and the transfer characteristic of the inverter, respectively. In Figure 58 (b), the V_{tn} and the V_{tp} indicate threshold voltages of NMOS and PMOS, respectively. The transition of the transfer function doesn't occur when the following conditions are satisfied; the n-device is cut-off, and the p-device is linear in the region defined by $0 \leq V_{in} \leq V_{tn}$ or the p-device is cut-off, and the n-device is linear in the region defined by $V_{dd} \geq V_{in} \geq V_{dd} - V_{tp}$.

Depending on the mobility of both holes and electrons, β_n/β_p ratio, and temperature, the overall transfer characteristics can be shifted to the right or left direction[Weste, 1985]. By using the bidirectional properties, the flexible characteristic function of the inverter is becoming available. However, it is not easy to control those factors in reality since they associate with a fabrication process and a circumstance.

The simplest way to acquire the flexibility is to directly control the threshold voltage of the inverter with an EERPOM device. As mentioned earlier, the threshold voltage of the device is easily manipulated by writing and erasing operation. During the writing operation, the floating gate is charged negatively with electrons. The stored negative charge on the floating gate will lead the threshold voltage of a n-channel transistor to shift positive direction. On the other hand, during the erasing operation, the electrons are removed from the floating gate transistor. Then the threshold voltage of the n-channel transistor is shifted to the negative direction. The direction of the shift on a p-channel transistor during the programming is opposite to that on a n-channel transistor. By using the properties, we have implemented a 8 bit serial A/D converter. Figure 59 shows the transfer characteristics of the converter.



(a)



(b)

Figure 58. (a) CMOS Inverter (b) CMOS Inverter DC Characteristic

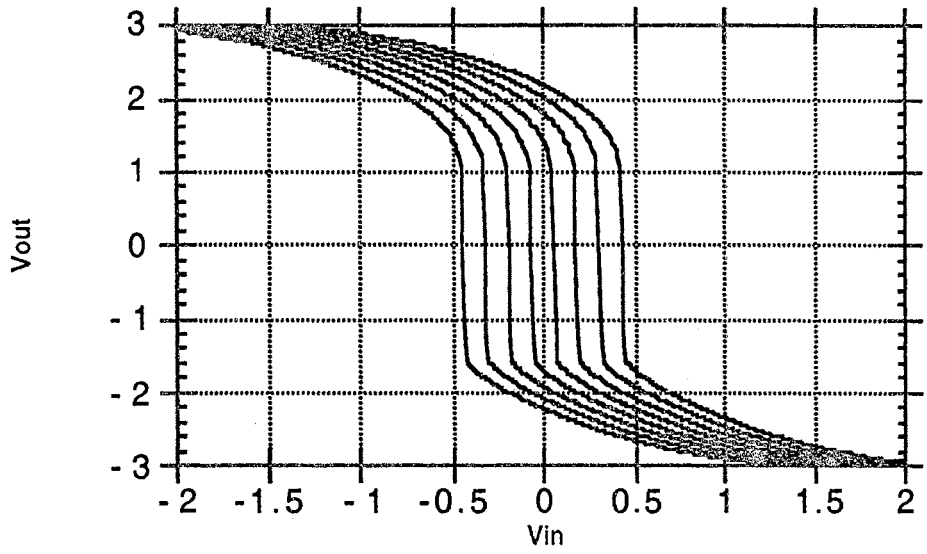


Figure 59. Characteristics of Eight Inverters with Different Transitions

Operating Scheme of Programmable A/D Converter

Figure 60 shows the schematic diagram of the programmable 8 bit serial A/D converter in which the conversion is done sequentially. The converter consists of a NMOS folded, a PMOS folded differential amplifier, a counter, a bias selection circuit, and 8 programmable floating gates MOSFET for both n and p channel transistors.

There are two modes in the converter: programming and operating mode. During the programming mode, each floating gate MOSFET of 8 bits will be charged or discharge depending on the level of the significant bit.

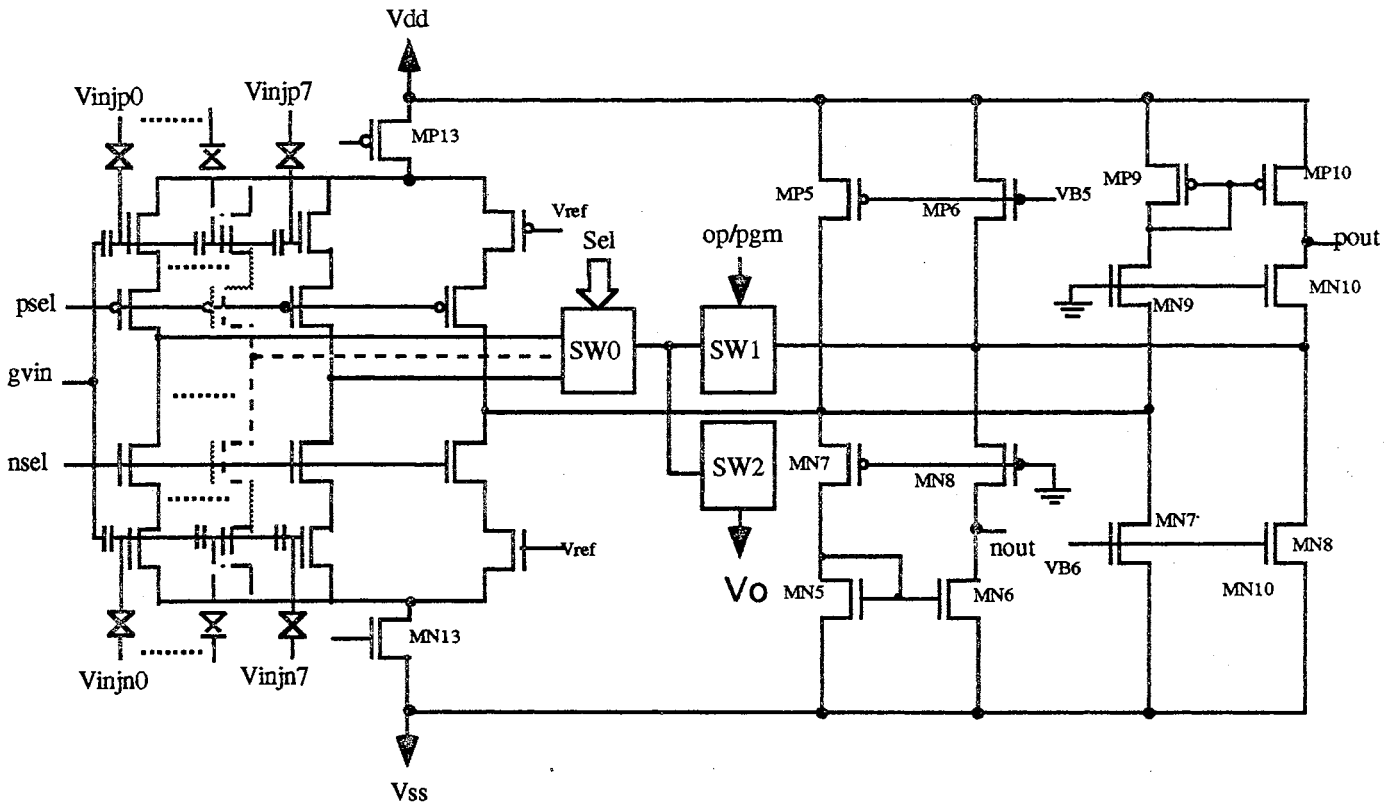


Figure 60. Circuit Diagram of the serial A/D converter

Note that the programming operation must take place separately on NMOS or PMOS, since the two folded differential amplifiers cannot be working simultaneously in this structure. Consequently, the total number of the programming will be 16.

The switches, sw1 and sw2, shown in Figure 60 represent complementary switches. With the sw1 being "on" and sw2 being "off", the converter is under the programming mode. Alternatively, the converter is under the operating mode.

After the programming mode is done, the 8 bit serial converter, which contains 8 different transition functions shown in Figure 59, is ready to receive an analog input signal and to perform the conversion. During the operating mode, the converter looks like the parallel connection of the cascode CMOS type inverters shown in Figure 61. It is well known that there is no DC current in this type of an inverter. The power consumption occurs only during the transition time of the transfer function shown in Figure 58 (b). It will greatly reduce the power consumption of the converter. In addition, the cascode type of the inverter increases the output resistance of each inverter. Consequently, it will cause the cascode configuration to have flatter characteristics in a saturation region[Allen, 1987]. It is also expected to reduce the variability of the transfer function.

Programming Mode

We now consider the programming mode in detail. The programming operation in this project is similar to the previous operation in the analog memory. The folded but not cascode amplifier has been used for a comparator during the programming. Recall that the circuitry compares a floating gate voltage with a reference voltage.

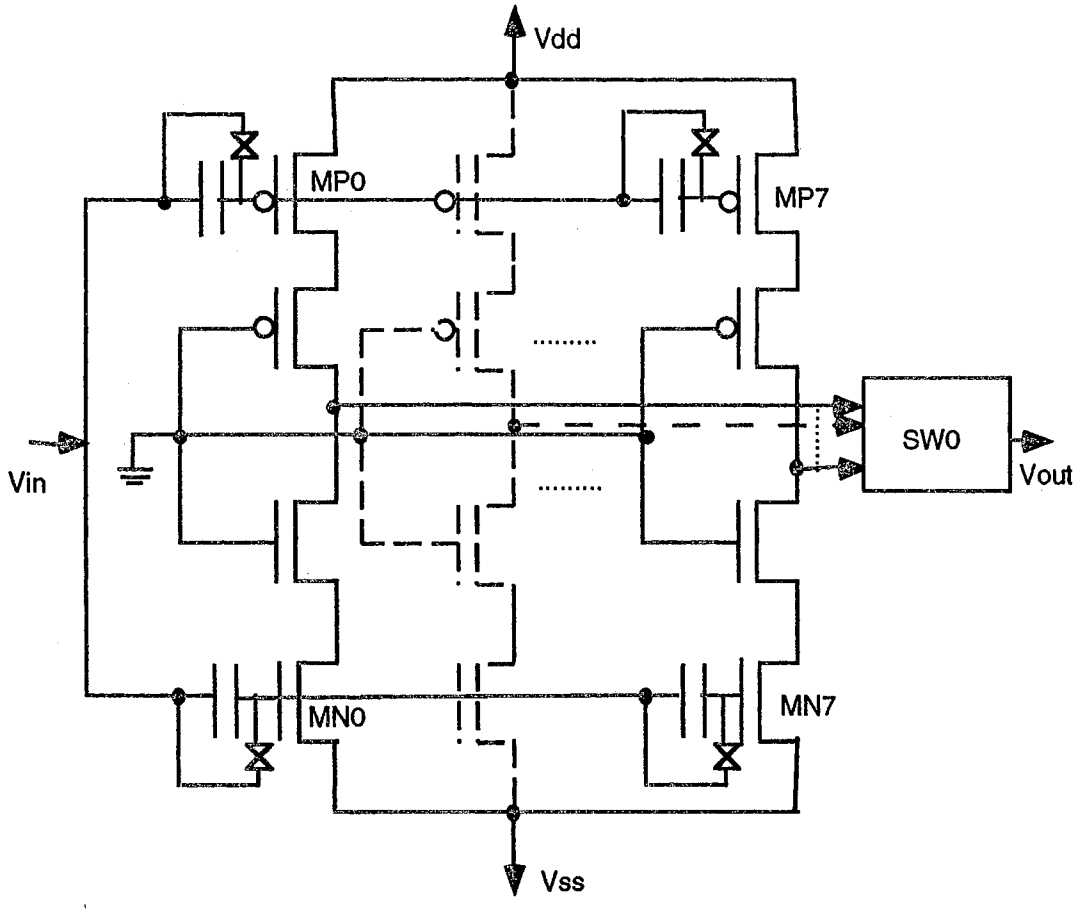


Figure 61. Schematic Diagram of a 8 Bit Serial Inverter

The difference between now and then is that the comparator contains both n and p channel folded amplifiers since the A/D converter is comprised of CMOS inverters. However, during the programming, only one type of amplifier has to be alive to avoid the disturbance of each other. Alternatively, the programming operation should be collapsed. Thus, a particular circuit separating one from the other is required. This is shown in Figure 62.

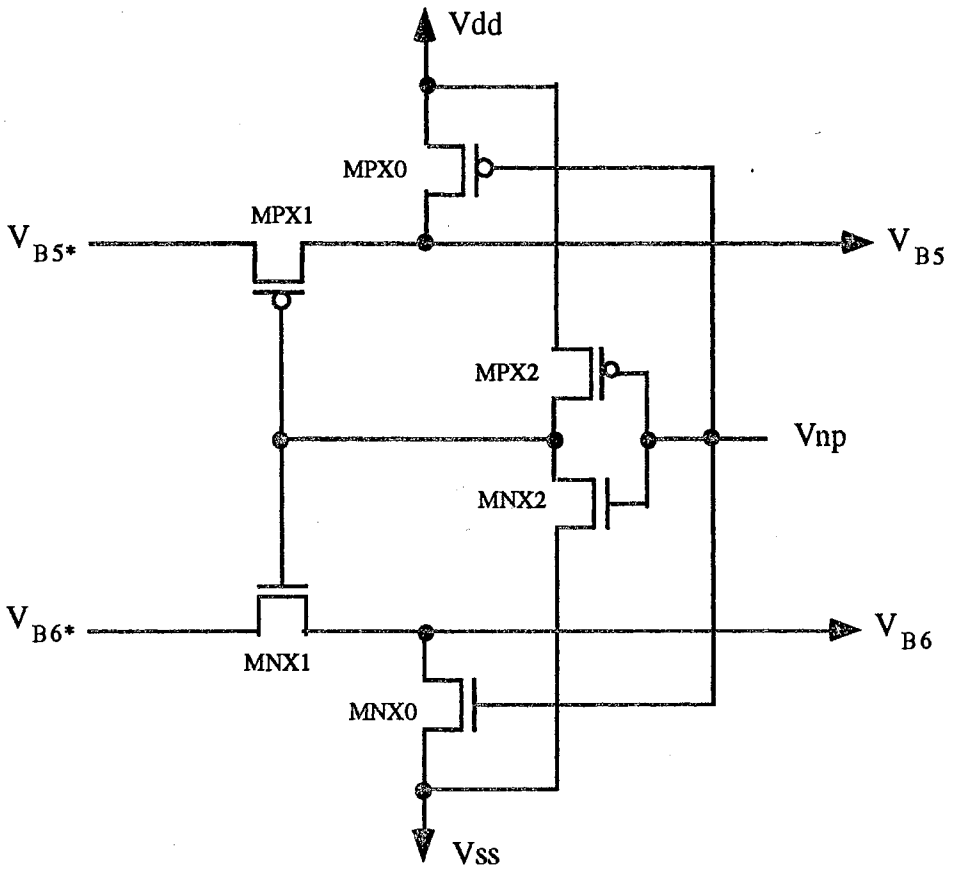


Figure 62. A Circuit Diagram Selecting a DC Bias Voltage

The operation of this circuit is not deliberate. The on of the input signal designated with Vnp which is applied externally causes the drain of MNX0 to be pulled down. Thus the DC bias voltage V_{B6} will be tied to ground no matter what the voltage state of V_{B6*} is. Meanwhile, since the MPX0 is off, the output V_{B5} should follow the state of the V_{B5*} which is generated from the bias circuit. Consequently, the p-channel folded amplifier with the Vss bias will be shut down, and the n-channel folded amplifier will be alive. Similarly, the off state of the input signal Vnp will activate the p-channel and deactivate the n-channel folded amplifier.

In addition to the circuitry determining the type of the amplifier, there exist 8 complementary switches for distinguishing one from the other seven floating gate MOSFETs. Alternately, the amplifier configuration could not be a differential amplifier so that we cannot expect proper programming operation. In order to route an appropriate floating gate MOSFET among 8 bit lines, we provide a 8 bit serial static shift register with 2 clocking scheme. Figure 63 shows just one register[Weste, 1985].

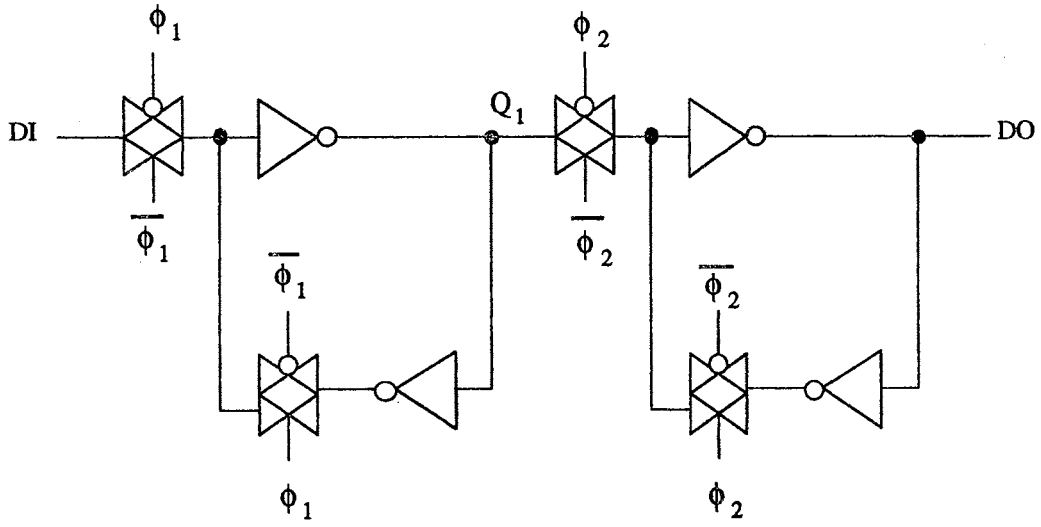


Figure 63. 2 Phase Static Flip Flops

Thus a total of 8 registers are cascaded in the counter. Here, data can be taken from the counter in serial form. In order to illustrate its operation, we consider the data word 10000000 as the input. We assume that the register is initially cleared. The shifting takes place upon the on state of the two clock signals. The reason why we uses two clock phases is to avoid a clock race condition. The first flip-flop in Figure 63 transfers data to the next flip-flop on the positive value of the clock pulse $\overline{\phi_1}$. Thus, on the positive value of the clock pulse, the first bit appears at Q_1. But the value still does not reach DO, the output of the first register. On the positive value of the clock pulse $\overline{\phi_2}$, the bit at the first flip-flop is transferred to the DO. Similarly, the bits of data are passed on so the first bit in is the first bit out.

With the same input data format as in the above example, the counter can activate just one of eight switches so that the selected floating gate will be one input node in the folded amplifier. Finally, the configuration of the n-type(p-type) amplifier with the proper switches being setting, becomes to be a standard folded amplifier shown in Figure 64.

Once the circuit is constructed, then the trimming operation should be followed to adjust each floating gate voltage. In this project, the input analog voltage range is 1V. As a result, the threshold voltage difference between the adjacent bit of the floating gate is 0.125V. The amplifier has been designed to allow the input threshold voltage range from -0.5V to 0.5V. Thus the threshold voltage increment of the first floating gate MOSFET will be made from -0.5V to -0.375V. Figure 59 shows the desired transfer function characteristics of the 8 bit serial inverter. One after the other floating gate should be charged or discharged according to the characteristics. The way of the programming is similar to that in the previous chapter.

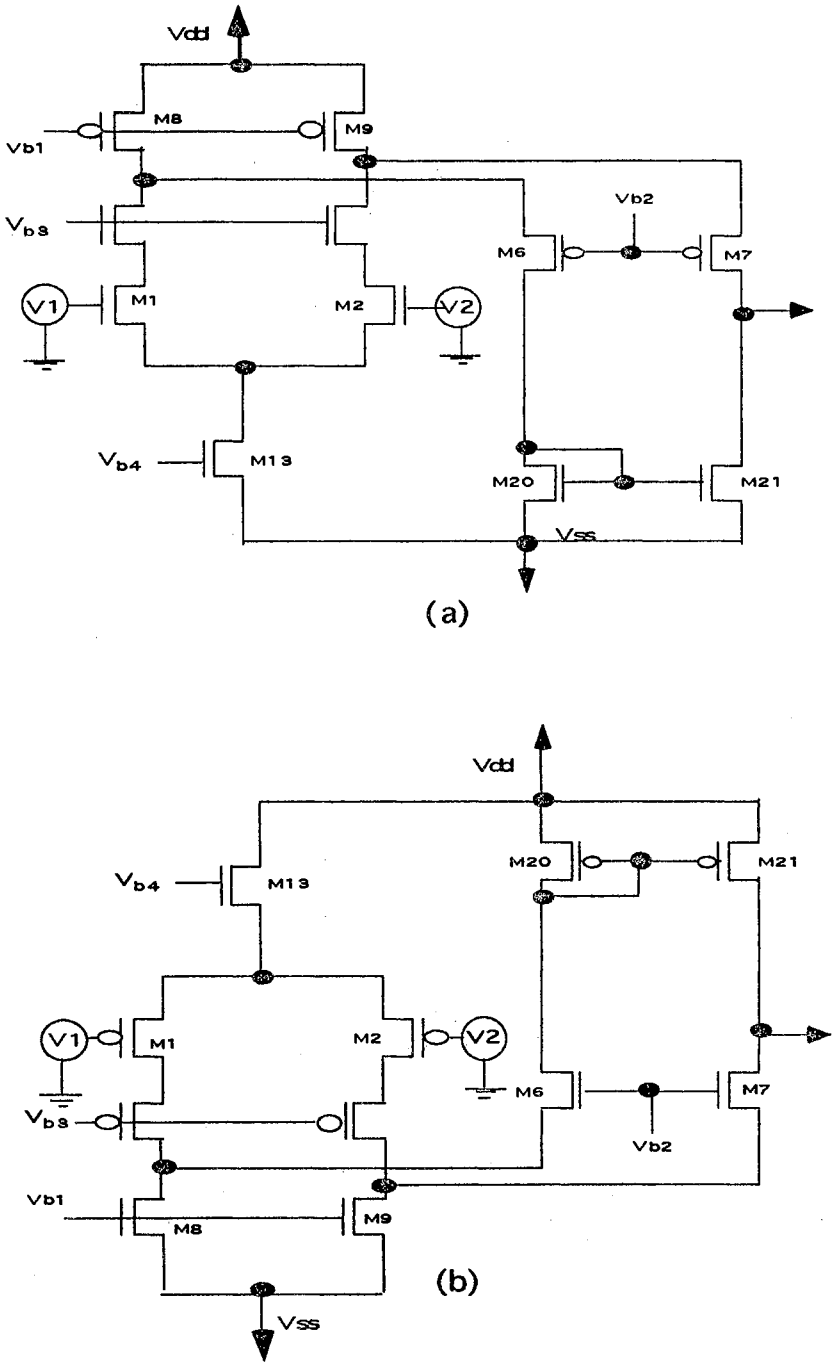


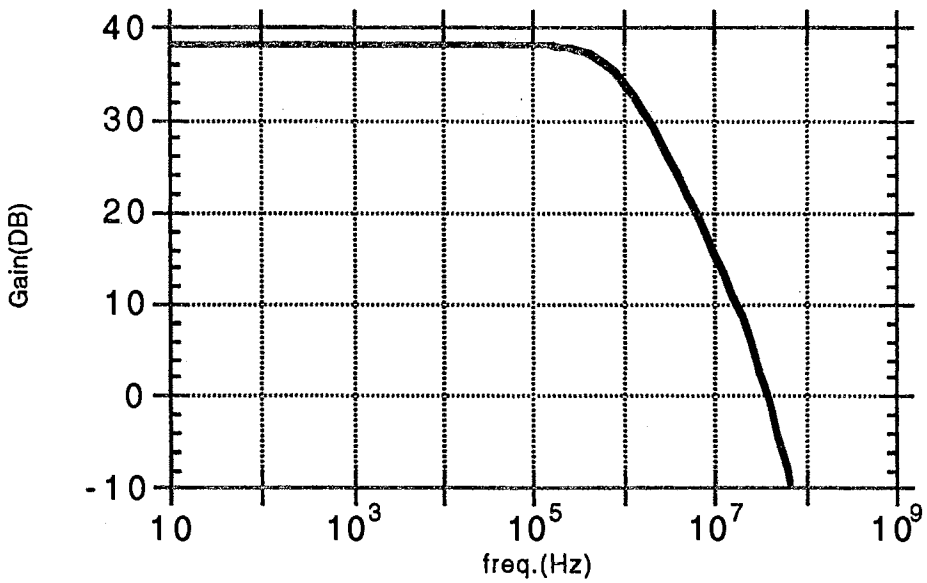
Figure 64. (a) N-Type Folded Differential Amplifier
(b) P-Type Folded Differential Amplifier

That is, the high programming pulses are applied to a control gate and ground to an injector during the writing operation. On the other hand, the programming pulse and the ground will be connected to an injector and a control gate during erasing operation. The programming operation will continue until the floating gate voltage reaches the Vref in a predetermined voltage range.

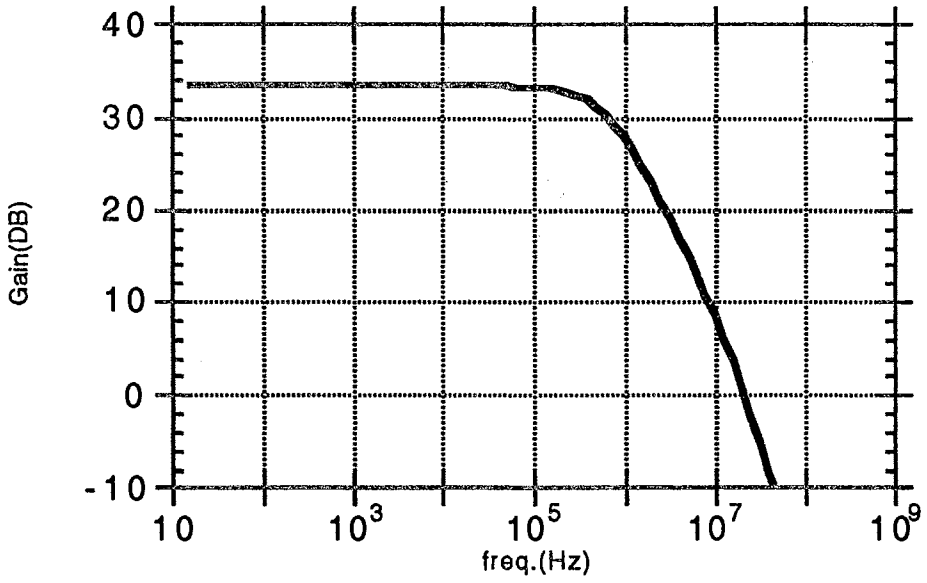
Note the gain of the amplifier has been designed not to be very large. Since the output of the comparator is directly connected to the scanner, the sudden change of the output is not helpful to the adjustment. Figure 65 shows AC characteristic functions of the N and P type comparator. It is seen that the gain of the N and the P type amplifier is around 38DB and 34DB, respectively. The offset of both amplifier are less than or equal to 10mV.

Operating Mode

Once the programming is finished, then the amplifier circuit should be removed from the overall circuitry with off and on of the sw1 and sw2 shown in Figure 60. Finally, the 8 bit programmed A/D converter illustrated in Figure 61 is accomplished. Assume that we assign various levels of threshold voltages starting with the lowest voltage and counting toward the highest to several inverters during the previous programming operation. Then each inverter has different level of transition voltages shown in Figure 59. The combining of the 8 characteristics implies a sequence of binary numbers. Toward an analog input signal, the converter provides an appropriate digital word that best represents the analog input signal.



(a)



(b)

Figure 65. (a) Open-Loop AC Transfer Characteristic of N-Type Amplifier (b) Open-Loop AC Transfer Characteristic of P-Type Amplifier

The procedure of the conversion is very simple. The analog input signal which is fed through a sample-and-hold circuit, is applied to the CMOS inverter. The input of the incoming signal is common to the inverters. The switch, sw0, shown in Figure 60 is implemented to select one appropriate output depending on the significant bit level. The counter, which has been used for selecting the particular line during the programming operation, also provide an enable signal to choose one interesting line among 8 lines. At the beginning of the operating mode, the most significant bit transition characteristic will be chosen to meet the analog input signal. As long as the value of the analog input signal is not great enough to pull down the relevant inverter, its output will be logically 1 and, allowing the clock pulse from the counter to select the next significant bit. With each count, the incoming signal will be compared with an inverter with one step lower threshold voltage. This count will continue until the value of the input voltage exceeds the middle of the transition state of the inverter. At this time, the comparator output drops to zero. It will stop the count operation. This type of counter is comparatively slow since it is a serial counter. However, we can expect an accurate converter implemented with the floating gate MOSFET.

CHAPTER VI

SUMMARY and FUTURE WORKS

The properties of the field enhancement bumps have been investigated using several different structures and electrical measurements, including the effect of the number of corners and sharpness of the electric field. The bumps, and general artificial angles are all features that contribute to the lowering of the dielectric strength of SiO2 through an increase of local electric fields at the corners.

In order to achieve the local enhancement, we designed a bump structured injector instead of the simple bar shape suggested by Thomsen[Thomsen, 1991]. It had been observed through four test chips that the injector with bumps contributes to reducing the programming voltage even though there exist some differences in magnitude. Second, we found that the edge also strengthens the local enhancement. Thus, during the erasing operation, the difference between type 1 and the others are not distinctive. Since the difference between each type during the erasing operation is negligible compared to that during the writing operation, some irregularities between each type were found. Third, we showed that the number of corners and the length of the edge do not influence the reduction of the programming voltage so much as the corner itself did. Fourth, we speculated that the size of the tip influences the electric field. The tip violating the design rule provided by

MOSIS is considered to strengthen the local electric field due to the sharp angle. Fifth, it was found that the longer the length of bump, the sharper the angle of tip. Finally, the sharp tip structure will give us an extra space when we design a cell due to its small occupied area. In spite of the fact that the insignificant number of testing and the non-flat oxide thickness sometimes provided undesirable results in our speculation, we found that there was a consistency of the trends mentioned above.

At the beginning of the experiment, we were afraid that the retention time of the EEPROM with the innovative injector structure would be deteriorated by the thinning of the oxide at each corner. But the measurement of the retention time for type 7 shows 0.1% leakage of charge for 61 years at the room temperature. Consequently, the floating gate MOSFET with the sharp bump injector reduces the programming voltage as well as the size of the memory implemented with the floating gate without any trade-off.

With the specific floating gate MOSFET, we designed a 2x2 analog array. This array is expandable to nxn. Here, n is any integer. In order to trim a desired voltage with a digital level controller, the system includes a feedback and voltage translators circuits.

The resolution of the memory has been found to be 10mvolt in the range 1.25volts to 2volts. It will substitute for 6bit digital cells. The goal of our project is to replace 1 byte digital memory with one cell of an analog memory. It was found that the high common mode gain of the comparator deteriorates the resolution. A long channel or a cascode type of transistor comprising the tail current of the comparator will reduce the problem since it increases CMRR. We have also demonstrated that the resolution is a function of duration and magnitude of programming pulse. But the effect of those to the resolution was a trivial. Instead of those factors, the resolution is strongly comparator gain

dependent. Recall that the low gain due to the common mode gain deteriorates the performance of the analog memory very much. Assuming that the CMRR of the amplifier is infinite, then the voltage range covered by the analog memory is expanded from 2volt to up to 4volt. The expansion makes it possible for the memory to substitute for 1 byte of digital memory with one cell of the analog memory. We can deduce that the comparator implemented with a boosted folded cascode amplifier will improve the performance of the memory even better.

We have also seen that the magnitude of the programming pulse affects the programming time distinctively. It means that the analog memory implemented with an appropriate magnitude of the programming pulse is applicable to a real time system without any serious errors.

It has been speculated that the amount of charge injection decreases during the writing operation as the reference voltage increases. It is due to the fact that the stored charges generate electric field, which is in the opposite direction to the applied electric field. The property will be an obstacle when the analog memory is applied to a real time system since it will delay the programming time. The analog memory with adaptive programming voltage, where the programming voltage magnitude is adjusted depending on the difference between a floating gate voltage and a reference voltage will solve the problem. It won't be difficult to construct the variable programming generator since the programming generator consumes insignificant amount of current. Thus just a simple serial diode connection can produce several different levels of programming voltage.

Finally, we proposed an A/D converter implemented with 16 floating gate MOSFETs. The ideal trimming operation will remove the offset problems which has been considered to deteriorate the performance of the converter.

Additionally, the converter implemented with CMOS inverters will remove dc current flow. As a result, it will reduce the total power of the converter. Finally, the cascode type implementation of the converter increases the output resistance of each inverter. Consequently, it will cause the cascode configuration to have flatter characteristics in a saturation region.

REFERENCES

Allen, P. E. and D. R. Holberg. (1987). "CMOS analog circuit design," New York, HRW.

Anderson, R. M. and Kerr, D. R. "Evidence for surface asperity mechanism of conductivity in oxide grown on polycrystalline silicon," J. Appl. Phys. vol.48(no.11): pp.4834-4836.

Bhattacharyya, A. (1984). "Modeling of write/erase and charge retention characteristics of floating gate EEPROM devices," Solid-State Electronics, vol.27(no.10): pp.899-906.

Blyth, T. e. a. (1991). "A non-volatile analog storage device using EEPROM technology," J. ISSCC Digest of Technical Papers : pp.192-193.

Chai, Y. Y, and L. G. Johnson(1993). "Floating gate MOSFET with reduced programming voltage," Electronics Letters, pp.1536-1537

Carley, L. R. (Dec. 1989). "Trimming Analog Circuits Using Floating-Gate Analog MOS Memory," IEEE J. Solid-State Circuits, vol.24: pp.1569-1575.

Durfee, D. A. and F. S. Shoucair. (May 1992). "Comparison of Floating Gate Neural Network Memory Cells in Standard VLSI CMOS Technology," IEEE Trans. in neural networks Vol.3: pp.347-353.

Ellis, R. K., H. A. R. Wegener and J. M. Caywood. (1982). "Electron tunneling in nonplanar floating gate memory structure," IEDM Tech. Dig. : pp.749-752.

Ellis, R. K. (Nov. 1982). "Fowler-Nordheim emission from non-planar surfaces," IEEE Device Lett. EDL-3: pp.330-332.

Frohman-Bentchkowsky, D. (Apr. 1971). "Memory behavior in a floating-gate avalanche-injection MOS(FAMOS) structure," Appl. Phys. Lett. vol.18: pp.332-334.

Gray, P. R. and R. G. Meyer. (1988). Analysis and Design of Analog Integrated

Circuits. John Wiley & Sons, Inc.

Haznedar, H. (1991). Digital microelectronics. Redwood city, The Benjamin/Cummins publishing company.

Johnson, W. S., G. Perlegos, A. Renninger, G. Kuhn and T. Ranganath. (Feb. 1980). "A 16K electrically erasable non-volatile memory," ISSCC Digest of Technical Papers : p.152.

Khang, D. and S. M. Sze. (July-Aug 1967). "A floating gate and its application to memory devices," Bell Syst. Tech. : pp.1288-1295.

Kolodny, A. e. a. (1983). "Analysis and modeling of floating-gate E2PROM cells," IEEE Trans. Electron Devices, vol.ED-30(n0.11): p.1572.

Kolodny, A., S. T. K. Nieh, B. Eitan and J. Shappir. (June 1986). "Analysis and Modeling of Floating-Gate EEPROM Cells," IEEE Trans. Electron Devices, vol. ED-33: pp.835-844.

Kupec, J., W. Gosney, V. McKenny and V. Kowshik. (1980). "Triple level poly silicon E2PROM with single transisotr per bit," IEDM Technical Digest : p.148.

Lee, B. W., B. J. Sheu and H. Tang. (June 1991). "Analog Floating-Gate Synapses for General-Purpose VLSI Neural Computation," IEEE Trans. Circuits and Systems, vol. 38: pp.654-658.

Lenzlinger, M. and E. H. Jullien. (Jan. 1969). "Fowler-Nordheim tunneling into thermally grown SiO2," J. Appl. Phys., vol.40: p.278.

Mann, J. R., "Floating gate circuits in MOSIS," MIT Lincoln Labs., Lexington, MA, TR-824, Nov. 1990

McConnell, M., B. A. R. Bussey, M. Gill, S.-W. Lin, D. McElroy, J. F. Schreck, P. Shah, H. Stiegler, P. Truong, A. L. Esquivel, J. Paterson and B. Riemenschneider. (April 1991). "An experimental 4-Mb Flash EEPROM with sector erase," IEEE J. Solid-State Circuits, vol.26(no.4): pp.484-489.

Mielke, N., A. Fazio and H.-C. Liou. (1987). "Reliability comparison of FLOTOX and textured-polysilicon E2PROMs," IRPS : pp.85-92.

Nozawa, H. e. a. (1982). "A thermionic electron emission model for charge retention in SAMOS structures, Japaness Journal of Applied Physics, vol.21(no.2): L111-L112.

Nozawa, H., N. Matsukawa and S. Morita. (1986). "An EEPROM cell using low barrier height tunnel oxide," IEEE Trans. Electron Devices, vol.ED-33: p.275.

Ong, T.-C., P. K. Ko and C. Hu. (Sept. 1989). "The EEPROM as an Analog Memory Device," IEEE Trans. on Electron Device, Vol 36: pp.1840-1841.

Ramacher, U. (1993). VLSI design of neural networks. Kluwer academic publishers.

Sackinger, E. and W. Guggenbuhl. (Dec. 1988). "An analog trimming circuit based on a floating-gate device," IEEE Electron Device Lett., vol.23(no.6): pp.1437-1440.

Sethi, R. B., U. S. Kim, I. Johnson, P. Cacharelis and M. Manley. (May 1992). "Electron barrier height change and its influence on EEPROM cells," IEEE Electron Device Lett., vol.13(no.5): p.244-246.

Shimabukuro, R. L., R. E. Reedy and G. A. Garcia. (August 1988). "Dual Polarity Non-volatile MOS Analogue cell for Neural-Type Circuitry." Electronics Lett. . Vol. 24: pp. 1231-1232.

Solomon, P. (1977). "High-field electron trapping in SiO2," J. Appl. Phys., vol. 48(No.9): pp.3843-3849.

Spindt, C. A. e. a. (Dec. 1976). "Physical properties of thin-film field emission cathodes with molybdenum cones," J. Appl. Phys., vol.47(no.12): pp.5248-5263.

Sweeney, J. and R. Geiger. (Sept. 1989). "Very high precision analog trimming

using floating gate MOSFETs," in European conf. Circuit Theory and Design : pp.652-655.

Tarui, Y., Y. Hayashi and K. Nagai. (1974). "Invited: electrically reprogrammable nonvolatile semiconductor memory," Proceedings of the 5th conf. on Solid-State Devices : pp.348-355.

Terada, Y. e. a. (1990). "High speed page mode sensing scheme for EPROM's and flash EEPROM's using divided bit line architecture," Symposium on VLSI circuits : pp.97-98.

Thomsen, A. and M. A. Brooke. (March 1991). "A Floating-Gate MOSFET with Tunneling Injector Fabricated Using a Standard Double-Polysilicon CMOS Process," IEEE Electron Device Lett., vol.12(no.3): pp.111-113.

Tsividis, Y. P. (1987). Operation and modeling of the MOS transistor. New York, McGraw-Hill Int.

Wallmark, J. T. e. a. (1969). "Switching and storage characteristics of MIS memory transistors," RCA Rev.: pp. 335-365.

Wang, S. T. (1980). "Charge retention of floating-gate transistors under applied bias conditions," IEEE Trans. on Electron Devices: pp.297-299.

Weste, N. and K. Eshrahgian. (1985). Principle of CMOS VLSI Design. Menlo Park, Addison-Wesley.

Yaron, G., S. J. Prasad, M. S. Ebel and B. M. K. Leong. (Oct. 1982). "A 16K E2PROM Employing New Array Architecture and Designed-In Reliability Features," IEEE J. Solid-State Circuits, vol.SC-17(no.5): pp.833-840.

VITA

Yong Yoong Chai

Candidate for Degree of

Doctor of Philosophy

Thesis: A 2x2 ARRAY ANALOG MEMORY IMPLEMENTED WITH A SPECIAL LAYOUT INJECTOR

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Seoul, Korea, August 16, 1958, son of Mr. Nam Young Chai and Mrs. Chun Ja Lee

Educational: Graduated from Kyung Ki High School, Seoul, Korea 1981; received Bachelor of Electronic Engineering from SoGang University in July 1985; received Master of Science degree from Oklahoma State University in May, 1990; completed requirements for Doctor of Philosophy at Oklahoma State University in December, 1994.

Professional Experience: Research Assistance(1994-Present), Dept, of Electrical Engr., OSU; Teaching Assistant (1990-Present), Dept, of Electrical Engr., OSU; Research Assistance(1989-Spring), Dept, of Electrical Engr., OSU.
Researcher, (1985-1987) GoldStar Telecommunication Ltd., Anyang, Korea
Researcher, (1987-1988) GoldStar Semiconductor Ltd., Anyang, Korea