NOVEL CAPACITORLESS SINGLE-TRANSISTOR DRAM TECHNOLOGIES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Mehmet Günhan Ertosun

May 2010

This dissertation is online at: http://purl.stanford.edu/wj698dt9238

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Krishna Saraswat, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Yoshio Nishi

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Philip Wong

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The dynamic random access memory (DRAM) industry has achieved miracles packing more and more memory bits onto ever smaller silicon die. But, the scaling of the conventional 1Transistor/1Capacitor (1T/1C) DRAM is becoming increasingly difficult, in particular due to the capacitor which has become harder to scale, as device geometries shrink.

Recently the capacitorless single-transistor (1T) DRAMs have attracted attention, due to its ability to achieve higher memory cell density and to solve the problems associated with the scaling of the capacitor. The information is stored as different charge levels at a capacitor in conventional 1T/1C DRAM, whereas the 1T DRAM employs floating body effects within the transistor to store the information without the need of the capacitor.

The absence of the capacitor is advantageous in terms of scalability, process and fabrication complexity, and compatibility with the logic processing steps, device density, yield and cost. Due to all these advantages of the capacitorless DRAM, and to solve the scaling problem of conventional 1T/1C DRAM, this work is focused on creating novel single transistor DRAM technologies. .

The first device that is studied is vertical double-gate (DG) capacitorless single-transistor DRAM. This device has advantages of being vertical with a small footprint, it

can be integrated on bulk Si and being a double-gate device allows better electrostatic control of the channel and higher intrinsic device retention.

The second device that is investigated is a novel DRAM device: capacitorless single-transistor quantum well DRAM, which employs energy-band engineering approach within the body of the transistor in order to create a “hole storage pocket” and “carrier distribution control layer” which results in superior device characteristics in terms of scalability and memory sensing window.

The third and the final device that is studied is also a novel DRAM device: capacitorless single-transistor charge trap DRAM. The body of this device is engineered with the intentional charge traps so as to obtain a memory effect. This novel DRAM relies on the existence and absence of electrons within its body and uses a charge-trapping mechanism in its memory operation, unlike conventional 1T DRAMs that rely on holes and employ floating-body effects in their memory operation.

This page is intentionally left blank

Yeniliğe Doğru

Her gün bir yerden göçmek ne iyi
Her gün bir yere konmak ne güzel
Bulanmadan, donmadan akmak ne hoş

Dünle beraber gitti cancağızım
Ne kadar söz varsa düne ait
Şimdi yeni şeyler söylemek lazım

-Mevlana Celaleddin Rumi

Something New

Every day
How wonderful it is to move from one place,
And alight upon another.
How wondrous is this flowing,
Neither blurry, nor frozen
Just flowing…
How pleasing!

All gone
My dear,
All the words of yesterday
Are gone.
Now is the time to say something new.

-Rumi
(Translation: Mehmet Günhan Ertosun)

This page is intentionally left blank

Acknowledgements

This thesis would not have been possible without the help, contribution and support of several people to whom I am greatly indebted.

First of all, my most sincere and deepest gratitude goes to my advisor, Prof. Krishna C. Saraswat. He has been a great mentor, a great advisor and source of inspiration. I have been extremely fortunate to work with the finest advisor that one could possibly hope for. Without his guidance and support, this work which you are about to read would not have existed.

I am also deeply indebted to my co-advisor, Prof. H.-S. Philip Wong, who has been an excellent role model and mentor. I have learned a lot from his vast knowledge and wide industry experience. I am also thankful to him for contributing his time to serve on my thesis committee.

I am also very thankful to Prof. Yoshio Nishi for his generous support and invaluable guidance. He has been very kind and generous to share his years of success and experience in the field. I am grateful to him for kindly accepting to be on my thesis committee.

I would like to thank Prof. David A. B. Miller for serving as the chair of my PhD oral defense.

I am also grateful to Paul Kirsch, and Kwan-Yong Lim of SEMATECH; and Hoon Cho and Pawan Kapur of Stanford for their invaluable collaboration.

I was extremely fortunate to get to know Prof. Haldun Ozaktas during my undergraduate studies at Bilkent University, and I also had a chance to be able to work with him. He was a great mentor from whom I have learned a lot, and it has been extremely rewarding experience to work with him.

Special thanks to the administrative staff - Irene, Gail, Gabrielle, Fely, Marjorie, Maureen, Miho, Natasha, Debby, Diane - for being very helpful and supportive throughout my stay at the Center for Integrated Systems and Electrical Engineering Department of Stanford.

I was also very lucky to have a great group of friends who have enriched my life and made my Stanford experience a very enjoyable one. I would like to thank all of them for their friendship and support.

I also would like to thank to Stanford University for awarding me with the very generous Stanford Graduate Fellowship. I also acknowledge the support of Non-Volatile Memory Technology Research Initiative (NMTRI) of Stanford and its member companies, and SEMATECH.

When I first arrived in the United States, I was very lucky to get to know Barbara & Bob Simpson through the Stanford's Homestay program offered to international students in which local families welcome international students to their home for a five

day stay prior to moving into their university residences. It was a pleasure for me to get to know them and their family.

No words would ever do justice to express my deepest thanks and gratitude to my family – my parents and my brother. I will be eternally grateful to them for being there for me at all times. Their continuous love, sacrifice, support and encouragement have allowed me to pursue my ambitions.

Lastly, but most importantly, I am everlastingly grateful and thankful to the ONE…

"I can no other answer make, but, thanks, and thanks."
William Shakespeare

This page is intentionally left blank

Dedication

I would like to dedicate this dissertation

To the memories of my grandfathers Ahmet, Hasan Cahit,

&

To my mother Duyçen.

This page is intentionally left blank

Contents

This page is intentionally left blank

List of Tables

This page is intentionally left blank

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

This page is intentionally left blank

Chapter 1

Introduction

1.1 Motivation

Dynamic random access memory (DRAM) is a type of random access memory. DRAM cell consists of one transistor and one capacitor. Capacitor stores the information in terms of the charge and the transistor is used to write and read the stored information. Each bit of data is stored in a separate capacitor within an integrated circuit. Due to the leakage of the capacitors, the information eventually fades unless the capacitor charge is refreshed periodically. This refresh requirement makes DRAM a dynamic memory as opposed to a static memory, such as, SRAM (static random access memory).

The advantage of DRAM is its structural simplicity: only one transistor and a capacitor are required per bit, compared to six transistors in SRAM. This allows DRAM to reach very high density.

The DRAM industry has achieved miracles packing more and more memory bits per unit area in a silicon die. But, the scaling of the conventional 1Transistor/1Capacitor (1T/1C) DRAM is becoming increasingly difficult, in particular due to the capacitor which has become harder to scale, as device geometries shrink.

The problem of scaling and leakage – as well as device size – rests fundamentally with the basic transistor/capacitor building block. While the transistor element is theoretically scalable – at least for the foreseeable future – the capacitor is not. Capacitors can be fabricated as high stacks above the wafer surface or deep trenches inside the wafer to maximize the surface area and thus the capacitance per unit footprint area. However, if the overall bitcell size shrinks due to increased density or a smaller process node, then the capacitor will have to be made higher or deeper in order to maintain the minimum charge required for reliable operation. We are fast-approaching the scaling limits for the capacitor element, and a new approach or a DRAM replacement will be needed [1].

In recent years, DRAM cell structure has been migrating from trench to a stack capacitor cell. Trench DRAM cell could not survive future scaling due to its

manufacturing difficulties and getting performance of the memory cell. But, even the stack capacitor cell also has many technology challenges for 40 nm or smaller size DRAM [2].

Recently the capacitorless single-transistor (1T) DRAMs have attracted attention, due to the lack of the capacitor and the problems associated with the scaling of the capacitor, and due to its ability to achieve higher device density. The information is stored as different charge levels at a capacitor in conventional 1T/1C DRAM, whereas the 1T DRAM emlpoys floating body effects within the transistor to store the information without the need of an extra extrernal capacitor.

The absence of the capacitor is advantageous in terms of scalability, process and fabrication complexity, compatibility with the logic processing steps, device density, yield and cost.

Due to all these aforementioned advantages of being capacitorless, and to solve the scaling problem of conventional DRAM, and so as to come up with the –soon to be needed- DRAM replacements, this work is focused on creating novel single transistor DRAM technologies.

1.2 Thesis Organization

The thesis is composed of six chapters including this introduction chapter. Chapter 2 gives a background of dynamic random access memory (DRAM). In chapter 2, at first, basics of DRAM are introduced. Subsequently, future challenges and scaling problems of DRAM is discussed.

In chapter 3, vertical double-gate (DG) capacitorless single-transistor DRAM is presented. This device has advantages of being vertical and hence can be integrated on bulk Si, is a double-gate device, which allows better electrostatic control, higher device density and higher intrinsic device retention.

In chapter 4, a novel DRAM device is introduced and investigated: capacitorless single-transistor quantum well DRAM, which employs energy-band engineering approach within the body of the transistor in order to create a “storage pocket” and a carrier distribution control layer which results in superior device characteristics in terms of scalability and memory sensing window.

In chapter 5, another novel DRAM device is presented and discussed: capacitorless single-transistor charge trap DRAM. The body of this device is engineered with intentional charge traps so as to obtain memory effect. This novel DRAM relies on the existence and absence of electrons within its body and uses a charge-trapping

mechanism in its memory operation, unlike conventional 1T DRAMs that rely on holes and employ floating-body effects in their memory operation.

Finally, chapter 7 presents a summary and suggests possible future work.

1.3 References

[1] Serguei Okhonin**,** "In Search of a Better DRAM", Innovative Silicon, Inc. [http://www.innovativesilicon.com/en/pdf/In_Search_of_a_Better_DRAM_white_paper.pdf]

[2] International Technology Roadmap for Semiconductors (ITRS) 2009 Edition

Chapter 2

Overview

2.1 MOS Memories

An overview of existing memory devices is shown in **Fig. 2.1**, classified in the way the information is stored. If the information is stored nonvolatile, then the power-supply can be turned off and the information remains stored. But, this is not the case for two remaining memory groups. These can be distinguished by their clock frequency. The static memories have no requirement on the slowest clock rate, whereas the dynamic ones require a periodic clock to refresh the stored information. The memories, shown in **Fig. 2.1**, may be used as standalone or in embedded form in all kinds of controllers and microprocessors. The meaning of the individual memory names is summarized in **Table 2.1**.

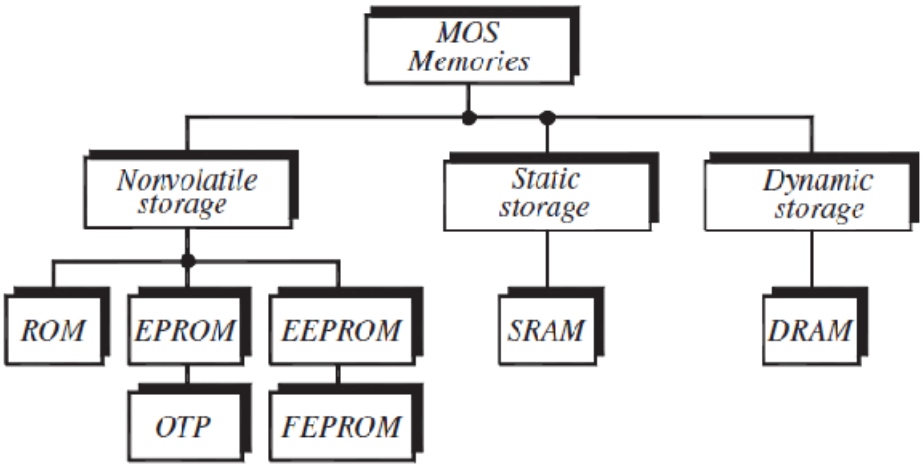


Fig 2.1 MOS memory classification [2.1]

Table 2.1 Commonly used memory names

Name	Memory
ROM	Read only memory
EPROM	Electrically programmable ROM
OTP	One time programmable EPROM
EEPROM	Electrically erasable programmable ROM
FEPROM	Flash erasable PROM
SRAM	Static random access memory
DRAM	Dynamic random access memory

2.2 Dynamic Memories

These are memories where the information is stored as different charge levels at a capacitor. Due to diverse leakage currents the storage of charge can only be guaranteed for a particular time at the capacitor. This requires a periodic refresh to maintain the cell's charge integrity.

2.2.1 Dynamic random access memory (DRAM)

Dynamic random access memory (DRAM) is a type of memory that stores each bit of data in a separate capacitor within an integrated circuit. The basic DRAM cell consists of one transistor and one capacitor as shown in **Fig. 2.2**. Due to the leakage of the capacitors, the information eventually fades unless the capacitor charge is refreshed periodically. This refresh requirement makes DRAM a dynamic memory as opposed to SRAM (Static random access memory).

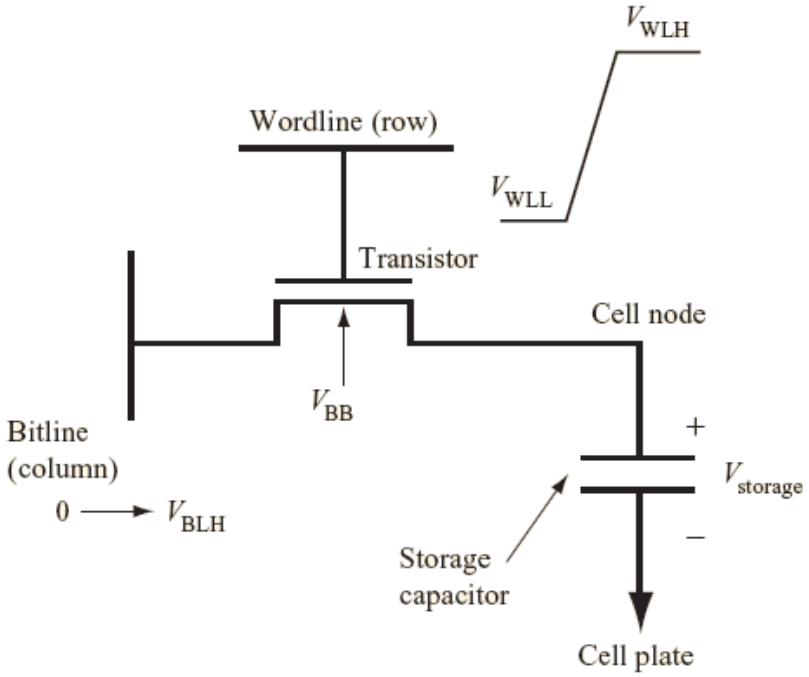


Fig2.2 Schematic of a DRAM cell. The array device (transistor) is addressed by switching the wordline voltage from V_{WLL} (wordline-low) to V_{WLH} (wordline-high), enabling the bitline and the capacitor to exchange charge. In this example, a data state of either a "0" (0 V) or a "1" (V_{BLH}) is written from the bitline to the storage capacitor. V_{BB} is the electrical bias applied to the p-well. [2.2]

The advantage of DRAM is its structural simplicity: only one transistor and a capacitor are required per bit, compared to six transistors in SRAM (**Fig. 2.3**). This allows DRAM to reach very high density.

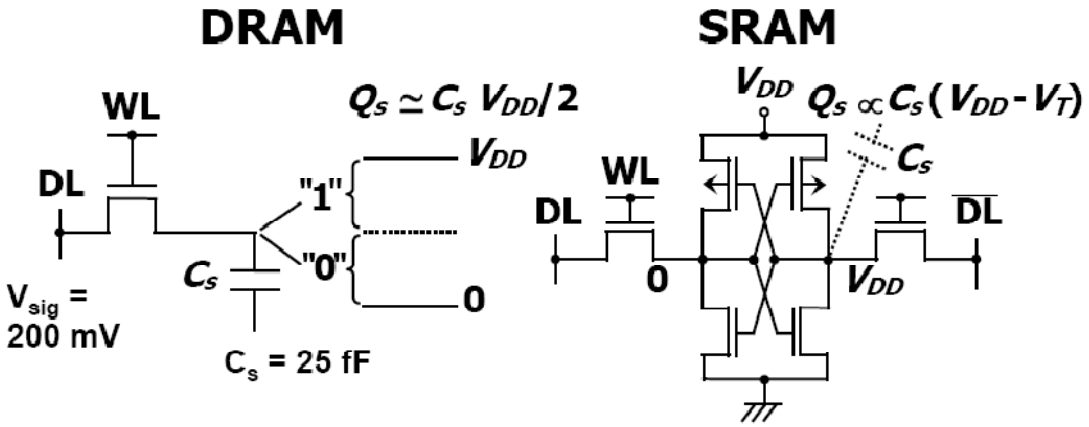


Fig 2.3 DRAM and SRAM cell comparison [2.3]

DRAM was invented in 1966 by Dr. Robert Dennard at the IBM Thomas J. Watson Research Center and he was awarded U.S. patent number 3,387,286 in 1968 (**Fig 2.4**).

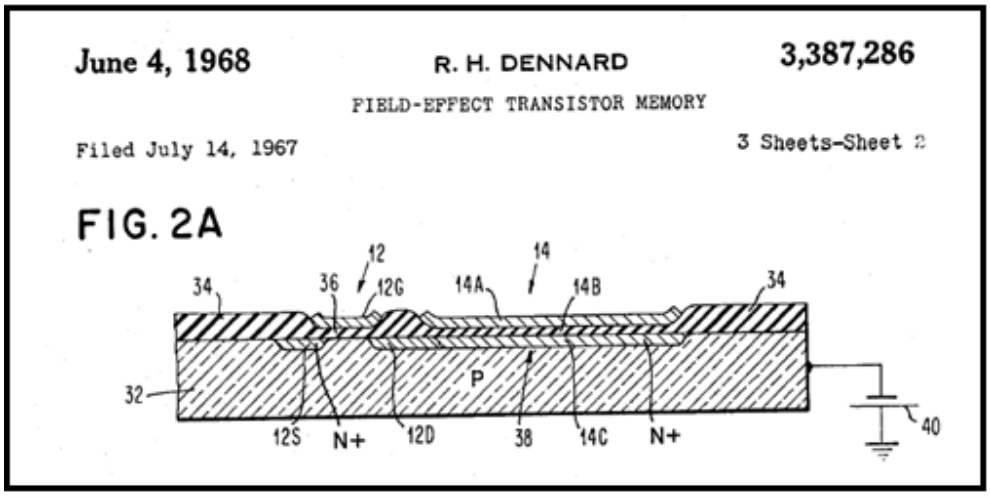


Fig 2.4 Drawing of original designs of DRAM patented in 1968. [2.4]

In 1970 Intel released the first DRAM chip, a 1K PMOS device. Since that time, the basic DRAM building block has consisted of a single transistor and an increasingly complex capacitor.

The DRAM industry has achieved miracles packing more and more memory bits per unit area in a silicon die. But, the scaling of the conventional 1Transistor/1Capacitor (1T/1C) DRAM is becoming increasingly difficult, in particular due to the capacitor which has become harder to scale, as device geometries shrink.

Apart from the problems associated with the scaling of the capacitor, scaling introduces yet another major problem for the DRAM manufacturers which is the leakage current. In both the memory cell as well as the supporting circuitry, leakage becomes more significant as complementary metal-oxide-semiconductor (CMOS) processing nodes progress from 90nm, through 78nm, 50nm and 45nm. Memory chips are already being discussed at 32nm, at which point leakage in traditional designs will grow to become very difficult and prohibitively expensive to counter, and will require new architectures, changes to standard operating specifications and significant process changes [2.8].

The problem of scaling and leakage – as well as device size – rests fundamentally with the basic transistor/capacitor building block. While the transistor

element is theoretically scalable – at least for the foreseeable future – the capacitor is not. The cell capacitance is the key parameter that determines the sensing signal margin, sensing speed, data retention time and endurance against the soft error. It is generally accepted that the minimum cell capacitance should be 25fF/cell regardless of minimum feature size, density, and chip size. The requirement of memory cell capacitance over 25fF/cell is a practical design guideline rather than a theoretical limit. This requirement imposes a great challenge on giga-bit scaled DRAMs because capacitor area is scaled down with the square of the minimum feature size [2.8]. Capacitors can be fabricated as high stacks (**Fig. 2.5**) or deep trenches (**Fig. 2.6**). However, if the overall memory cell size shrinks due to increased density or a smaller process node, then the capacitor will have to be made higher or deeper in order to maintain the minimum charge required for reliable operation. We are fast-approaching the scaling limits for the capacitor element, and a new approach or a DRAM replacement will be needed [2.9].

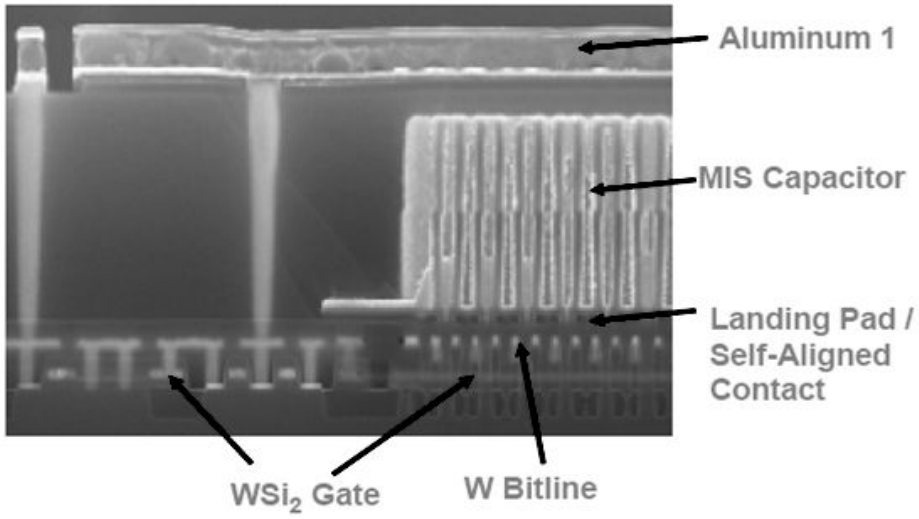


Fig 2.5 DRAM with stack capacitor [2.5] [Courtesy: H.-S. P. Wong]

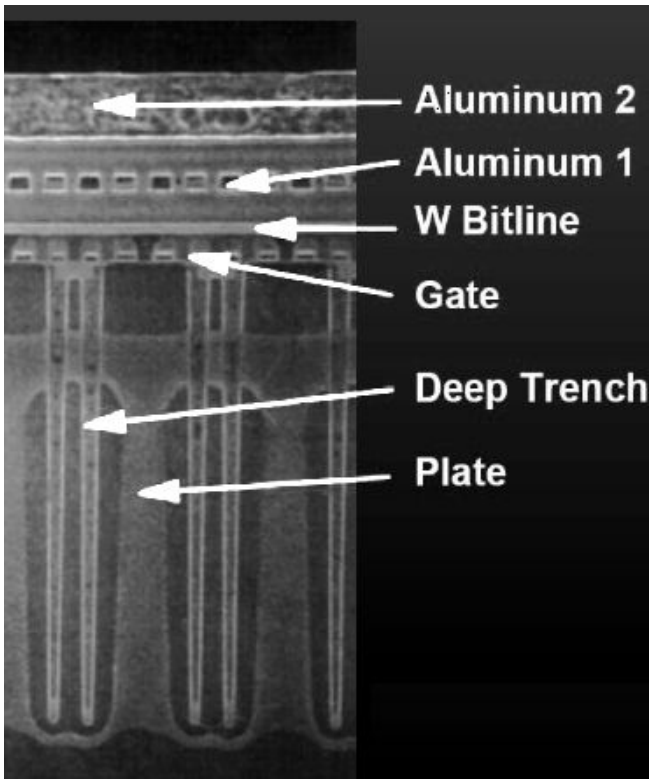


Fig 2.6 DRAM with trench capacitor **[2.6]** [Courtesy: H.-S. P. Wong]

2.3 Description of Challenges

There are several challenges of DRAM and SRAM scaling [2.7]:

2.3.1 Near-Term ($L_g \geq 16$ nm)

For DRAM, a key issue is implementation of high-κ dielectric materials in order to get adequate storage capacitance per cell even as the cell size is shrinking. Also important is controlling the total leakage current, including the dielectric leakage, the storage junction leakage, and the access transistor source/drain subthreshold leakage, in order to preserve adequate retention time. The requirement of low leakage currents causes problems in obtaining the desired access transistor performance. Deploying low sheet resistance materials for word and bit lines to ensure acceptable speed for scaled DRAMs and to ensure adequate voltage swing on word line to maintain margin is critically important**.** The need to increase bit density and to lower production cost is driving toward $4F^2$ cell size, which will require high aspect ratio and non-planar structures. Novel solution to have a capacitorless cell would be highly beneficial.

For SRAM scaling, difficulties include maintaining both acceptable noise margins in the presence of increasing random threshold voltage (V_t) fluctuations and random telegraph noise, and controlling instability, especially hot-electron instability and negative bias temperature instability (NBTI). There are difficult issues with keeping the leakage current within tolerable targets, as well as difficult lithography and etch process issues with scaling. Solving these SRAM challenges is critical to system performance, since SRAM is typically used for fast, on-chip memory.

2.3.2 Long-Term (L_g < 16 nm)

Increasing difficulty is expected in scaling DRAMs, especially in continued demand of scaling down the footprint of the storage capacitor. Thinner dielectric equivalent oxide thickness (EOT) utilizing ultra-high-κ materials and attaining the very low leakage currents and power dissipation will be required. A DRAM replacement solution getting rid of the capacitor all together would be a great benefit. The current 6-transistor SRAM structure is area consuming, and a challenge is to seek a replacement which would be highly rewarding solution.

These challenges are summarized in Table 2.2.

Table 2.2: Summary of challenges and issues [2.7]

Difficult Challenges for Lg ≥ 16 nm	***Summary of Issues***
Scaling of DRAM and SRAM	DRAM— Adequate storage capacitance with reduced feature size; implementing high-κ dielectric Low leakage in access transistor and storage capacitor Low resistance for bit and word lines to ensure desired speed Improve bit density and to lower production cost in driving toward $4F^2$ cell size SRAM— Maintain adequate noise margin and control key instabilities and soft-error rate Difficult lithography and etch issues

Difficult Challenges for Lg < 16 nm	Summary of Issues
Identification and implementation of new memory structures	Scaling storage capacitor for DRAM DRAM and SRAM replacement solutions

2.4 Architecture: Embedded DRAM

The integrated circuit (IC) industry has begun to deploy architectural techniques such as multiple cores and multiple threads that exploit parallelism to improve the overall chip performance, and to enhance its functionality while maintaining power density and total power dissipation at a manageable level. In a multi-core chip with more than one central processing unit (CPU), the cores can be clocked at a lower frequency while still getting better overall chip performance. Thus, there is a trend for system designers to emphasize integration level, which enables more cores to be put on a chip, instead of raw transistor speed in optimizing system-level performance. In addition, system designers are integrating ever more cache memory onto the processor chip in order to minimize the system performance penalty associated with finite-cache effects. As DRAM cells are significantly smaller than SRAM cells, another high-performance system technology trend is to integrate DRAM cells onto a processor chip

for use in higher-level cache memory (Fig 2.7). With scaling, it is expected that these techniques will be more and more heavily exploited [2.7].

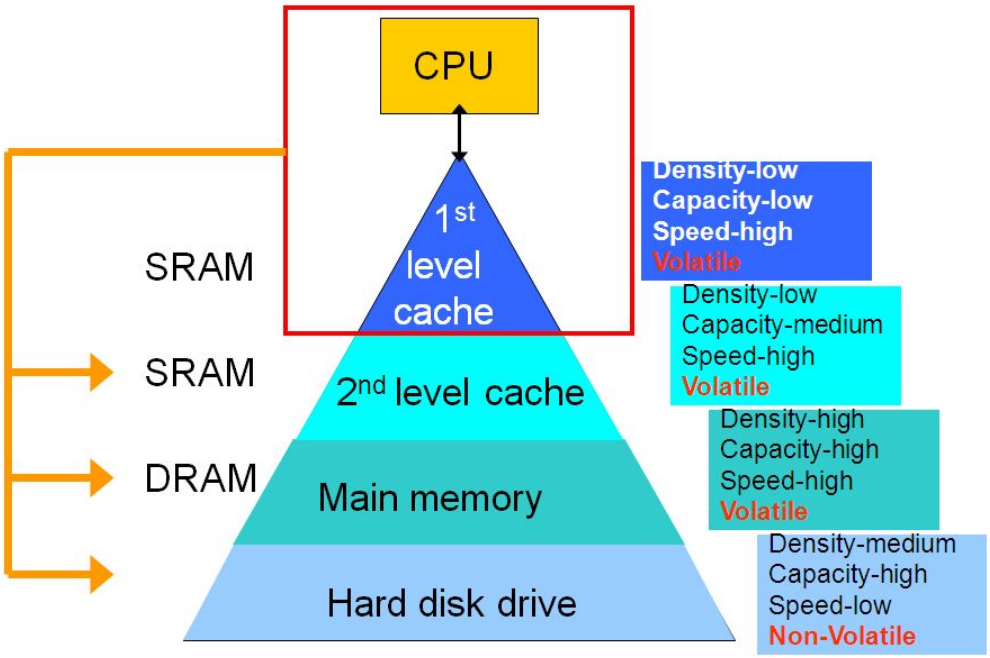


Fig 2.7 Memory hierarchy for a typical computer system [Courtesy: Y. Nishi, H.-S. P. Wong]

2.5 Memory technology requirements and potential solutions for DRAM

Technical requirements for DRAMs have become more difficult with scaling [2.7]. As an example lithography to define smaller dimensions is becoming increasingly more difficult. The process associated with 193 nm argon fluoride (ArF) immersion high-numerical aperture lithography and double patterning technology are keys for 40 nm or smaller half-pitch DRAMs.

In recent years, DRAM cell structure has been migrating to stack capacitor cell. Trench DRAM cell could not survive future scaling due to its difficulties of getting the adequate process and performance of the memory cell. But, even the stack capacitor cell has many technology challenges for 40 nm or smaller size DRAM.

Table 2.3: DRAM technology requirements [2.7]

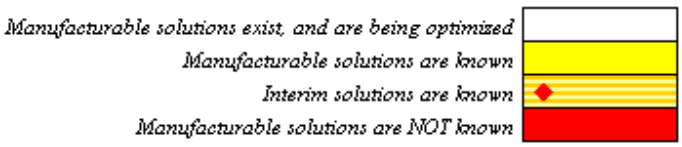
Year in Production	*2009*	*2010*	*2011*
DRAM ½ Pitch (nm) [1]	**50**	**44**	**40**
DRAM cell size (μm^2) [2]	**0.01500**	**0.01162**	**0.00640**
DRAM storage node cell capacitor dielectric: equivalent oxide thickness EOT (nm) [3]	**0.80**	**0.60**	**0.50**
DRAM storage node cell capacitor voltage (V) [4]	**0.60**	**0.55**	**0.55**
Equivalent Electric field of capacitor dielectric, (MV/cm) [5]	**7.5**	**9.2**	**11.0**
DRAM cell FET structure [6]	**RCAT**	**FinFET**	**FinFET**
DRAM cell FET dielectric: equivalent oxide thickness, EOT (nm) [7]	**4.5**	**4.50**	**4.00**
Maximum Word line (WL) level (V) [8]	**2.7**	**2.7**	**2.7**
Negative Word line (WL) use [9]	***yes***	***yes***	***yes***
Equivalent Electric field of cell FET device dielectric (MV/cm) [10]	**6.00**	**6.00**	**6.75**
Cell Size Factor: a [11]	**6**	**6**	**4**
Array Area Efficiency [12]	**0.56**	**0.56**	**0.50**
Minimum DRAM retention time (ms) [13]	**64**	**64**	**64**
DRAM soft error rate (fits) [14]	**1000**	**1000**	**1000**
V_{int} (support FET voltage) [V] [15]	**1.2**	**1.1**	**1**
Support nMOS EOT [nm] [16]	**2.6**	**2.6**	**2.5**
Support PMOS Gate Electrode [17]	**P+Poly/W**	**P+Poly/W**	**P+Poly/W**
Support Gate Oxide [18]	**SiON**	**SiON**	**SiON**
Support min. L_{gate} for NMOS FET, physical [nm] [19]	**80**	**75**	**65**
Support I_{sat-n} [$\mu A/\mu m$] (25C, $V_g = V_d = V_{int}$) [20]	**450**	**450**	**410**
Support min. V_{tn} (25C, $G_{m,max}$, V_d =55mV) [21]	**0.40**	**0.40**	**0.40**
Support I_{sat-p} [$\mu A/\mu m$] (25C, $V_g = V_d = -V_{int}$) [22]	**220**	**210**	**170**
Support min. V_{tp} (25C, $G_{m,max}$, V_d =55mV) [23]	**-0.40**	**-0.40**	**-0.40**

Table 2.3 continued:

2012	*2013*	*2014*	*2015*	*2016*	*2017*	*2018*
36	31	27	24	21	18	16
0.00518	0.00384	0.00310	0.00230	0.00176	0.00130	0.00102
0.40	0.30	0.30	0.30	0.30	0.30	0.40
0.55	0.50	0.50	0.45	0.45	0.40	0.40
13.8	16.7	16.7	15.0	15.0	13.3	10.0
FinFET	FinFET	FinFET	FinFET	FinFET	FinFET	FinFET
4.00	4.00	4	4.00	3.5	3.5	3.5
2.7	2.6	2.6	2.5	2.4	2.3	2.3
yes	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
6.75	6.50	6.50	6.25	6.86	6.57	6.57
4	4	4	4	4	4	4
0.50	0.50	0.50	0.50	0.50	0.50	0.50
64	64	64	64	64	64	64
1000	1000	1000	1000	1000	1000	1000
1	1	1	0.9	0.9	0.9	0.9
2.5	2.3	2	2	1.8	1.6	1.5
P+Poly/W	TiN	TiN	TiN	TiN	TiN	TiN
SiON	HfSiON	HfSiON	HfSiON	HfSiON	HfSiON	HfSiON
65	55	48	42	38	34	32
400	400	400	440	450	450	450
0.40	0.40	0.40	0.40	0.40	0.40	0.40
170	175	170	190	200	200	200
-0.40	-0.40	-0.40	-0.40	-0.40	-0.40	-0.40

Table 2.3 continued:

2019	*2020*	*2021*	*2022*	*2023*	*2024*
14	13	12	10	9	8
0.00078	0.00068	0.00058	0.00040	0.00032	0.00026
0.25	0.20	0.15	0.12	0.10	0.10
0.35	0.35	0.35	0.35	0.35	0.35
14.0	17.5	23.3	29.2	35.0	35.0
FinFET	FinFET	FinFET	FinFET	FinFET	FinFET
3.5	3.5	3.5	3.5	3	3
2.2	2.2	2.2	2.2	2	2
yes	*yes*	*yes*	*yes*	*yes*	*yes*
6.29	6.29	6.29	6.29	6.67	6.67
4	4	4	4	4	4
0.50	0.50	0.50	0.50	0.50	0.50
1000	1000	1000	1000	1000	1000
0.8	0.8	0.7	0.7	0.7	0.7
1.4	1.3	1.3	1.2	1.2	1.2
TiN	TiN	TiN	TiN	TiN	TiN
HfSiON	HfSiON	HfSiON	HfSiON	HfSiON	HfSiON
27	24	21	19	17	16
450	450	450	450	450	450
0.40	0.40	0.40	0.35	0.35	0.35
200	200	200	200	200	200
-0.40	-0.40	-0.40	-0.40	-0.40	-0.40



However, there exist several significant process flow issues from a production standpoint, such as process steps of capacitor formation or high aspect ratio contact etches requiring photoresist that can stand up for a prolonged etch time. To overcome these challenges, the technology related to photoresist with a hard mask layer for pattern transfer is gaining importance. Furthermore, continuous improvements in lithography and etch will be needed [2.7].

On the other hand, with the scaling of peripheral CMOS devices, a low-temperature process flow is required for process steps after formation of these devices. This is a challenge for DRAM cell processes which are typically done after the CMOS devices are formed, and therefore are limited to low-temperature processing. In addition, the planar access device (cell field-effect-transistor (FET)) for the one transistor-one capacitor (1T/1C) cell is becoming difficult to design due to the need to maintain a low level of both subthreshold leakage and junction leakage current to meet the retention time requirements. To attain that, recessed channel cell FET is being adapted and optimization work have been done under half-pitch scaling. But below the 40’s nm half-pitch, FinFET or 3-D type FET will be required to get the high drive current and low-voltage operation. Another challenge is a highly reliable gate insulator. A highly boosted gate voltage is required to obtain higher drain current with the relatively high threshold voltage adopted for the cell FET to suppress the subthreshold leakage current [2.7].

The scaling of the DRAM cell FET dielectric, maximum word-line (WL) level, and the electric field in the cell FET dielectric are critical points for gate insulator reliability concern. To keep the electric field low in the dielectric and the drain region with scaling, process requirements for DRAMs such as front-end isolation, low-resistance materials for the word lines, self-aligned and high aspect ratio etches, planarization, and Cu interconnection will be all needed for future high-density DRAMs.

Since the DRAM storage capacitor gets physically smaller with scaling, the effective oxide thickness (EOT) must scale down sharply to maintain adequate storage capacitance. To scale the EOT, dielectric materials having high relative dielectric constant (κ) will be needed. Therefore MIM (metal insulator metal) capacitors have been adopted using high-κ ($HfSiO/Al_2O_3$, κ~10-25) as the dielectric of 70 nm half-pitch DRAM, and ZrO/HfO as that of 50 nm half-pitch DRAM. And this material evolution will be continued and eventually ultra high-κ (perovskites $\kappa > 50$) material will need to be implemented. Also, the physical thickness of the high-κ insulator should be scaled down to fit the minimum feature size. Due to that, capacitor 3-D structure will be changed from cylinder to pedestal shape.

All in all, maintaining sufficient storage capacitance will pose an increasingly difficult requirement for continued scaling of DRAM devices [2.7].

2.6 Capacitorless DRAM

As maintaining sufficient storage capacitance will pose an increasingly difficult requirement for continued scaling of DRAM devices: “Novel solution to have a capacitorless cell would be highly beneficial” and “A DRAM replacement solution getting rid of the capacitor all together would be a great benefit.” (ITRS 2009) [2.7]

The time evolution of bit cost for DRAM is shown in **Fig. 2.8**. For years, the DRAM bit costs have reduced with Moore’s Law, but now it is hitting the physical limit of scaling.

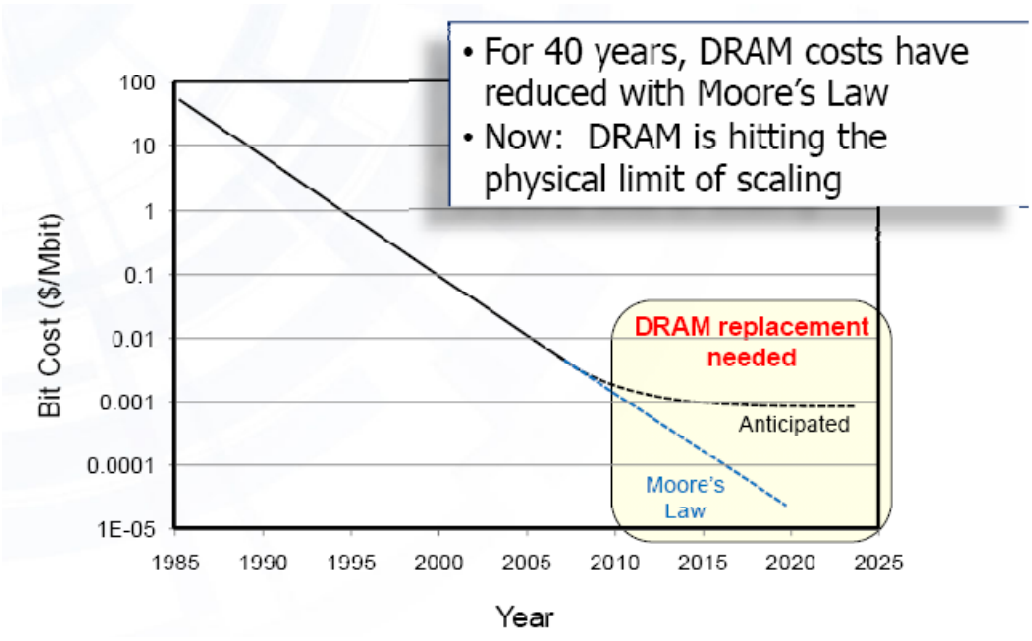


Fig 2.8. Bit cost evolution of DRAM [Source: Innovative Silicon]

Recently the capacitorless single-transistor (1T) DRAMs have attracted attention, due to the lack of the capacitor and the problems associated with the scaling of the capacitor and also its ability to achieve higher device density. The information is stored as different charge levels at a capacitor in conventional 1T/1C DRAM, whereas the 1T DRAM emlpoys floating body effects within the transistor to store the information (**Fig 2.9)**.

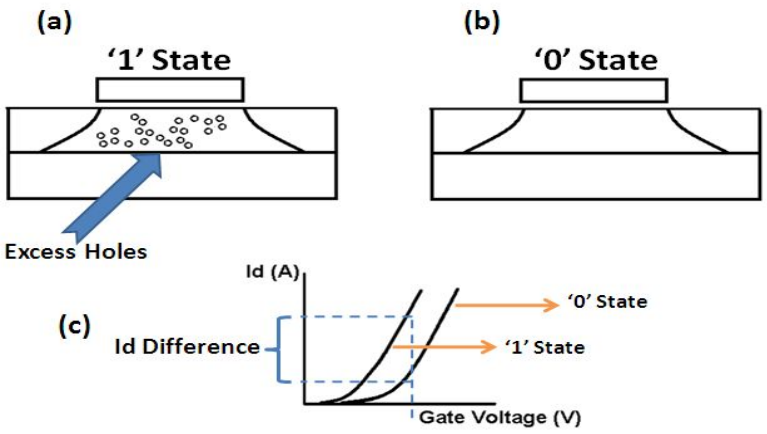


Fig 2.9 1T DRAM device schematics. In (a) "1" state, there are excess holes in the body, whereas in (b) "0" state, there is not. The device state is determined by sensing (c) the drain current. The excess holes in the body cause an decrease in threshold voltage corresponding to state "1"

Fig. 2.10 compares the scanning electron microscope (SEM) image of a conventional 1Transistor/1Capacitor (1T/1C) DRAM cell with a single transistor (1T) DRAM cell. The absence of the capacitor is advantageous in terms of scaling, process

and fabrication complexity, compatibility with the logic processing steps, device density, yield and cost (**Fig. 2.11)**.

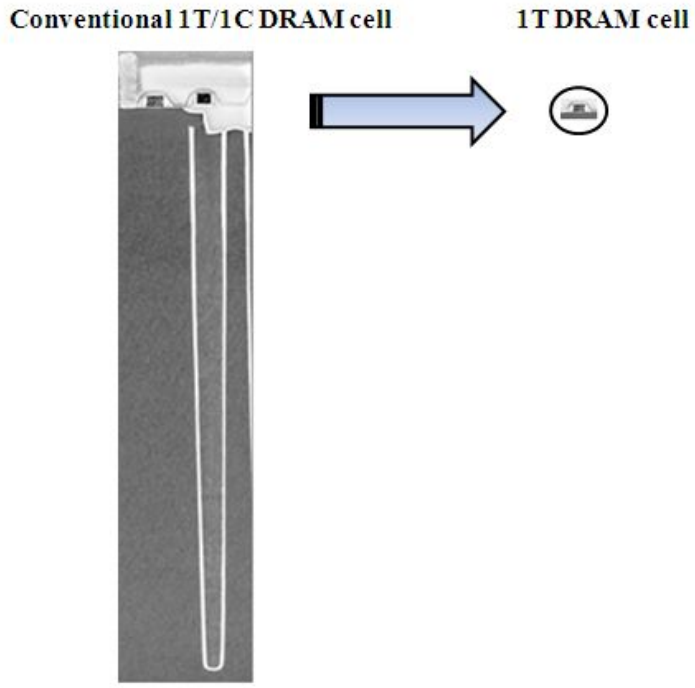


Fig 2.10 Comparison of a conventional 1Transistor/1Capacitor (1T/1C) DRAM cell with a single transistor (1T) DRAM cell

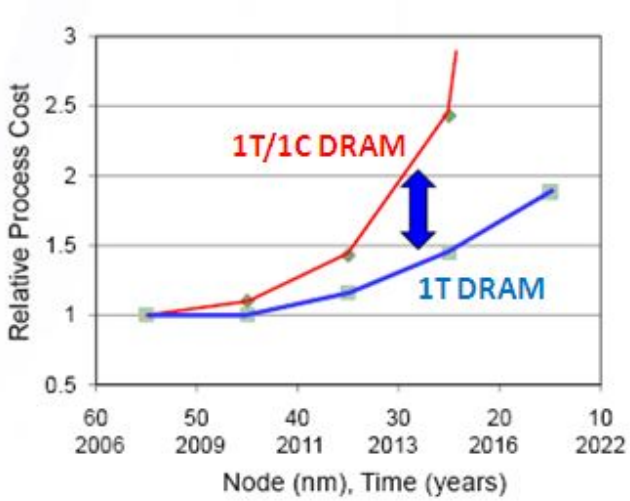


Fig 2.11 Processing costs comparison for conventional 1Transistor/1Capacitor (1T/1C) DRAM with single transistor (1T) DRAM [Source: Innovative Silicon]

Due to all these advantages of being capacitorless, and to solve the scaling problem of conventional DRAM, and to come up with capacitorless DRAM replacements, this work is focused on creating novel single transistor DRAM technologies.

2.7 References

[2.1]Kurt Hoffmann,"System Integration: From Transistor Design to Large Scale Integrated Circuits", Wiley; 1st edition (April 30, 2004)

[2.2] J. A. Mandelman, R. H. Dennard, G. B. Bronner, J. K. DeBrosse, R. Divakaruni, Y. Li, and C. J. Radens, “Challenges and future directions for the scaling of dynamic random-access memory (DRAM)” IBM J. RES. & DEV., vol. 46 no. 2/3 March/May 2002

[2.3] K. Itoh, “VLSI Memory Chip Design”, Springer-Verlag, March 2001

[2.4] Robert H. Dennard, "Field-Effect Transistor Memory," U.S. Patent No.3387286, filed on 14 July 1967 and approved on 4 June, 1968

[2.5] H.S. Kim, D.H. Kim, J.M. Park, Y.S. Hwang, M. Huh, H.K. Hwang,N.J. Kang, B.H. Lee, M.H. Cho, S.E. Kim, J.Y. KIM, B.J. Park, J.W. Lee, D.I. Kim, M.Y. Jeong, H.J. Kim, Y.J. Parkand Kinam Kim, "An Outstanding and Highly Manufacturable 80nm DRAM Technology" IEEE IEDM 2003, pp. *IEDM* 03-411--03-414

[2.6] G. Bronner, H. Aochi, M. Gall, J. Gambino, S. Gernhardt, E. Hammerl, H. Ho, J. Iba, H. Ishiuchi, M. Jaso, R. Kleinhenz, T. Mii, M. Narita, L. Nesbit, W. Neumueller, A. Nitayama, T. Ohiwa, S. Parke, J. Ryan, T. Sato, H. Takato, S. Yoshikawa , "A fully planarized 0.25 μm CMOS technology for 256 Mbit DRAM and beyond," Symposium on VLSI Technology, 1995. Digest of Technical Papers. 1995, pp.15-16, 6-8 Jun 1995

[2.7] International Technology Roadmap for Semiconductors (ITRS) 2009 Edition

[2.8] Kinam Kim, “Perspectives on giga-bit scaled DRAM technology generation, Microelectronics Reliability”, Volume 40, Issue 2, Pages 191-206, 28 February 2000

[2.9] Serguei Okhonin, “In Search of a Better DRAM”, Innovative Silicon, Inc. [http://www.innovativesilicon.com/en/pdf/In_Search_of_a_Better_DRAM_white_paper.pdf]

Chapter 3

Vertical Double-Gate (DG) Capacitorless Single-Transistor DRAM

3.1 Introduction

Capacitorless DRAM, because of its smaller size, has been researched aggressively recently. However, to date, most published results show either only simulations [3.1]–[3.3] or limited and expensive experimental approaches such as wafer bonding approach [3.4] or SOI substrate [3.5]. We experimentally demonstrate a double-gate (DG) capacitorless single-transistor (1T) DRAM on a bulk silicon wafer (**Fig. 3.1** and **Fig. 3.2**), which is a big advantage in terms of processing and process integration. In this device, one of the MOS gates is used as a conventional switching transistor, whereas the other (back) gate is used to create the floating-body storage

node. By reverse biasing the back gate, one can keep the excess holes in the body and obtain memory operation even for the highly scaled devices under fully depleted condition. This vertical source/drain structural configuration does not suffer from width discretization issues as experienced with FINFETs and provides a simple back gate option stemming from electrical isolation between two gates. Also, since the proposed structure has a low body impurity concentration, it has a lower junction leakage current, hence longer retention time. Finally, a DG structure effectively terminates the field lines from the drain; one can use a relatively thicker gate oxide than what is used for the logic transistors, which helps increase the body coefficient. A fully depleted SOI MOSFET's body coefficient is defined as $dV_t/dV_{BS} = -3T_{ox}/T_{Si}$ [3.2]. Hence, with a thinner silicon film thickness (T_{Si}) and thicker gate oxide (T_{ox}), significant threshold voltage (V_t) change can be obtained by using relatively smaller body potential changes [3.6].

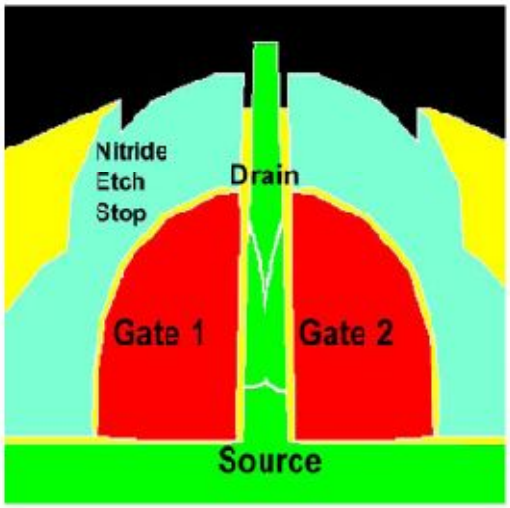


Fig. 3.1 Schematic of the vertical DG 1T DRAM from TSUPREM4^{TM} simulation (final channel doping = 4.7×10^{15} cm^{-3}).

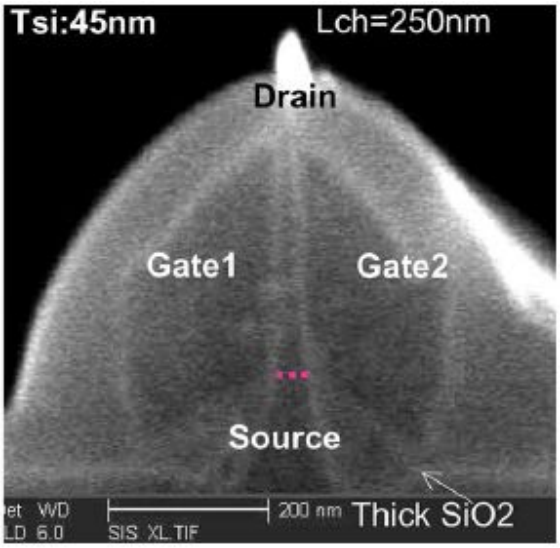


Fig. 3.2 Scanning electron microscope (SEM) image of the vertical DG 1T DRAM clearly depicting that there is no electrical connection between gate 1 and gate 2.

A similar DG DRAM is reported by using a FINFET device [3.1], albeit in simulations only. In addition, Ban et al. [3.5] report an independently controlled DG floating-body cell on SOI. We not only provide both simulation and experimental results but also our vertical DG structure has the added advantage of being able to be built on bulk silicon wafer, resulting in cost reduction as well as easier process integration.

3.2 Memory Operation

3.2.1 Program

Programming in our structure is achieved by using impact ionization at the drain side. The body can be saturated with holes in a few nanoseconds [3.7]–[3.9]. During programming, the impact-ionization-generated holes create a quasi-floating-body at the back interface (Gate 2) [3.10].

3.2.2 Hold

The retention time depends on how long the carriers (holes, in this case) generated by impact ionization can be kept in the floating-body. Applying a negative bias to the back gate helps keep the holes in the floating-body storage node. The memory charge loss occurs primarily due to carrier recombination and leakage current through tunneling. Because tunneling-induced leakage depends exponentially on electric field and because DG structures exhibit low vertical electric field, the charge loss through leakage is naturally mitigated. Furthermore, by using a combination of a low gate bias, thicker gate oxide, and thin body, the total electric field at the junctions can be reduced [3.2]. This, in turn, results in reduced band-to-band tunneling, whereas a thin silicon body results in very low drain-induced barrier lowering. Thus, a DG structure is capable of much higher retention times.

3.2.3. Read

The read operation is performed by turning on the front MOS structure without disturbing the memory state stored in. The sensed drain current is altered by the presence or absence of charge on the floating-body through a change in threshold voltage (V_t). In the Programmed/"1" state, cells have excess holes at the back interface. Excess holes created by impact ionization during programming operation, due to the increased body potential, lower V_t, thus resulting in a higher drain current. A sense amplifier can sense the current difference between "1" and "0" states [3.11].

3.3 Fabrication

The process flow developed by Hoon Cho and described in detail in [3.12, 3.13] is highlighted by several key innovations, such as the Si_3N_4/poly-silicon based robust spacer process, which was recently demonstrated to yield pillars down to 5-nm thickness [3.13]. The spacer process hard mask is subsequently used to etch a specifically tailored flared-out (source-end) pillar profile in Si to obtain low parasitic resistance at the source end. The implant sequence and conditions were narrowed by using extensive TSUPREM4^{TM} simulations. A key innovation, which enabled the separation of the two gates around the body, comprises of a novel self aligned process, which does not require lithography and is highly reproducible. It entails depositing a blanket thin low temperature SiO_2 LTO hard mask, followed by a N+ poly-Si

deposition. A subsequent curvature-dependent differential oxidation rate of poly-Si results in its complete consumption at the top of the pillar while leaving its thin unoxidized part at the sidewall, corner, and the bottom of the pillar. This thin film is used to etch the underline LTO, which, in turn, is used as a hard mask to remove the underlying poly-Si gate selectively from the top of the pillar, thereby separating the two gates. Another process highlight is the drain contact technology. This entails depositing and spacer etching a Si_3N_4 film prior to depositing the passivation. The nitride layer serves as a selective etch stop for the drain contact etch. Although, we demonstrate ~200-nm channel length (L_{ch}) as a first pass, no intrinsic scaling limit down to sub-20 nm is foreseeable.

3.3.1 Process Flow

The process flow for a vertical double-gate transistor including spacer, isolation and the contact steps are shown in **Fig. 3.3**. The body doping of $3e15cm^{-3}$ is aimed for a fully depleted operation.

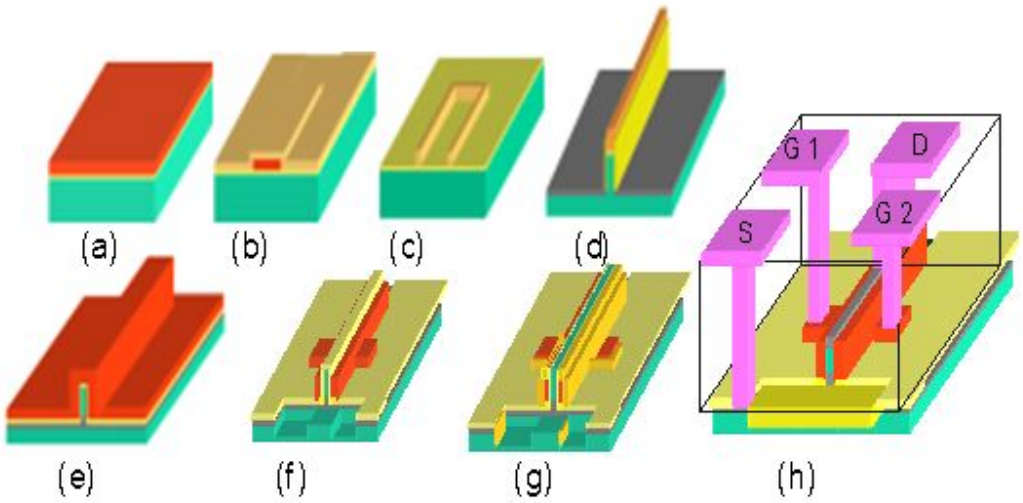


Fig. 3.3 Process flow for the vertical DG DRAM [3.12]

(a) p-type substrate, dry pad oxidation, poly-Si block deposition

(b) Poly block etch, few nm thick (~T_{Si}) nitride deposition

(c) Blanket nitride spacer etch, remove poly block by TMAH etch

(d) Silicon pillar (fin) etch, LTO sidewall, Source implant

(e) Remove all nitride and pad oxide, anneal, the thick bottom oxide process including gate oxidation/deposition

(f) Self align process, Isolation trench etch, dry oxidation

(g) Nitride etch stop deposition and blanket etch for drain contact, Drain implant

(h) PSG deposition, RTA, contact etch, metal deposition and etch [Courtesy: H. Cho].

3.3.2 Spacer process, Si pillar etch, source implant

The spacer process is shown in steps (a) -(c) in **Fig. 3.3**. This involves defining a spacer block using coarse lithography, depositing a spacer material of desired thickness, etching it anisotropically, and finally, selectively wet etching the block material, leaving a free standing

spacer to be subsequently used as a hard mask. Si_3N_4 is picked for the spacer material due to its high uniformity and conformal step coverage, and polysilicon is chosen as the block material.

A 20nm of SiO_2 pad layer was a starting point, and this thickness is chosen, by keeping in mind the SiO_2 etch-selectivity to both polysilicon block and Si_3N_4 spacer etch steps. The polysilicon block, which dictates the spacer height, is chosen to be three times its width. This rule of thumb provides a balance between a minimum dictated by hard-mask thickness requirement and a maximum determined by its mechanical stability (**Fig. 3.3 (a)**). After the polysilicon block etch, different Si_3N_4 spacer thicknesses ranging from 5nm to 50nm were deposited (**Fig. 3.3 (b)**) and etched anisotropically. The polysilicon block was removed using TMAH, leaving nitride spacer behind (**Fig. 3.3 (c)**).

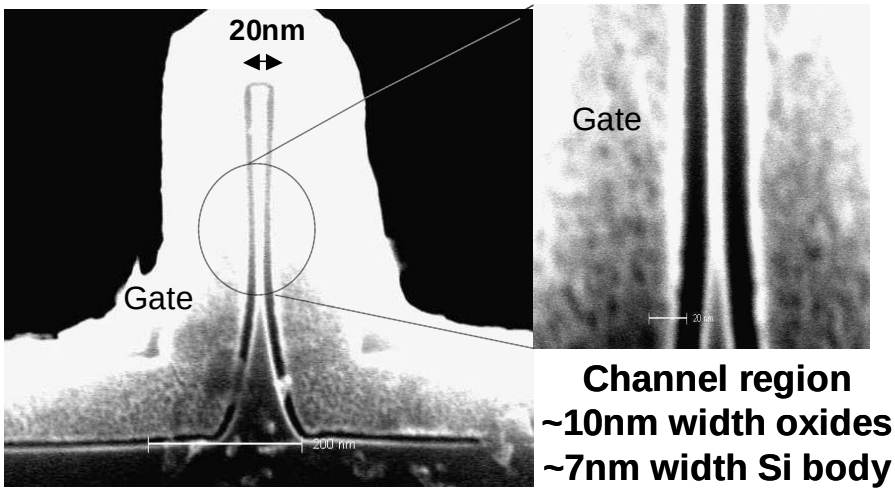


Fig. 3.4 Cross-section SEM picture of the vertical structure with 7nm silicon body thickness. Gate oxide is selectively etched by HF dip for SEM contrast. [Courtesy: H. Cho [3.12]]

Subsequently, a Si fin is etched in a $Cl_2/HBr/O_2$ chemistry with the Si_3N_4 spacer as the hard mask. The etch is carefully engineered to yield a fanned-out shape at the base of the fin,

where the thickness of the fin gradually increases with depth (**Fig. 3.4**). This results in a lower source resistance. This is followed by an arsenic source implant step (7^{o} angle, $5x10^{15}$ cm^{-2} dose and 30keV energy). Before implant a 45nm thick sacrificial low temperature oxide (LTO) is deposited by low pressure CVD to protect the channel, whereas the original Si_3N_4 spacer mask protects the drain (**Fig. 3.3 (d)**). Finally, a 5min, 1000^{o}C anneal is performed for dopant activation and diffusion after removing the nitride and oxide films.

Normally, the silicon etch and source implant process should be followed by the gate oxidation step. However, stress effects at the concave surface and the orientation difference between the bottom and the sidewall (110 vs. 100) results in a significant oxide growth retardation at corners and bottom of the fins. This, in turn, results in a large leakage current as well as unwanted parasitic capacitance with the bottom substrate. To overcome this classic problem of vertical structures, and get thicker field and corner oxide, a manufacturable novel process for the thick/bottom oxide is employed.

3.3.3 Self-aligned process for thick corner/bottom oxide

This process creates a polysilicon hard mask which is self-aligned to a vertical topography without using lithography and the associated masks. The idea is to exploit the curvatures and crystallographic orientation differences on a vertical topography to obtain significantly varying oxidation rates for a doped polysilicon film. The areas of faster oxidation rates result in complete polysilicon consumption, while, slower rate areas have a thin polysilicon left.

This highly controllable left-over layer serves as a self-aligned hard mask after the oxide is removed (a simple HF dip etch). In addition to being useful for obtaining thick bottom and corner oxide, some of the other applications of this process are 1) separation of gates and 2) creating self-aligned drain contact for thin vertical devices without any mask or CMP [3.15]. The process is highly reproducible and uniform as explained in [3.16].

This aforementioned process is employed to create thick bottom and corner SiO_2 film. A 40nm thick LTO is deposited on the vertical fin made using the spacer process, followed by an in-situ n^+ doped, 65nm thick, polysilicon layer. Subsequently, the oxidation process (wet ambient, 800^{o}C for 14min) is performed to create the self-aligned hard.

The oxidation temperature is limited to 800^{o}C to avoid dopant diffusion and segregation. The hard mask is used to selectively remove LTO from the sidewalls and the top of the fin (wet HF dip etch), while protecting it at the bottom and the corners.

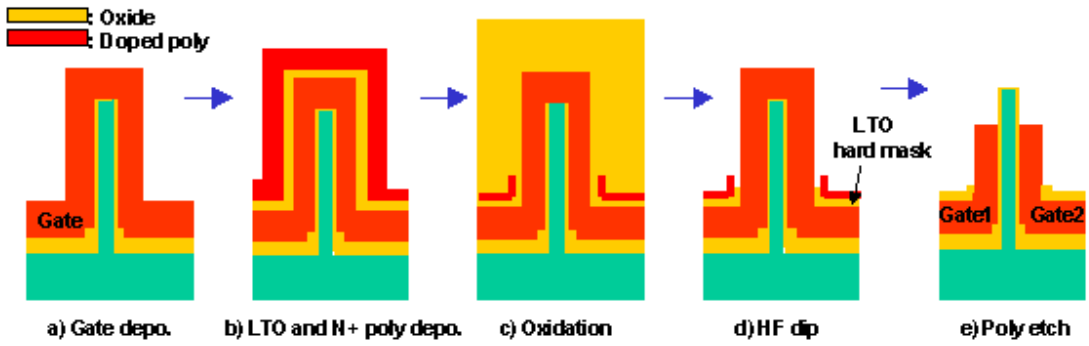
This is followed by a sacrificial gate oxide to clean-up the vertical MOS interface. This process also consumes the remaining polysilicon hard mask from the previous step. A final gate stack consisting of a 8nm thick gate oxide (sidewall) and an in situ n^+-doped poly-Si is deposited (**Fig. 3.3 (e)**).

3.3.4 Self-aligned process for the gate separation

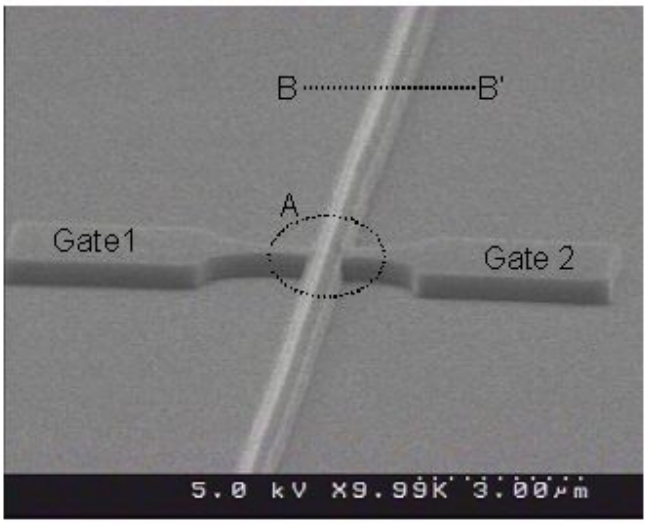
The self-aligned process that have been developed can also be used for the gate separation process and is one of the key novel unit process steps of this vertical DG device. The

gate is selectively etched only above the pillar to electrically isolate the two cells on the same pillar.

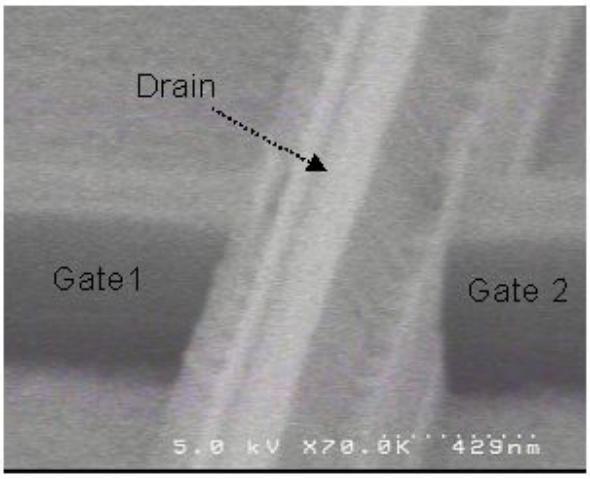
Fig 3.5 (a) shows the gate separation process using the self-aligned process. A 40nm thick LTO SiO_2 is deposited on the gate followed by a 40nm thick n+ doped polysilicon. During wet oxidation at 800^{o}C, a faster oxidation at the top of the pillar results in complete polysilicon consumption, while a slower oxidation leaves a thin polysilicon at the bottom and corners (**Fig 3.5 (a) step c**). This remaining polysilicon layer is used as a hard mask during HF dip to protect the underlying SiO_2 (**Fig. 3.5 (a) step d**), which in turn, is used as a hard mask to etch the gate selectively from just the top of the pillar.



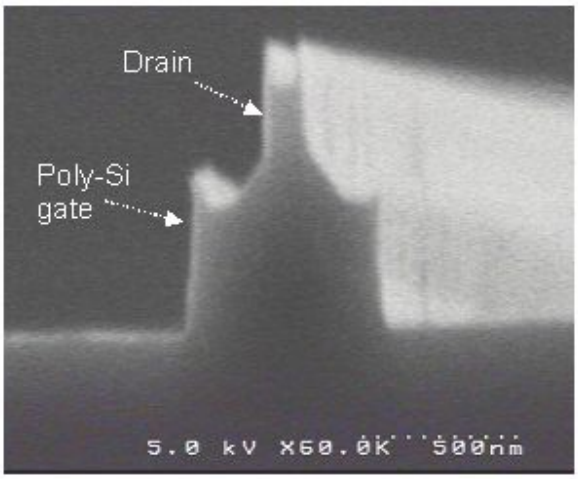
(a)


B
B'
A
Gate1
Gate 2
5.0 kV X9.99K 3.00µm

(b)


Drain
Gate1
Gate 2
5.0 kV X70.0K 429nm

(c)



(d)

Fig. 3.5 Gate isolation process using the self-align process. (a) the process flow [3.13], (b) bird-view SEM picture of the vertical DG device after the gate separation process, (c) SEM: inside of the 'circle A' showing the separated two gates, (d) cross-section view SEM along the B-B' cut line. [Courtesy: H. Cho].

The polysilicon hard mask layer is also etched away during the gate etch, which simplifies the process. **Fig. 3.5 (b)** shows the SEM picture at the gate etch stage. In **Fig. 3.5 (c)** and **(d)**, the zoomed-in SEM pictures clearly show the two gates separated by the novel process, maintaining the gate contact region. The "crowns" at the topside of gate edges are due to native oxide during the gate etch which is very anisotropic and selective to oxide. The crown can be eliminated by a short oxidation and HF dip. The gate definition step is followed by active area isolation using a trench etch process, and dry oxidation to prevent leakage.

3.3.5 Drain Contact Process

For the drain contact process the first step is to deposit a compressive Si_3N_4 layer that is slightly thicker than gate polysilicon, and subsequently spacer-etch it. This layer serves several purposes in the vertical DG device structure. Its main purpose is to serve as an etch-stop for the drain contact hole etch and consequently, provide control over channel length and contact/series resistance. It retains about 6:1 etch selectivity over the passivation oxide, phosphosilicate glass (PSG). The Si_3N_4 layer also serves as a mask, protecting the channel from the subsequent drain implant. Without this layer, the device will have a large gate-drain overlap, increasing parasitic capacitance between the gate and the drain. Finally, the Si_3N_4 layer also plays an important role in dictating the stress in the channel region, which impacts the transistor transport properties.

The Si_3N_4 layer deposition and etch is followed by the drain implant **(**arsenic, $5x10^{15}$ cm^2 dose, 30keV energy) (**Fig. 3.3 (g)**), the PSG layer deposition, and the RTA step (12sec at 1050^oC) for dopant activation. The high selectivity of Si_3N_4 to oxide etch allows a simultaneous contact hole etch for drain, source, and gate. Finally, aluminum is deposited and patterned (**Fig 3.3 (h)**), and a forming gas anneal (FGA) for 30 minutes at 400^oC is performed.

3.4 Electrical Results

The impact-ionization-generated holes (during programming) are swept to the back gate, where they alter the device V_t. A back-gate negative bias holds the holes during retention. The presence and absence of holes alter V_t, and the difference in the

drain current (I_d) is sensed to distinguish between the two memory states. **Fig. 3.6** shows a prominent kink in the I_d versus V_d curves, indicating V_t alteration on a device with the following parameters: channel length (L) = 250 nm, oxide thickness (T_{ox}) = 8 nm, and silicon thickness (T_{Si}) = 50 nm. An explicit DRAM operation is demonstrated by using SENTAURUS^{TM} simulations with the device parameters of channel length (L) = 250 nm, oxide thickness (T_{ox}) = 8 nm, and silicon thickness (T_{Si}) = 100 nm.

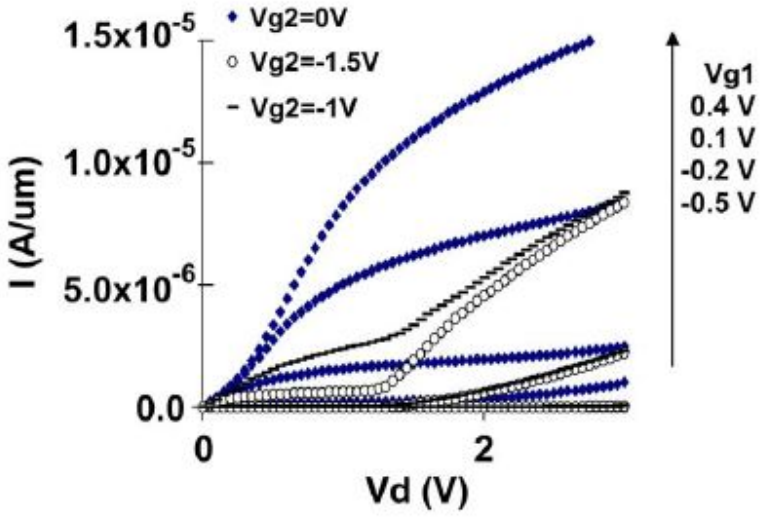
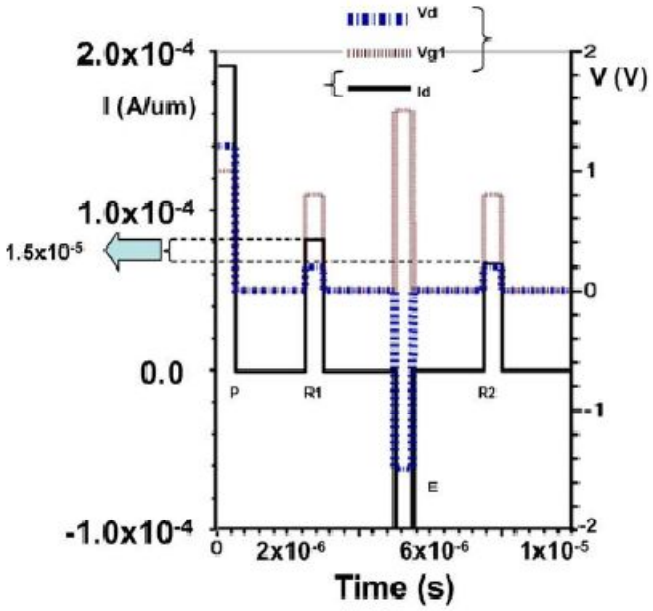


Fig. 3.6. Experimental data showing current (I_d) versus drain voltage (V_d) for different gate 1 voltages (V_{g1}) of −0.5, −0.2, 0.1, and 0.4 V. Three sets of curves correspond to different back gate voltages (V_{g2}) (0, −1, and −1.5 V). Figure shows the onset of kink effect caused by excessive hole accumulation aided by negative V_{g2}.

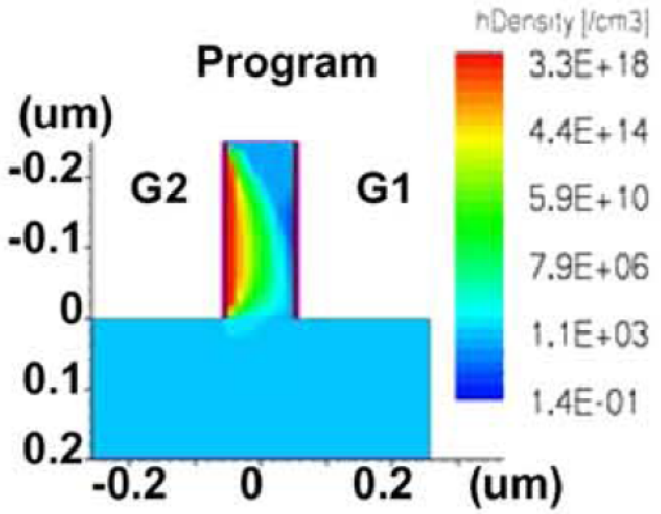
Table 3.1 Operating Voltages for the DRAM cell

	Program (Write "1")	Erase (Write "0")	Read	Hold
Gate 1 Voltage (V)	1	1.5	0.8	0
Drain Voltage (V)	1.2	-1.5	0.2	0
Gate 2 Voltage (V)	-1.5	-1.5	-1.5	-1.5

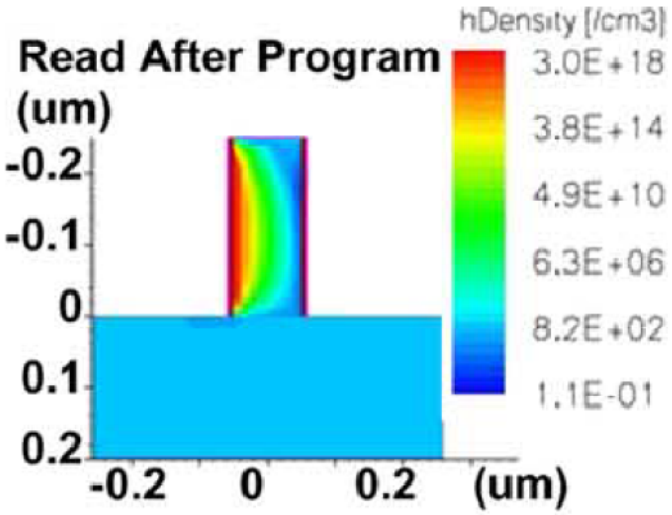
Table 3.1 shows the voltages on terminals during different memory operations. **Fig. 3.7(a)** shows the temporal profile for the applied V_d, V_{g1}, and the resulting I_d during program (P), during read after program (R1), during erase (E), and during read after erase (R2). A clear difference in I_d(≅15 µA/µm) for R1 and R2 confirms a memory operation. **Fig. 3.7 (b-e)** show the hole densities during program (P), during read after program (R1), during erase (E), and during read after erase (R2). **Fig. 3.7(b)** specifically shows the simulated enhancement in the hole density after programming (P) and its significant reduction after erase cycle (E) **Fig. 3.7(d)**.



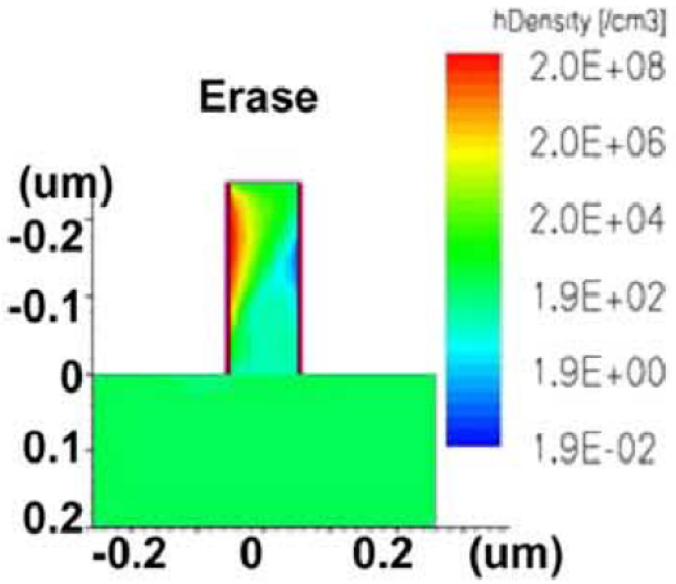
(a)


hDensity [/cm3]
Program
(um)
-0.2
-0.1
0
0.1
0.2
G2
G1
3.3E+18
4.4E+14
5.9E+10
7.9E+06
1.1E+03
1.4E-01
-0.2
0
0.2
(um)

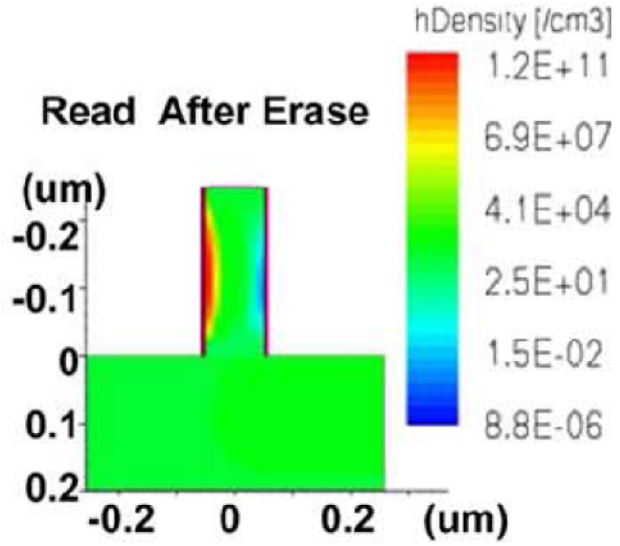
(b)


hDensity [/cm3]
Read After Program
(um)
-0.2
-0.1
0
0.1
0.2
3.0E+18
3.8E+14
4.9E+10
6.3E+06
8.2E+02
1.1E-01
-0.2
0
0.2
(um)

(c)



(d)



(e)

Fig. 3.7 (a) Simulation result showing the temporal profiles for the applied voltages (Gate1 voltage, V_{g1}; drain voltage, V_d). The cell is first programmed ("P" in the figure), then read (R1),

erased (E), and read again (R2). Figure shows the difference in the reading drain current for the programmed and erased states (15 μA/μm).

(b-e) SENTAURUS^{TM} simulations showing 2-D hole density profile at four time instances corresponding to:

(b) program (P),

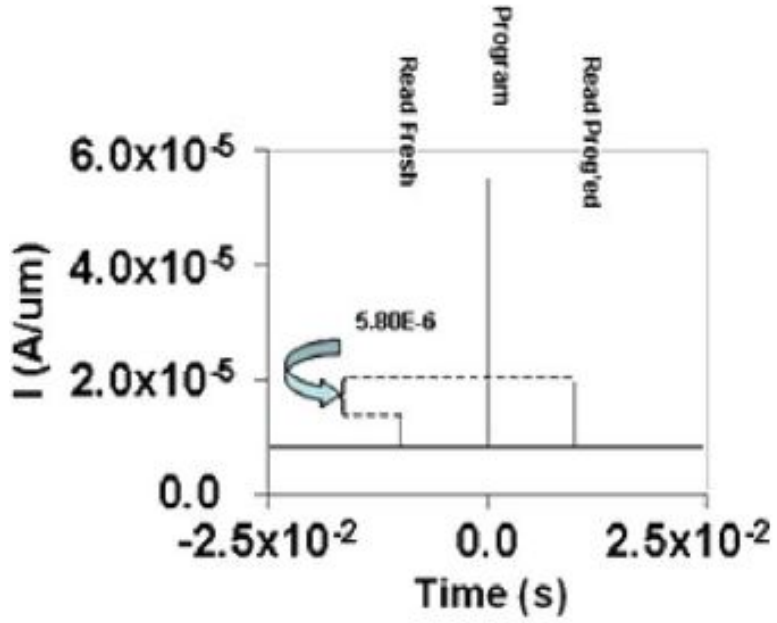
(c) read after program (R1),

(d) erase (E)

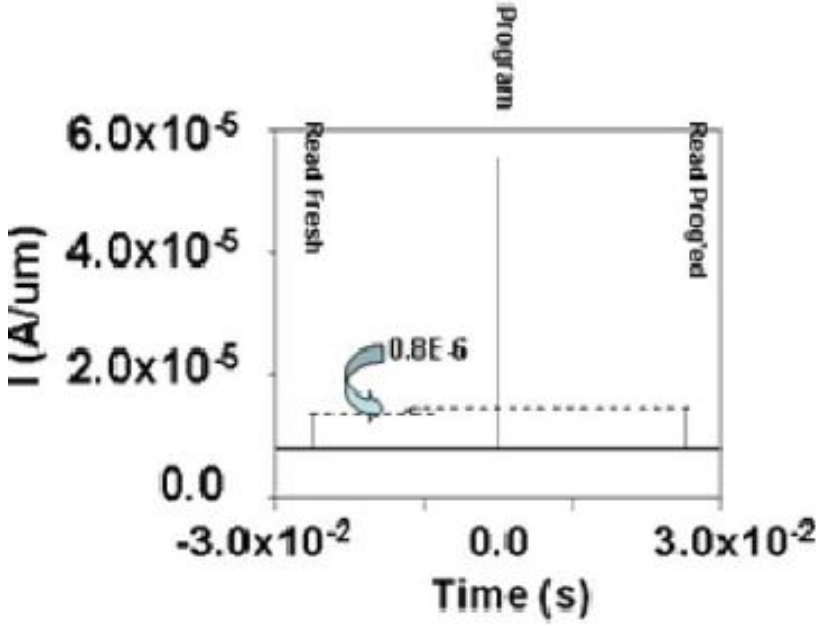
(e) read after an erase (R2).

A clear increase in hole density is observed near the back gate after programming.

Fig. 3.8 (a) and **(b)** show experimentally measured I_d during a fresh cell read, during programming, and during a second read after cell programming on a device with the parameters of channel length (L) = 250 nm, oxide thickness (T_{ox}) = 8 nm, and silicon thickness (T_{Si}) = 50 nm at room temperature. In experimental case, lower back gate voltage is used compared to the simulations in order to obtain an adequate drain current. The time between program and second read is varied and corresponds to the retention time. A 3–6-μA/μm I_d difference between fresh and programmed cell reads is maintained up to 10 ms, whereas a sensed current difference is approximately 1 μA/μm after 25 ms of retention.


Read Fresh
Program
Read Prog'ed
6.0x10-5
4.0x10-5
2.0x10-5
0.0
I (A/um)
5.80E-6
-2.5x10-2
0.0
2.5x10-2
Time (s)

(a)


Program
Read Fresh
Read Prog'ed
6.0x10-5
4.0x10-5
2.0x10-5
0.0
I (A/um)
0.8E-6
-3.0x10-2
0.0
3.0x10-2
Time (s)

(b)

Fig. 3.8 (a) Experimental data showing the drain current during the stages of a fresh cell read, cell programming, and read after cell programming. The program time was 500 ns, and the retention corresponding to the time after programming and before the second read was also 10 ms. The figure (arrow) clearly indicates the difference in the I_d(5.80×10^{-6} A/μm) for fresh and programmed cell reads, corresponding to these two states. (b) Similar experimentally measured plot as in Fig. 3.8(a), except now, the time between program and the second read corresponding to retention is 25 ms. Now, the current difference between two states reduces slightly to 1 μA/μm (arrow), indicating that the device exhibits a reasonable retention of up to 25 ms.

3.5 Conclusions

We have experimentally demonstrated a DG capacitorless 1T DRAM on a bulk silicon wafer, with the measured retention times of up to 25 ms [3.17]. The scalability limits for conventional 1T-1C DRAM is the capacitor, whereas in the 1T DRAM, this limit is due to the lithography [3.1]. Our devices do not have capacitors, and their gate lengths are not defined by lithographic processes. In fact, various process innovations enable gate length definition using well-controlled and characterized etch processes. Because this process is capable of much smaller features, it facilitates scaling of the DRAM down to 22-nm technology node.

3.6 References

[3.1] T. Tanaka, E. Yoshida, and T. Miyashita, “Scalability study on a capacitorless 1T-DRAM: From single-gate PD-SOI to double-gate FinDRAM,” in *IEDM Tech Dig.*, Dec. 13–15, 2004, pp. 919–922.

[3.2] C. Kuo, T.-J. King, and C. Hu, “A capacitorless double-gate DRAM cell,” *IEEE Electron Device Lett.*, vol. 23, no. 6, pp. 345–347, Jun. 2002.

[3.3] E. Yoshida, T. Miyashita, and T. Tanaka, “A study of highly scalable DGFinDRAM,” *IEEE Electron Device Lett.*, vol. 26, no. 9, pp. 655–657, Sep. 2005.

[3.4] C. Kuo, T.-J. King, and C. Hu, “A capacitorless double-gate DRAM cell design for high density applications,” in *IEDM Tech. Dig.*, 2002, pp. 843–846.

[3.5] I. Ban, U. E. Avci, U. Shah, C. E. Barns, D. L. Kencke, and P. Chang, “Floating body cell with independently-controlled double gates for high density memory,” in *IEDM Tech. Dig.*, Dec. 11–13, 2006, pp. 1–4.

[3.6] C. Kuo, T.-J. King, and C. Hu, “A capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications,” *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2408–2416, Dec. 2003.

[3.7] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan, “A capacitor-less SOI 1T-DRAM concept,” in *Proc. Int. SOI Conf.*, 2001, pp. 153–154.

[3.8] T. Ohsawa, K. Fujita, T. Higashi, Y. Iwata, T. Kajiyama, Y. Asao, and K. Sunouchi, “Memory design using one-transistor gain cell on SOI,” in *Proc. ISSCC*, 2002, pp. 152–153.

[3.9] J. Gautier, K. A. Jenkins, and J. Y.-C. Sun, “Body charge related transient effects in floating-body SOI NMOSFETs,” in *IEDM Tech. Dig.*, Dec. 10–13, 1995, pp. 623–626.

[3.10] J. P. Colinge, “Reduction of kink effect in thin-film SOI MOSFETs,” *IEEE Electron Device Lett.*, vol. 9, no. 2, pp. 97–99, Feb. 1988.

[3.11] T. N. Blalock and R. C. Jaeger, “A high-speed clamped bit-line current mode sense amplifier,” *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 542–548, Apr. 1991.

[3.12] H. Cho, P. Kapur, P. Kalavade, and K. C. Saraswat, “A low power, highly scalable, vertical double gate MOSFET using novel processes,” *IEEE Trans. Electron Devices*, vol. 55, no. 2, pp. 632–639, Feb. 2008.

[3.13] H. Cho, P. Kapur, P. Kalavade, and K. C. Saraswat, “A novel spacer process for sub-10-nm-thick vertical MOS and its integration with planar MOS device,” *IEEE Trans. Nanotechnol.*, vol. 5, no. 5, pp. 554–563, Sep. 2006.

[3.14] D. B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon-I. Experiments," *IEEE Trans. Elec. Dev.*, vol. ED-34, p.1008, 1987.

[3.15] H. Cho, P. Kapur, P. Kalavade, and K. C. Saraswat, "Geometry dependence of poly-Si oxidation and its application to self-aligned, maskless process for nano-scale vertical CMOS structures," *Electro-Chemical Soc. Transactions*, vol. 3, no. 2, pp 403-414, 2006.

[3.16] H. Cho, “Low Power, Highly Scalable, Vertical Flash Memory Cell and MOSFET”, PhD Dissertation, Stanford University, CA, September 2007

[3.17] M. G. Ertosun, H. Cho, P. Kapur, K. C. Saraswat, “A Nanoscale Vertical Double-Gate Single-Transistor Capacitorless DRAM” IEEE Electron Device Letters, Volume 29, Issue 6, June 2008

Chapter 4

Capacitorless Single-Transistor Quantum Well DRAM

4.1 Introduction

Capacitorless DRAM, because of its smaller size, has been widely researched recently [4.1,4.2,4.3,4.4,4.5]. In chapter 3 experimental demonstration of a double gate (DG) capacitorless single transistor (1T) DRAM is described on a bulk Silicon wafer [4.6] which is a big advantage in terms of processing and process integration. By reverse biasing the back gate, one can keep the excess holes in the body and obtain a memory operation even under fully depleted condition. This source/drain structural configuration does not suffer from width discretization issues as experienced with FINFETs, and provides a simple back gate option stemming from electrical isolation between two gates.

Furthermore, the proposed structure has a low body impurity concentration, hence it has a lower junction leakage current, which in return yields a longer retention time. Finally, a DG structure with better electrostatics effectively terminates the field lines from the drain resulting in reduced drain induced barrier lowering. Therefore, one can use a relatively thicker gate oxide than what is used for logic transistors, which helps obtaining a higher body coefficient. For a fully depleted SOI MOSFET with body thickness of T_{si}, gate oxide of T_{ox} the body coefficient is defined as $dV_t/dV_{BS}= -3T_{ox}/T_{si}$ [4.2], where V_t is the threshold voltage and V_{BS} is the back gate bias. Hence, with a thinner T_{si} and thicker T_{ox}, substantial V_t change can be obtained using relatively smaller body potential changes [4.6]. A similar DG DRAM is reported using a FINFET device [4.1], albeit in simulations only. In addition, [4.5] reports independently controlled double-gate floating body cell on SOI. Our vertical DG structure has the added advantage of being able to fabricate on bulk silicon wafer, resulting in cost reduction as well as easier process integration and lastly but not least importantly, higher density due to its vertical 3D structure.

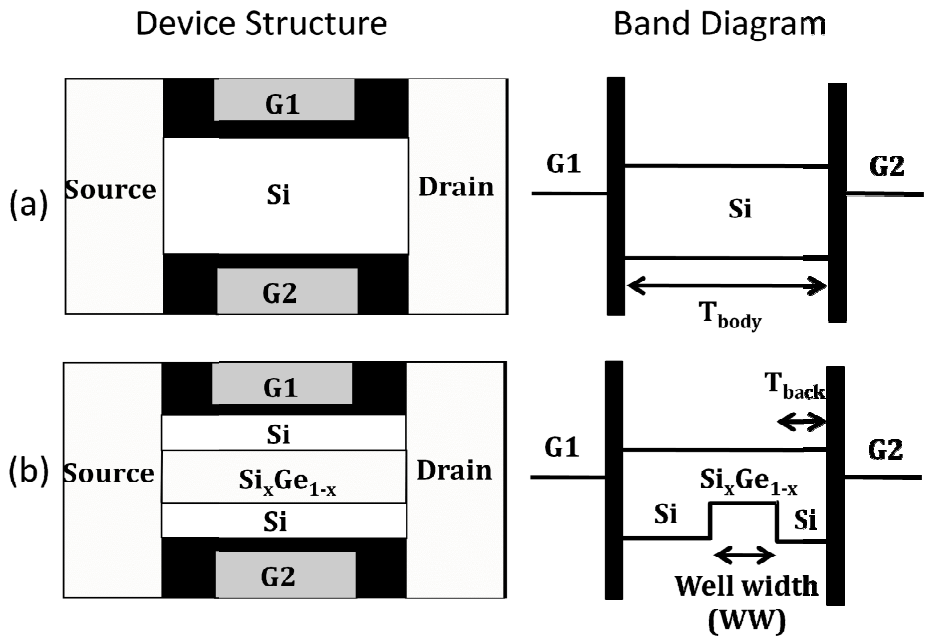


Fig. 4.1 The schematics of (a) the single transistor double gate DRAM **(b)** the novel single transistor double gate quantum well DRAM (1T QW DRAM) and the corresponding band diagram illustrating the quantum well in the body of the device

Table 4.1 Table shows the comparison of features of various memories

	1T QW DRAM	1T DRAM	DRAM	SRAM
Structure	1T	1T	1T/1C	6T
Cell Size	$4F^2$	$4F^2$	$8F^2$	$100F^2$
Storage	Quantum Well	Floating Body	Capacitor	Flip Flop
Speed	Fast	Fast	Fast	Ultra Fast
Read	Non destructive	Non destructive	Destructive	Non destructive
Scalability Issues	Lithography	Lithography Volume reduction	Capacitor	6T size
New Materials	Ge, SiGe, III-V	None	High K	None

We introduced a novel 1-Transistor (capacitorless) Quantum Well DRAM (1T-QW DRAM) [4.7]. First we use extensive simulations to find the optimum parameters for DG 1T DRAM schematic of which is shown in **Fig. 4.1 (a)**. Subsequently, we do a detailed analysis of the 1T-QW DRAM, one version of which using a SiGe or Ge quantum well in Si is shown in **Fig. 4.1 (b)**. This type of structure has been demonstrated experimentally for high performance p-MOS logic applications [4.8]. **Table 4.1** shows the comparison of features of various memories. Compared with 1T DRAM, this new memory has superior characteristics, such as the ability to produce higher V_t shifts, hence higher sensing margin, and also control over the distribution of the holes within the body. The distribution of stored holes can be successfully moved closer to the front gate, which is impossible in 1T-DRAM where holes are stored just close to the back gate interface. This property not only yields an increased V_t shift, but also a higher retention time. This novel memory (Heterostructure QW DRAM) can be realized as a horizontal Double-Gate structure [4.8], and possibly as a vertical Double-Gate structure [4.6], or as a FINFET device if a processing technique to build such vertical epitaxial layers could be realized.

4.2 Operation

4.2.1 Program

Operating voltages are given in **Table 4.2**. Program is done with the mechanism of impact ionization that occurs at the drain side. The body can be saturated with holes in a few nanoseconds [4.9,4.10,4.11].

Table 4.2 Table shows the operating voltages for the DRAM cell

	Program (Write "1")	Erase (Write "0")	Read	Hold
Gate 1 Voltage (V)	1	1.5	0.8	0
Drain Voltage (V)	1.2	-1.5	0.2	0
Gate 2 Voltage (V)	-1.5	-1.5	-1.5	-1.5

4.2.2 Hold

The retention time depends on how long the carriers (holes, in this case) generated by impact ionization can be kept in the floating body. Applying a negative bias to the back gate helps to keep the holes in the created floating body storage node.

The memory charge loss occurs primarily due to carrier recombination and leakage current due to tunneling. Because of the fact that tunneling induced leakage depends exponentially on electric field, and since DG structures exhibit low vertical electric field, the charge loss through leakage is naturally mitigated. Furthermore, by using a combination of a low gate bias, thicker gate oxide, and the thin body, the total electric field at the junctions can be reduced [4.2]. This, in turn results in reduced band to band tunneling (BTBT). Furthermore, a thin silicon body results in very low drain-

induced barrier lowering (DIBL). Thus, these advantages make a DG structure capable of much higher retention times.

4.2.3 Read

The reading is done by switching on the front MOS structure (shown as G1 in the device schematics) without disturbing the existing memory state and sensing the current and determining the state by the current difference. In the programmed/'1'state, cells have excess holes. Thus, due to the increased body potential, they have higher drain current due to the body effect. A sense amplifier can sense the current difference between the two states of '1' and '0' and hence the state of the device [4.12].

4.3 Results

Fig. 4.2 shows SENTAURUS^{TM} simulations explicitly demonstrating a DRAM operation for 1T QW DRAM: the drain voltage V_d, and drain current I_d corresponding to program (P), followed by read (R1), erase (E), and a second read (R2).

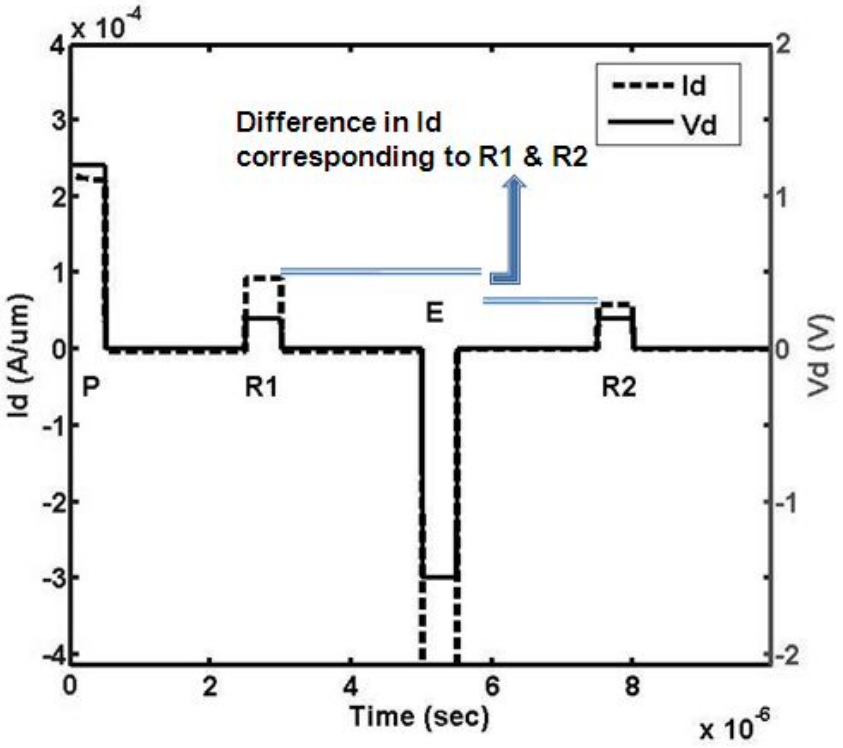


Fig. 4.2 Simulation results which show the drain current (I_d) and drain voltage (V_d) vs. time for 1T QW DRAM with $Si_{0.5}Ge_{0.5}$ Quantum Well. The cell is first programmed (denoted with "P" in the figure), then read (R1), and then erased (E) and then read again (R2).

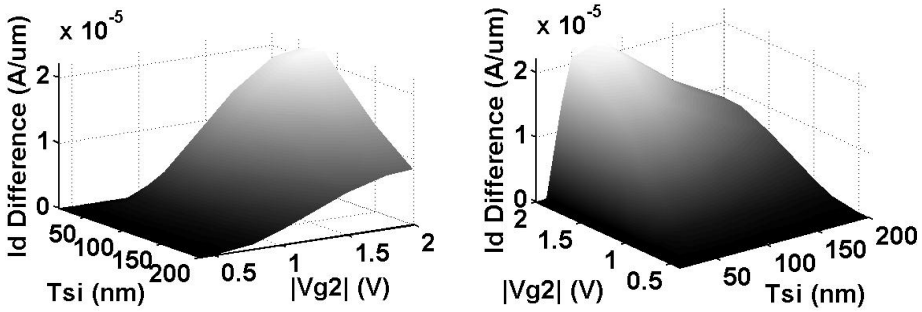


Fig. 4.3 Drain current difference between R1 and R2 as a function of Gate2 voltage (V_{g2}) and body thickness (T_{Si}) (For body doping =3e15 cm^{-3})

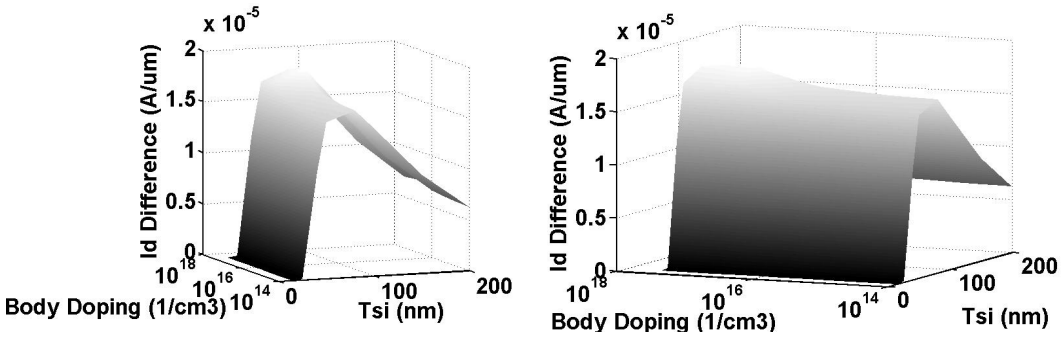


Fig. 4.4 Drain current difference between R1 and R2 as a function of body doping and body thickness (T_{Si}) (For V_{g2}=-1.5)

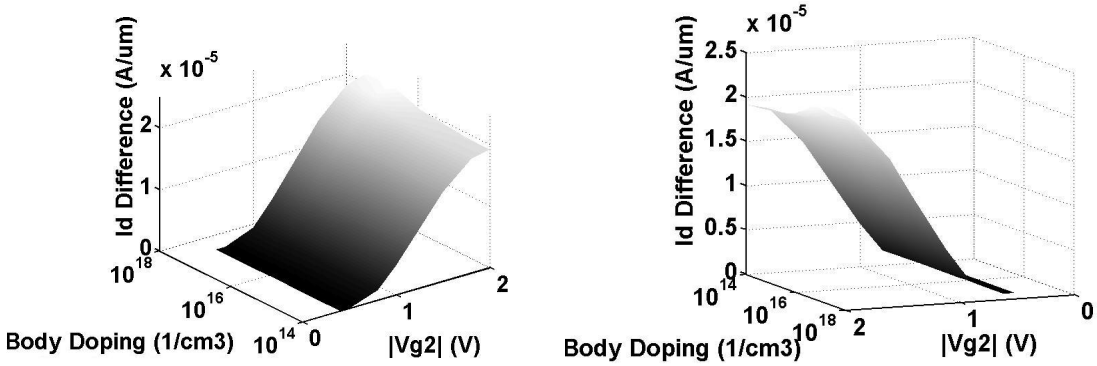


Fig. 4.5 Drain current difference between R1 and R2 as a function of body doping and Gate2 voltage (V_{g2}) (For T_{Si}=100nm)

Fig. 4.3 shows the drain current difference for DG 1T DRAM between states R1 and R2 as a function of back gate voltage (V_{g2}) and body thickness (T_{Si}). The same parameter is shown as a function of body doping and T_{Si} in **Fig. 4.4** and as a function of body doping and V_{g2} in **Fig. 4.5.** From these figures, it is seen that the body thickness around 80nm and back gate voltage around -1.5V yields the optimum results. Regarding the dependence on T_{Si}, there are two competing mechanisms: i) Body coefficient, ii)Storage capacity. The Body coefficient gets lower as T_{Si} increases. On the other hand, the lower the T_{Si} value, the less 'storage capacity' there is for the holes. Due to these competing mechanisms, there is an optimum point which is found to be around 80 nm. Regarding the back gate voltage (V_{g2}), in the case of low $|V_{g2}|$, holes cannot be kept in the body enough, so the V_t shift and hence, the I_d difference is reduced. For large $|V_{g2}|$ there will be some hole accumulation in the body at the back gate interface even for the erased state, hence the current difference between two states saturates after some point. Regarding the body doping, one should keep in mind that a large body doping results in threshold voltage fluctuations, and also trap-assisted tunneling and leads to tail bits with small retention times [4.4]. Here, the parameter of interest is the difference in I_d at R1 and R2, since it captures not only information related to the V_t shift, but also to the retention (since the reads are done after some time passes after programming and erase). This value is highly correlated with retention, whereas the V_t shift value only captures the effect of the presence of the holes on the V_t, but not how long that change could be maintained. Also, this difference in the current has to be in the range, which can be sensed by the sense amplifiers to determine the state of the memory. From the above results we see that

the performance is unsatisfactory for thin body devices, which might be a problem from the scalability perspective. We circumvent this problem by introducing a novel energy band engineered heterostructure 1-Transistor Quantum Well DRAM (1T-QW DRAM). **Fig. 4.1(b)** shows the new 1T-QW DRAM device and the corresponding band diagram. One example in which this concept can be realized is using Si and Ge (or SiGe). The concept is also applicable to the III-V material systems. In this work the QW width is taken as 5nm and Si/Ge/Si heterostructure is chosen. The gate length is 250nm, unless otherwise is stated.

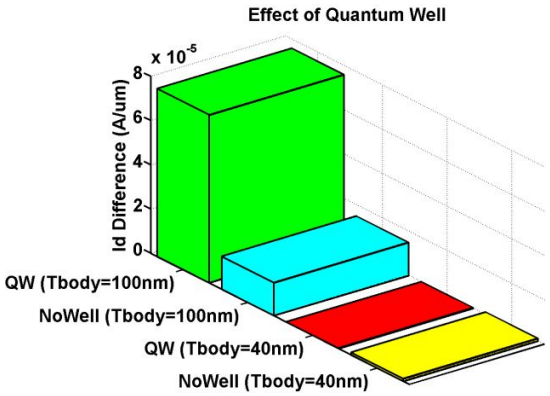


Fig. 4.6 Comparing the drain current difference between R1 and R2 of the devices with a quantum well in their body with the devices with no quantum well for T_{body}={100nm, 40nm}. The well width (WW) is 5nm and T_{back} is 5nm. The improvement in the 100nm body thickness device is 5x, where as in the device with smaller body thickness the performance gets worse.

Fig. 4.6 compares I_d difference between R1 and R2 for the case with and without QW for devices. The body thickness (T_{body}) of 100nm and 40nm is shown. For the

100nm T_{body}, the QW device exhibits a 5x improvement, whereas the QW device performance is worse for thinner T_{body}. This is due to the fact that the holes cannot be purged as effectively in the presence of the QW. The problem is solved by introducing the "PowerErase" in which the back gate voltage is switched to a positive value during erase, in order to push holes away effectively. The new operating voltages can be seen in **Table 4.3**. In **Fig. 4.7**, it is seen that with the introduction of PowerErase, the improvement due to QWs in the 100nm and 40nm T_{body} devices is 7x and 8x, respectively. Thus, PowerErase substantially improves the performance of the scaled QW device. Note that, in the case without a QW, PowerErase does not cause an improvement, since in that case holes are already purged far more easily.

Table 4.3 Table shows the operating voltages for the DRAM cell with PowerErase

	Program (Write "1")	Erase (Write "0")	Read	Hold
Gate 1 Voltage (V)	1	1.5	0.8	0
Drain Voltage (V)	1.2	-1.5	0.2	0
Gate 2 Voltage (V)	-1.5	1.5	-1.5	-1.5

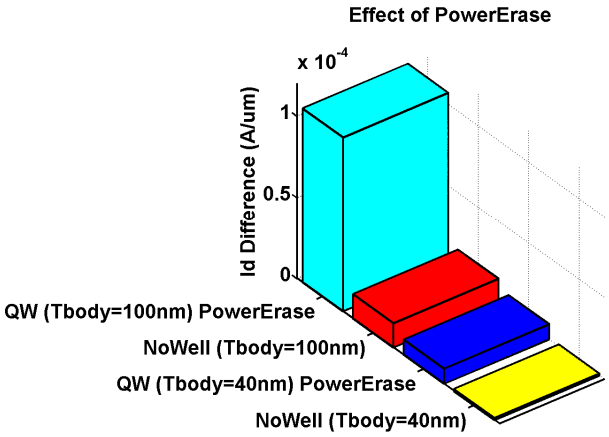
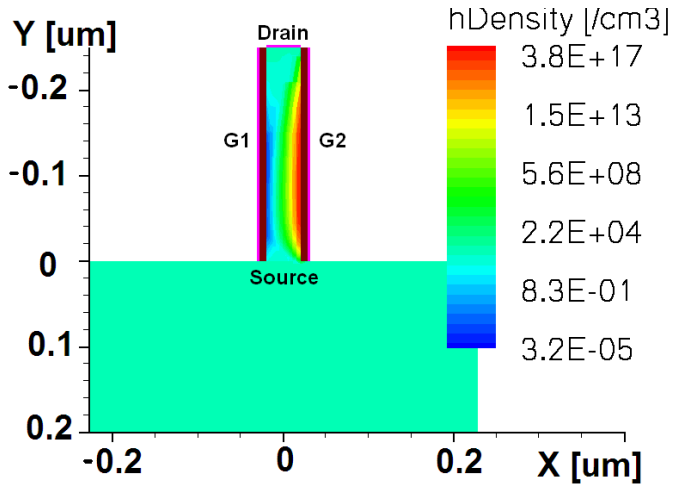


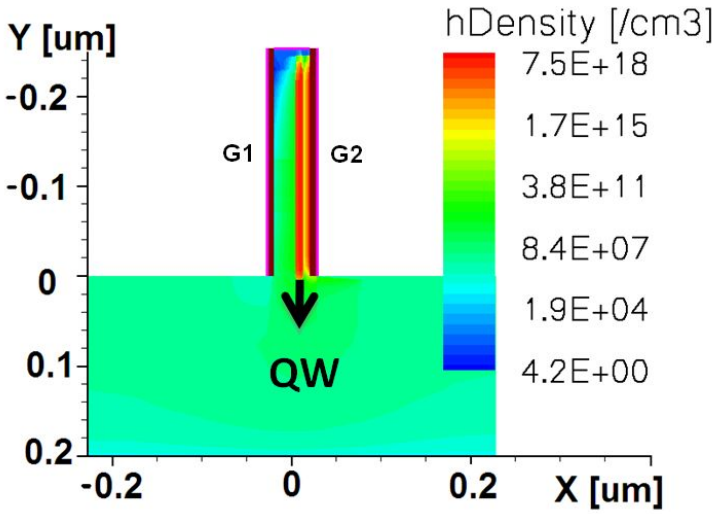
Fig. 4.7 Comparing the drain current difference between R1 and R2 of the devices with a quantum well in their body with the devices with no quantum well for T_{body}={100nm, 40nm} when PowerErase is used. The well width (WW) is 5nm and T_{back} is 5nm. The improvement in the 100nm body thickness device is 7x, and in the 40nm body thickness device it is 8x. The PowerErase substantially improves the performance in the scaled device.

Apart from introducing a "storage pocket" for holes in the body via QW, this device also allows the engineering of spatial distribution of the holes within the device body. **Fig. 4.8(a)** shows the hole density within a device without a QW during programming, whereas **Fig. 4.8(b)** shows the results when there is QW 10 nm away from the back gate interface (i.e., T_{back} =10 nm). The hole distribution is successfully shifted toward the front gate with the incorporation of the QW. **Fig. 4.8 (c)** shows that when the T_{back} is increased from 5nm to 10nm, the improvement in the 100nm body thickness device increases from 7x to 9x, and in the 40nm body thickness device it increases from

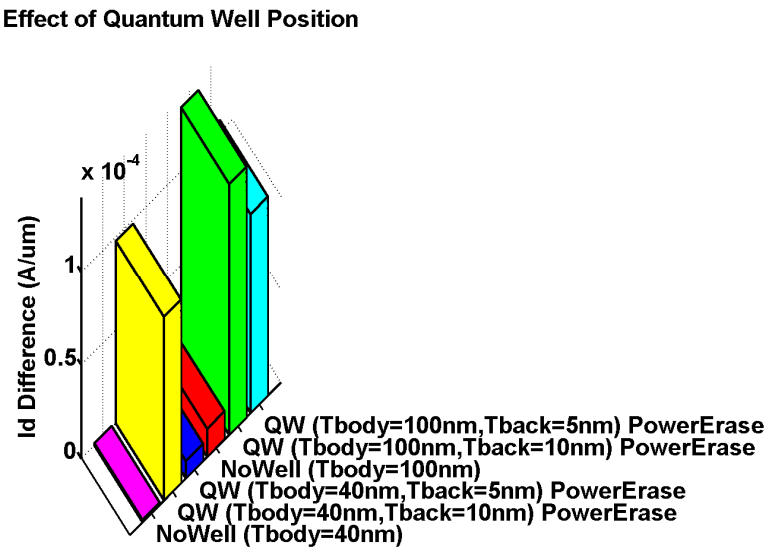
8x to 86x. The engineering of the spatial distribution of holes significantly improves the performance in the scaled device.



(a)



(b)



(c)

Fig. 4.8 (a) Hole density during program (T_{body} = 40 nm, No QW). (b) Hole density during program in the device with QW (T_{body} = 40 nm, T_{back} =10 nm). As it can be seen, by choosing the location of the QW, one can engineer the spatial hole distribution. (c) Comparing the drain current difference between R1 and R2 of the devices with a quantum well in their body with the devices with no quantum well for T_{body}={100nm, 40nm} and T_{back}={5nm, 10nm}. The well width (WW) is 5nm. When the QW is shifted 5nm towards the front gate the improvement in the 100nm body thickness device increases from 7x to 9x, and in the 40nm body thickness device it increases from 8x to 86x. The engineering of the spatial distribution of holes significantly improves the performance in the scaled device.

As it can be seen, the QW gives us a control over the spatial distribution of the stored holes, which is not possible in the other 1T DRAMs. With the engineering of the spatial distribution of the holes, it is possible to bring the stored holes closer to the front gate, thus, causing an increase in the amount of the V_t shift. Also, in numerous cases it is desirable to keep the stored holes away from the back oxide interface due to presence of traps and dangling bonds. This further helps in improving extrinsic retention. This property is especially advantageous for materials such as Ge and III-V material systems, where passivation of interface traps is problematic.

Fig. 4.9 shows the comparison of the effect of scaling on Quantum Well and No Quantum Well devices. As, it can be seen, the No Well devices have problems with scaling. Similar to the case of body thickness scaling, the reduction of the gate length reduces the volume of the devices to hold the excess holes. The QW devices are more scalable since the introduction of the extra ‘storage space’ by the QW helps to reduce the negative effects caused by reduced gate length to some extent.

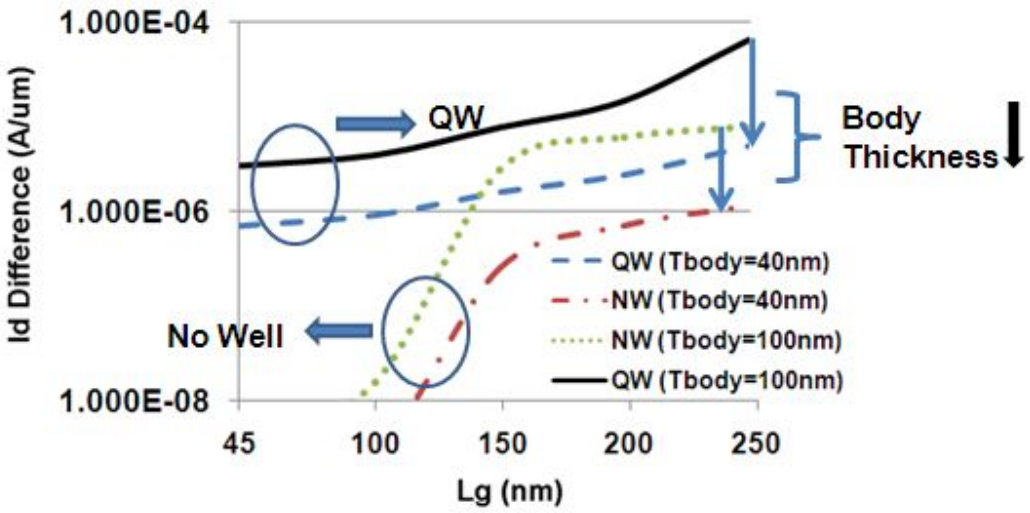


Fig. 4.9 Effect of scaling (T_{back}=5nm, WW=5nm)

The retention characteristics are also studied. **Fig. 4.10** shows the retention characteristics at 300K and 358K for the devices with QW under read, and **Fig. 4.11** under the hold (for 60nm body thickness devices). **Fig. 4.12** compares the retention characteristics for the devices with and without QW at 358K. It can be seen that the QW devices have better characteristics in terms of higher difference in I_d (ie. higher V_t shift) hence, higher sense margin (i.e. wider sense current window), and also higher retention time.

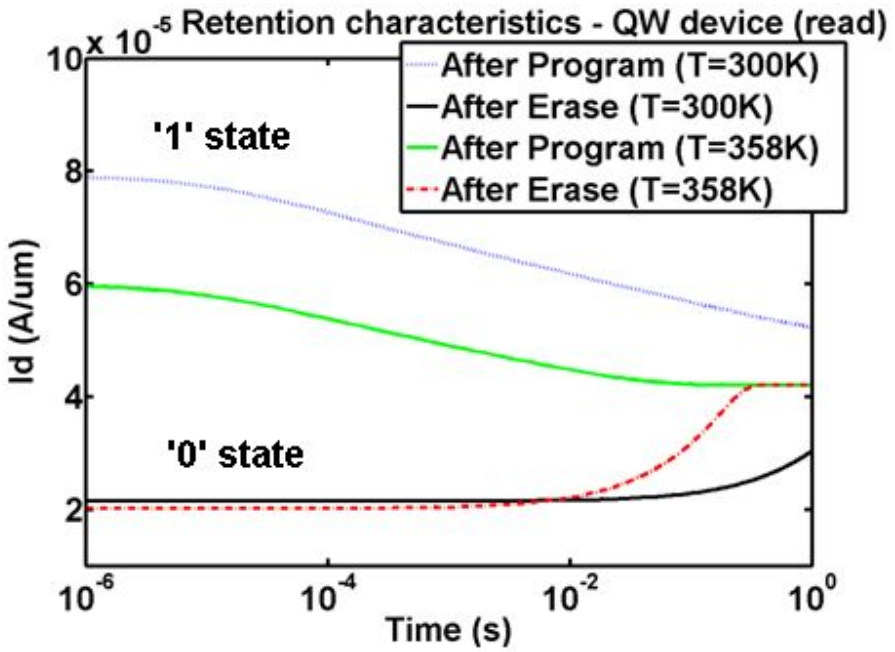


Fig. 4.10 Retention characteristics for devices with QW under read (T_{body}=60nm, WW=5nm)

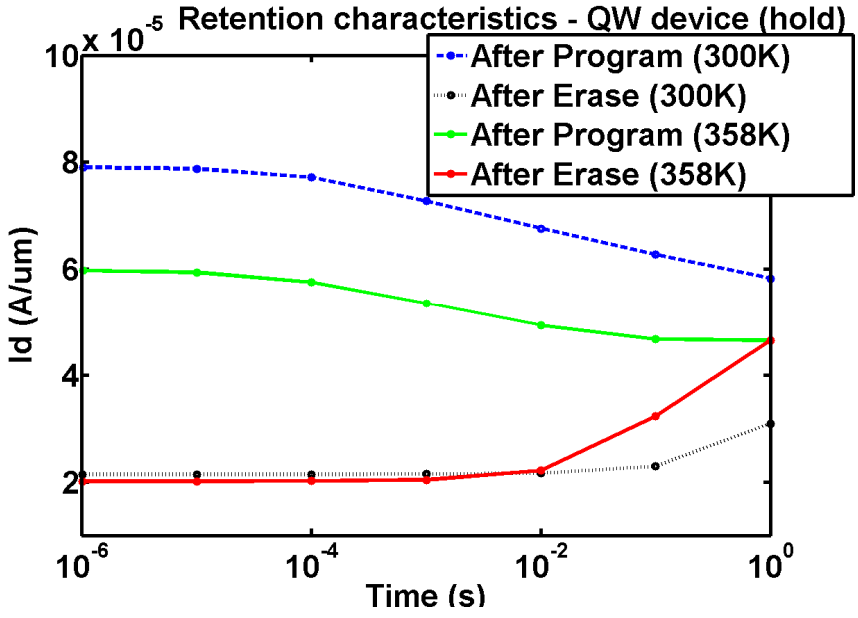


Fig. 4.11 Retention characteristics for devices with QW under hold (T_{body}=60nm, WW=5nm)

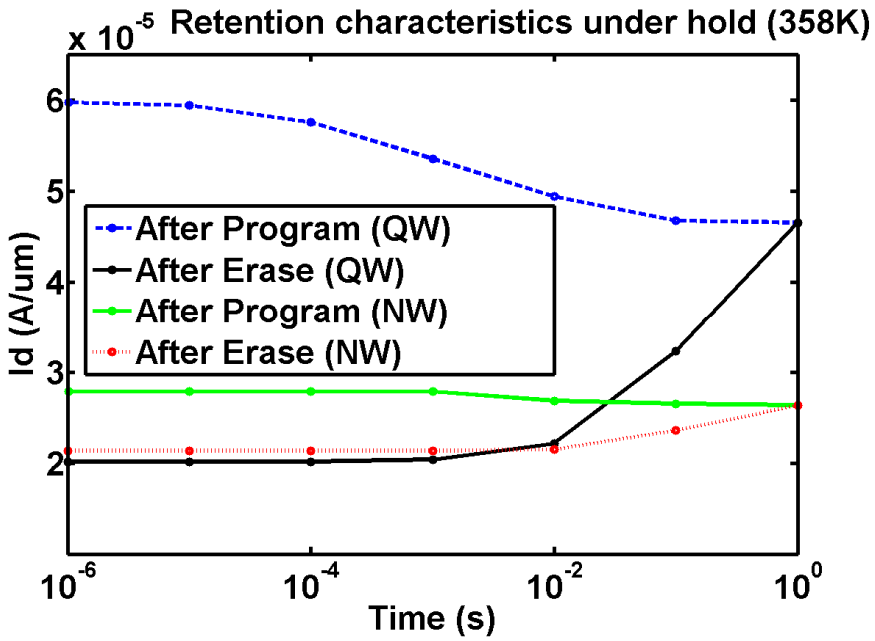


Fig. 4.12 Retention characteristics for devices with and without QW under hold at 358 K (T_{body}=60nm, WW=5nm)

The effect of the Germanium concentration is also investigated. The characteristics of the QW devices can be improved further by using SiGe instead of pure germanium, and by engineering the depth of the QW, the '0' state (i.e. erased state) behaviour is improved, hence the retention time. This is because of the fact that the '0' state (ie. erased state) retention degrades faster when the QW is too deep. Due to repulsion between holes and diffusion mechanism, holes tend to diffuse back into the QW well faster as the QW gets deeper at some point. Apart from our work [4.7], a similar band engineering approach has been also employed in [4.13] as $Si/Si_{1-x}Ge_x$ energy band

engineering but on bulk substrates for a single gate device, where similar relation between degradation and well depth was also observed. Furtheremore, from the power dissipation perspective, the erase operation becomes difficult when the well is too deep. **Fig. 4.13** reveals the retention characteristics when $Si_{(1-x)}Ge_{(x)}$ (x={1.0, 0.7, 0.5, 0.4, 0.3, 0.2}) is used to form the QW. From these results, it is seen that one can use $Si_{0.5}Ge_{0.5}$ instead of Germanium, and still have a similar sense margin (sensing current window) and even possibly a higher retention, with the added advantage of slower '0' state degradation and easier Erase operation. Also, better Si/SiGe interfaces can be fabricated compared to Si/Ge interfaces due to lesser mismatch in lattice, resulting in lower defect densities which also helps to improve the retention. Another point that needs to be noted is that at high Ge concentrations the trap density that could originate from epitaxial growth might be significant enough to effect state '1' (ie. programmed state) degradation, hence, this is another reason for not favouring high Ge concentrations from 'state '1' degradation point of view.

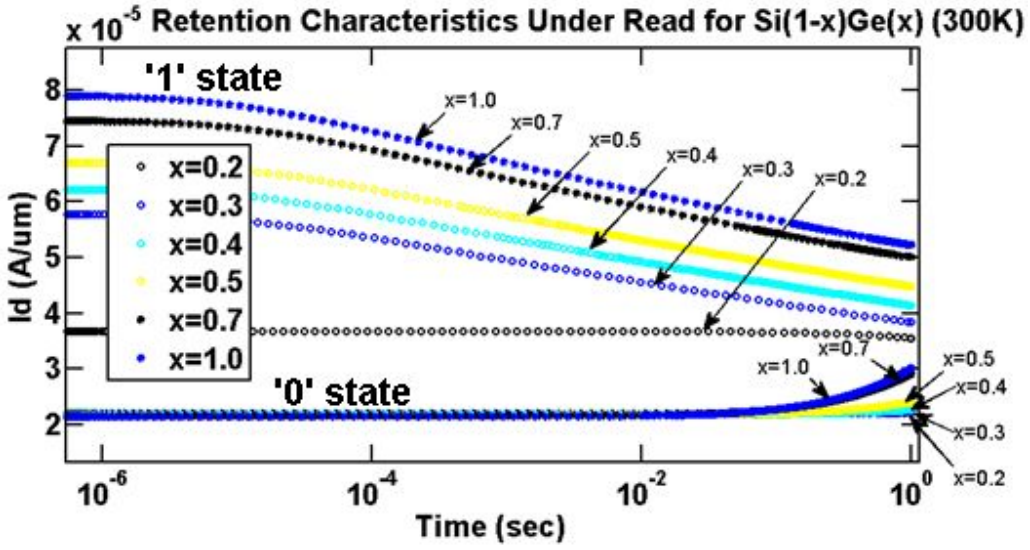


Fig. 4.13 Retention characteristics for devices with $Si_{(1-x)}Ge_{(x)}$ QW under read (T_{body}=60nm, WW=5nm)

Fig. 4.14 shows the hole density within a device with a $Si_{0.3}Ge_{0.7}$ QW during programming, whereas, **Fig. 4.15** shows the results within a device with a $Si_{0.5}Ge_{0.5}$ QW. As it can be seen the hole density is increased when the QW is deeper, hence causing a higher difference in I_d (ie. higher V_t shift) therefore, higher sense margin (ie. wider sense current window).

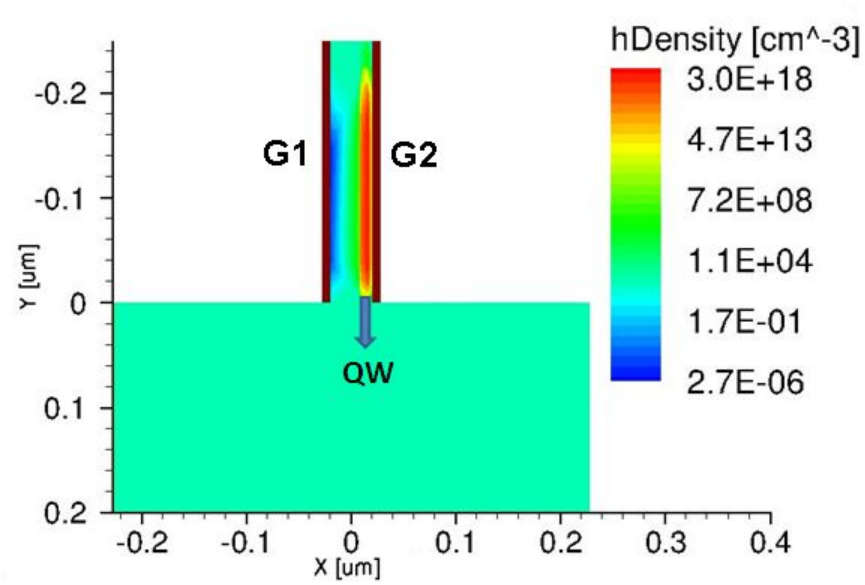


Fig. 4.14 Hole Density during Program for $Si_{(1-x)}Ge_{(x)}$ QW devices (x=0.7, T_{body}=40nm, WW=5nm)

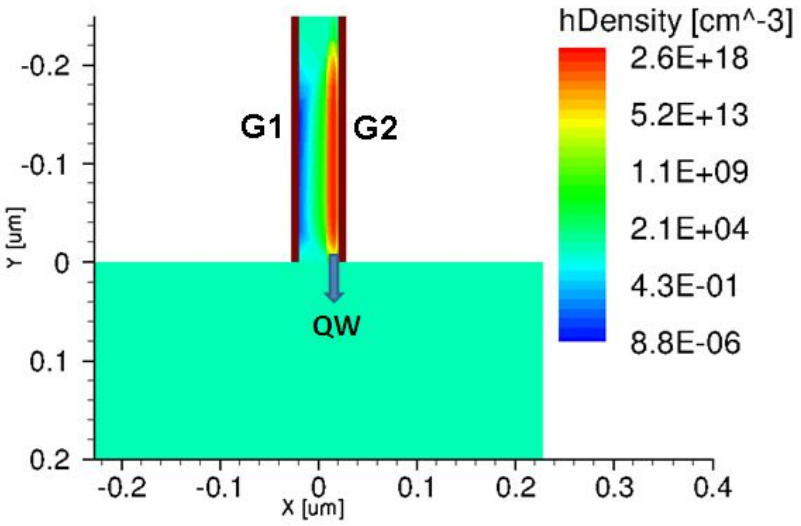


Fig. 4.15 Hole Density during Program for $Si_{(1-x)}Ge_{(x)}$ QW devices (x=0.5, T_{body}=40nm, WW=5nm)

Fig. 4.16 reveals that in SiGe QW devices, the regular erase does not come with as much penalty as in the case with pure germanium Quantum Well. In this case, since the well is less deep compared to the case with germanium, the regular Erase voltages would be able to be used instead of the PowerErase scheme, and this brings the added advantage of lesser power dissipation and also less complexity in array design.

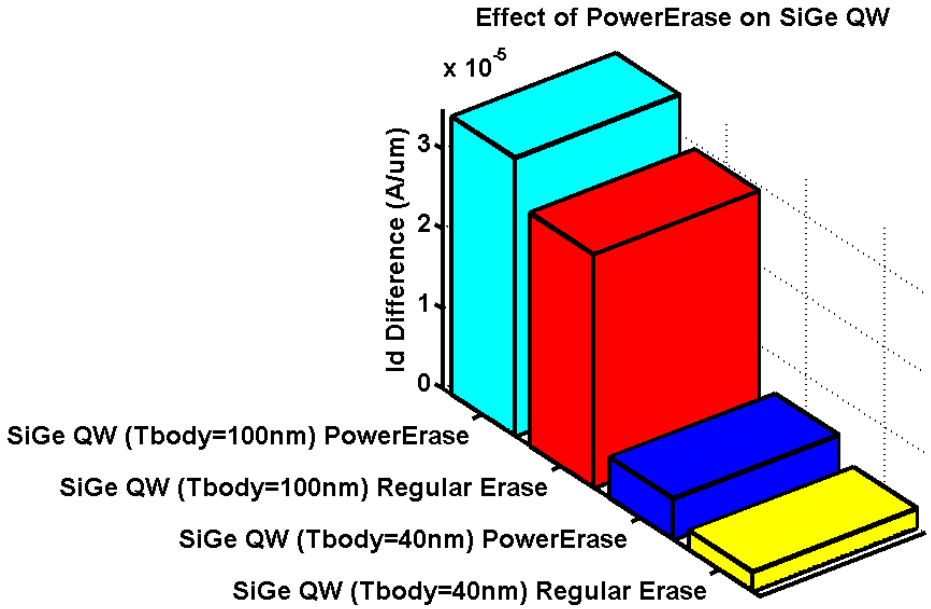


Fig. 4.16 The comparison of regular Erase and PowerErase in $Si_{(1-x)}Ge_{(x)}$ QW devices (x=0.5, T_{body}={40nm,100nm}, WW=5nm)

4.4 Conclusions

We have studied and characterized Double Gate Capacitorless DRAM, and in addition, we studied the novel single transistor double gate quantum well DRAM (1T QW DRAM) [4.7, 4.15]. This new DRAM has several advantages in terms of performance and scalability. One of them is introducing a "storage pocket" within the device. It also allows the possibility of engineering the spatial distribution of the holes in the body of the device, which is not possible in the other 1T-DRAMs. We studied the effects of the germanium content in the Quantum Well in terms of difference in I_d (ie. V_t shift), sensing current window, '0' state degradation and the erase mechanism. It is found that the use of SiGe instead of pure germanium for the Quantum Well has added advantages such as: Improvement in '0' state degradation, and hence in retention; the reduction in the need of PowerErase; the added ability to be able to form better interfaces with Si/SiGe heterostructure, compared to Si/Ge heterostructure; and ease in fabrication process. This new memory has the ability to have higher I_d difference values (i.e. wider sense margin) and retention values, and a better candidate for scaled new technology nodes. We successfully quantify these advantages.

4.5 References

[4.1] T. Tanaka, E. Yoshida, T. Miyashita, "Scalability study on a capacitorless 1T-DRAM: from single-gate PD-SOI to double-gate FinDRAM," Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International , pp. 919-922, 13-15 Dec. 2004

[4.2] C. Kuo, T.-J. King, C. Hu, "A capacitorless double-gate DRAM cell," Electron Device Letters, IEEE , vol.23, no.6, pp.345-347, Jun 2002

[4.3] E. Yoshida, T. Miyashita, T Tanaka, T., "A study of highly scalable DG-FinDRAM," Electron Device Letters, IEEE , vol.26, no.9, pp. 655-657, Sept. 2005

[4.4] C. Kuo, T.-J. King, C. Hu, "A capacitorless double-gate DRAM cell design for high density applications," Electron Devices Meeting, 2002. IEDM '02. Digest. International , pp. 843-846, 2002

[4.5] I. Ban, U. E. Avci, U. Shah, C. E. Barns, D. L. Kencke, P. Chang, "Floating Body Cell with Independently-Controlled Double Gates for High Density Memory," Electron Devices Meeting, 2006. IEDM '06. International , pp.1-4, 11-13 Dec. 2006

[4.6] M. G. Ertosun, H. Cho, P. Kapur, K. C. Saraswat, "A Nanoscale Vertical Double-Gate Single-Transistor Capacitorless DRAM," IEEE Electron Device Letters, vol.29, no.6, pp.615-617, June 2008

[4.7] M. G. Ertosun, P. Kapur, K. C. Saraswat, "A Highly Scalable Capacitorless Double Gate Quantum Well Single Transistor DRAM: 1T-QW DRAM," IEEE Electron Device Letters, vol.29, no.12, pp.1405-1407, Dec. 2008

[4.8] T. Krishnamohan, Z. Krivokapic, K. Uchida, Y. Nish and K. C. Saraswat, ``High Mobility, Ultra Thin (UT), Strained Ge MOSFETs On Bulk and SOI With Low Band To Band Tunneling (BTBT) Leakage : Experiments," IEEE Trans. Electron Dev. Vol. 53, No. 5, May 2006, pp. 990-999.

[4.9] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan, "A capacitor-less SOI 1T-DRAM concept," in Proc. Int. SOI Conf., 2001, pp. 153-154, 2001.

[4.10] T. Ohsawa, K. Fujita, T. Higashi, Y. Iwata, T. Kajiyama, Y. Asao, and K. Sunouchi, "Memory design using one-transistor gain cell on SOI," in Proc. ISSCC, pp. 152-153, 2002.

[4.11] J. Gautier, K. A. Jenkins, J.Y.-C. Sun, "Body charge related transient effects in floating body SOI NMOSFETs ," Electron Devices Meeting, 1995., International , vol., no., pp.623-626, 10-13 Dec 1995

[4.12] T. N. Blalock, R.C. Jaeger, "A high-speed clamped bit-line current-mode sense amplifier," Solid-State Circuits, IEEE Journal of , vol.26, no.4, pp.542-548, Apr 1991.

[4.13] Jin-Woo Han; Seong-Wan Ryu; Sungho Kim; Chung-Jin Kim; Jae-Hyuk Ahn; Sung-Jin Choi; Kyu Jin Choi; Byung Jin Cho; Jin Soo Kim; Kwang Hee Kim; Gi Sung Lee; Jae Sub Oh; Myong Ho Song; Yun Chang Park; Jeoung Woo Kim; Yang-Kyu Choi, "Energy band engineered unified-RAM (URAM) for multi-functioning 1T-DRAM and NVM," Electron Devices Meeting, 2008. IEDM 2008. IEEE International , pp.1-4, 15-17 Dec. 2008

[4.15] M. G. Ertosun; K. C. Saraswat; "Investigation of Capacitorless Double-Gate Single-Transistor DRAM: With and Without Quantum Well," IEEE Transactions on Electron Devices, vol.57, no.3, pp.608-613, March 2010

Chapter 5

Capacitorless Single-Transistor Charge Trap DRAM

5.1 Introduction

The use of capacitors is becoming increasingly problematic in conventional DRAMs from the scaling and fabrication points of view. Capacitorless single-transistor (1T) DRAMs, because of their smaller size, have been widely researched recently [5.1]–[5.5]. In these devices, excess holes are created in the body of transistors using impact ionization or gate-induced drain leakage mechanisms. The existence of holes in the floating-body reduces the threshold voltage (V_t), so a higher drain current (I_d) is obtained during a read operation compared to the case where there are no excess holes. A current-

sensing circuitry is employed to sense the I_d level, hence the state that is stored within the device. We propose and experimentally demonstrate a novel DRAM which uses the existence or absence of electrons in its body instead of holes that are traditionally used, and which also utilizes the charge-trap (CT) mechanism in its memory operation instead of floating-body effects. This is achieved by engineering the device to create an intentional electron-trapping zone (ETZ) within the body of the device. Furthermore, lower voltage and faster operation are achieved with this novel body-engineered charge trap (CT) device compared to previously explored gate-stack-engineered tunneling-based CT devices.

5.2 Operation Principle

Fig. 5.1 shows the operational principles of conventional floating-body single transistor (1T) capacitorless DRAM. The memory effect is caused by the excess holes in the floating body. **Fig. 5.1 (a)** shows the state "1" with excess holes, in which the V_t of the transistor is lowered due to holes in the floating body, whereas in state "0" (**Fig. 5.1 (b)**) there are no excess holes. During the read, a higher drain current (I_d) is obtained for state "1" compared to state "0" (**Fig. 5.1(c)**), and hence the memory state is sensed.

Fig. 5.2 shows the operational principles of the proposed novel charge trap (CT) memory. Here the main mechanism that is responsible for the memory effect is the charge trapping. In state "0" (**Fig. 5.2(a)**) there are electrons trapped within the device, whereas in state "1" (**Fig. 5.2(b)**) there is not. The existence of electrons increases the V_t of the transistor corresponding to state "0," so lower I_d is observed during a read (**Fig. 5.2(c)**). A current-sensing circuitry determines the state that is stored in the device. **Table 5.1** shows the characteristics of various memory types.

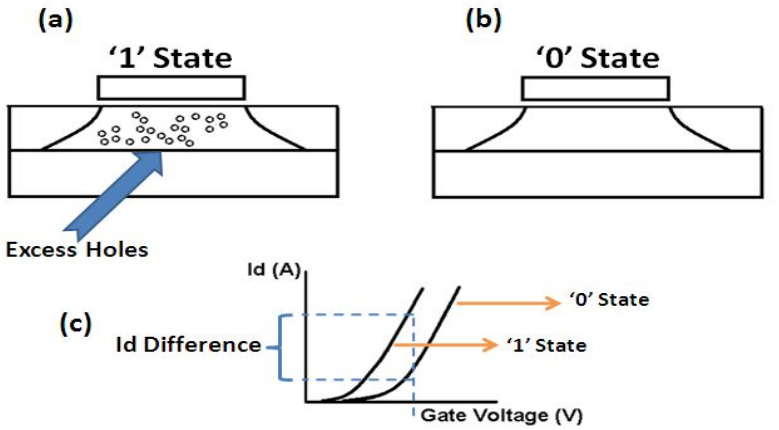


Fig. 5.1 Conventional (floating-body) 1T DRAM device schematics. In (a) "1" state, there are excess holes in the body, whereas in (b) "0" state, there is not. The device state is determined by sensing (c) the drain current. The excess holes in the body cause an decrease in threshold voltage corresponding to state "1"

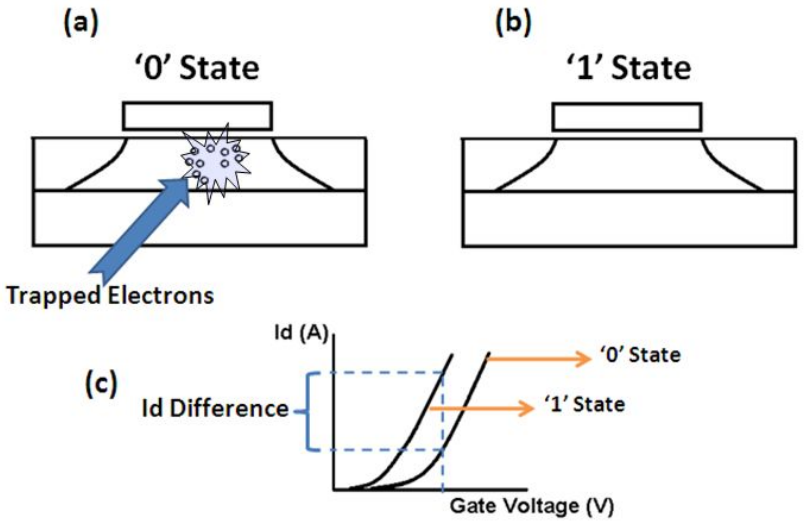


Fig. 5.2 1T CT DRAM device schematics. In (a) "0" state, there are electrons trapped in the body, whereas in (b) "1" state, there is not. The device state is determined by sensing (c) the drain current. The trapped electrons in the ETZ cause an increase in threshold voltage corresponding to state "0."

Table 5.1 Comparison of characteristics of various memory types

	1T CT DRAM	1T DRAM	DRAM	SRAM
Structure	1T	1T	1T/1C	6T
Cell Size	$4F^2$	$4F^2$	$8F^2$	$100F^2$
Storage	Charge Traps	Floating Body	Capacitor	Flip Flop
Speed	Fast	Fast	Fast	Ultra Fast
Read	Non destructive	Non destructive	Destructive	Non destructive
Scalability Issues	Lithography	Lithography	Capacitor	6T size
New Materials	Voids, Nano Crystals, Ge…etc	None	High K	None

5.3 Simulation Results

Fig. 5.3 shows simulation results for the capacitorless single-transistor charge-trap DRAM. **Fig. 5.3 (a)** depicts the device that is simulated along with the Electron Trapping Zone (ETZ). **Fig 5.3 (d)** shows the DRAM operation: The device is first programmed into a "0 state" (P) and then read after program (R1) is performed, then the device is erased (i.e. set into a "1 state"), and then read after erase (R2) is performed. As it can be seen, the trapped electrons increase the V_t in the "0 state" and a lower drain current is obtained during R1, compared to R2. The difference in the drain current corresponding to these two reads is $2x10^{-5}$ μA/μm in this case. **Fig 5.3 (b)** and **Fig 5.3 (c)** shows the trap occupancy in the ETZ during R1 and R2, respectively, confirming the charge trapping mechanism.

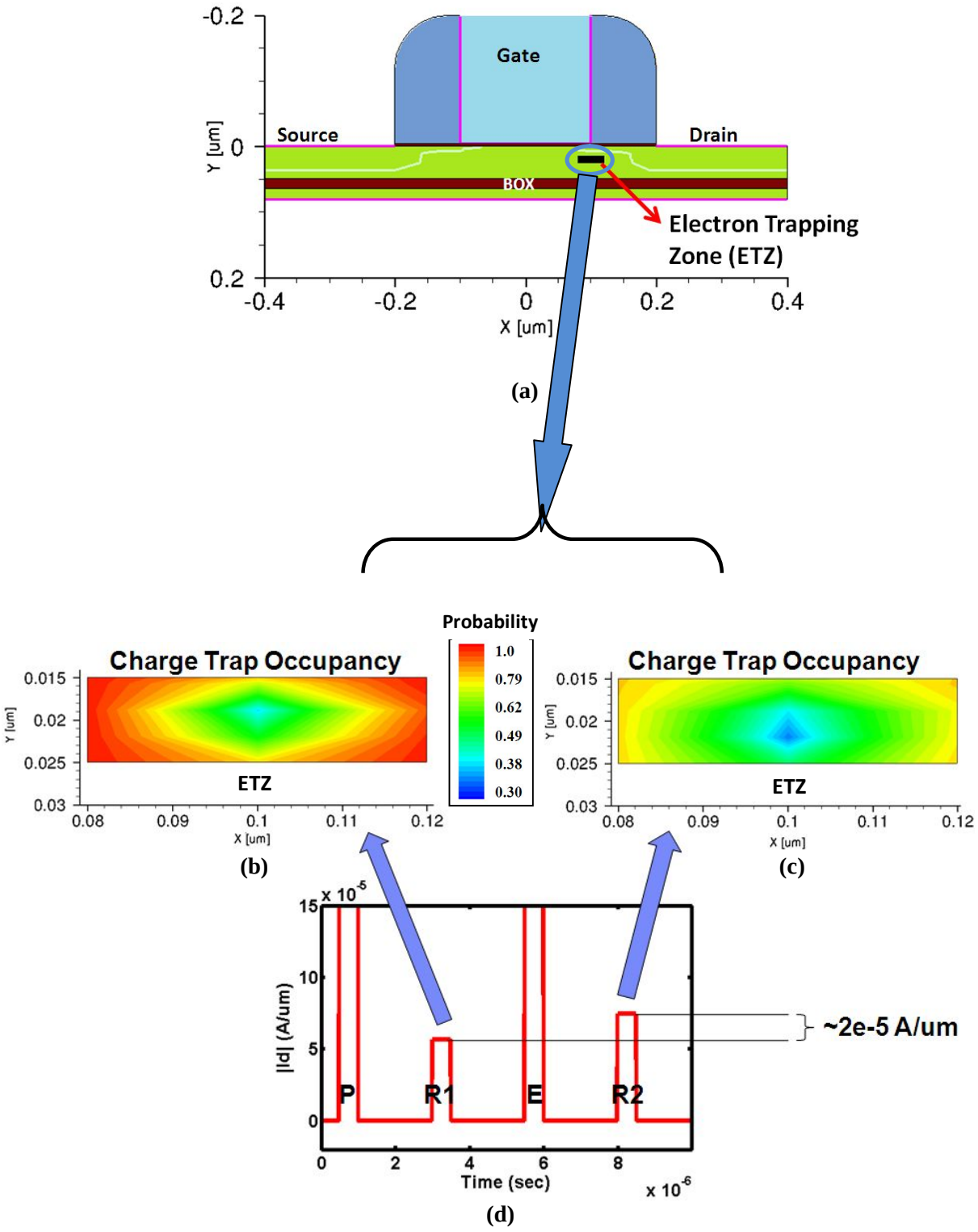

Gate
Source
Drain
BOX
Electron Trapping Zone (ETZ)
Y [um]
X [um]
(a)
Probability
1.0
0.79
0.62
0.49
0.38
0.30
Charge Trap Occupancy
ETZ
(b)
(c)
|Id| (A/um)
x 10^-5
x 10^-6
Time (sec)
P
R1
E
R2
~2e-5 A/um
(d)

Fig 5.3 Simulation results. (a) The simulated structure (L = 200 nm and T_{Si} = 50 nm). (b) Trap occupancy during read after program, (c) Trap occupancy during read after erase, (d) Drain current as a function of time during program (P), read after program (R1), erase (E), read after erase (R2).

5.4 Experimental Results

Fig. 5.4 shows a scanning electron microscope (SEM) image of the 1T CT DRAM. In this proof-of-concept demonstration the ETZ is created within the body of the device by employing a void structure. **Fig. 5.5** shows a scanning electron microscope (SEM) image of a device without the ETZ (i.e. a No-Trap (NT) Device) that is fabricated as a control device for the experiments.

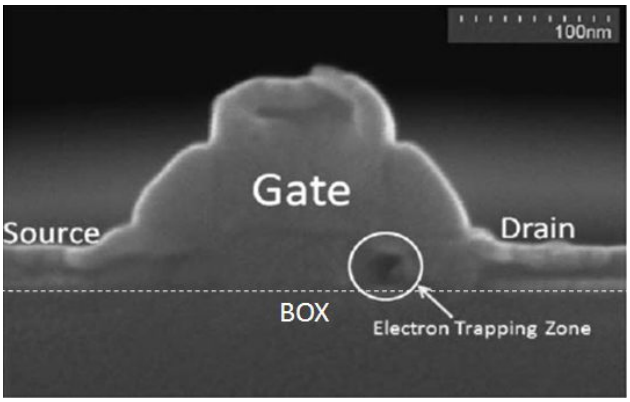


Fig. 5.4 SEM image of a CT DRAM that has an ETZ (L/W = 125 nm/10 μm and T_{Si} = 35 nm).

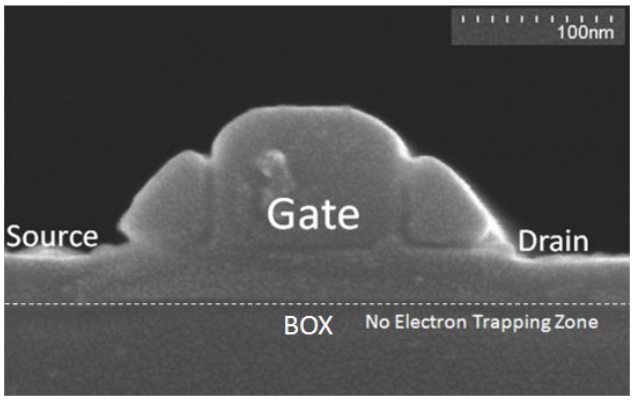


Fig. 5.5 SEM image of a no-trap (NT) device which does not have electron trapping zone (ETZ) (L/W=125nm/10µm, T_{Si}=35nm)

Fig. 5.6 shows the capacitance modeling of devices with an intentional ETZ (i.e., with a charge trap (CT)) and devices without an ETZ (no charge trap (NT)). The CT devices have extra capacitive element due to electron trapping zone (C_{ETZ}). The other capacitances are: C_{ox}: oxide capacitance, C_h: accumulation hole capacitance, C_e: inversion electron capacitance, C_b: space-charge region capacitance, C_{it}: interface trap capacitance. **Fig. 5.7** depicts a CV measurement of an NT device, whereas **Fig. 5.8** shows the same for a CT device. As can be seen, there is a difference in the inversion regime of a CV curve of a CT device, due to to the additonal capacitance associated with ETZ (C_{ETZ}). **Fig. 5.9** shows the equivalent capacitances in the inversion regime for CT and NT devices and explains the observed difference in the inversion regime of CV curve of a CT device, which indicates a phenomenon related to minority charges (electrons, in this case). **Fig.**

5.10 reveals the hysterisis in the CV curve of a CT device that reflects the memory behaviour.

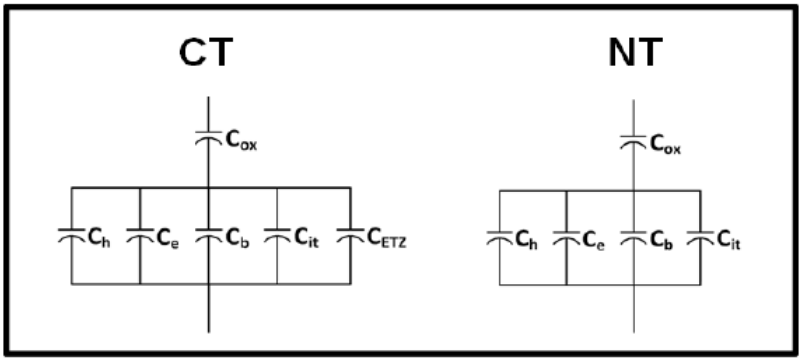


Fig. 5.6 The capacitor model for CT and NT devices. The CT devices have extra capacitive element due to electron trapping zone (C_{ETZ}). The other capacitances are: C_{ox}: oxide capacitance, C_h: accumulation hole capacitance, C_e: inversion electron capacitance, C_b: space-charge region capacitance, C_{it}: interface trap capacitance.

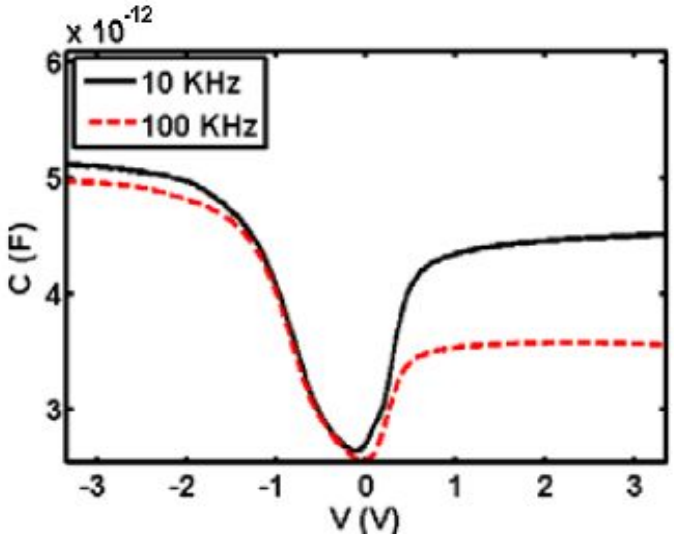


Fig. 5.7 CV measurement of an NT device.

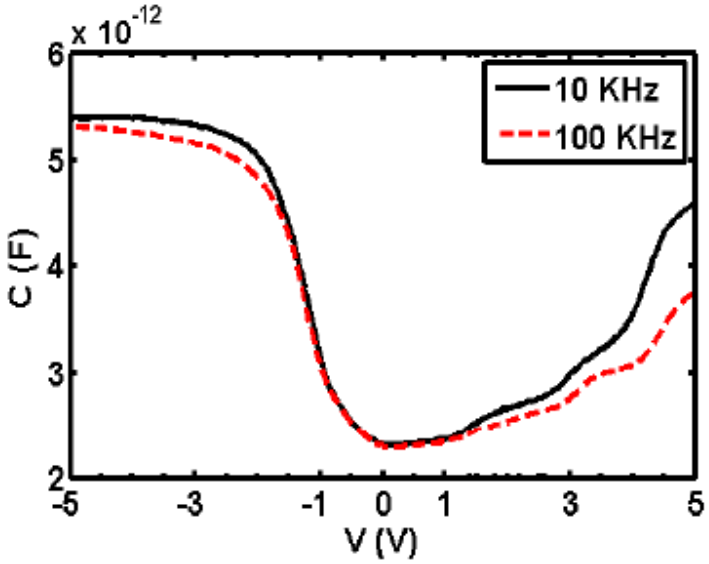


Fig. 5.8 CV measurement of a CT device. As it can be seen, the existence of C_{ETZ} lowers the capacitance in inversion compared to inversion capacitance of an NT device.

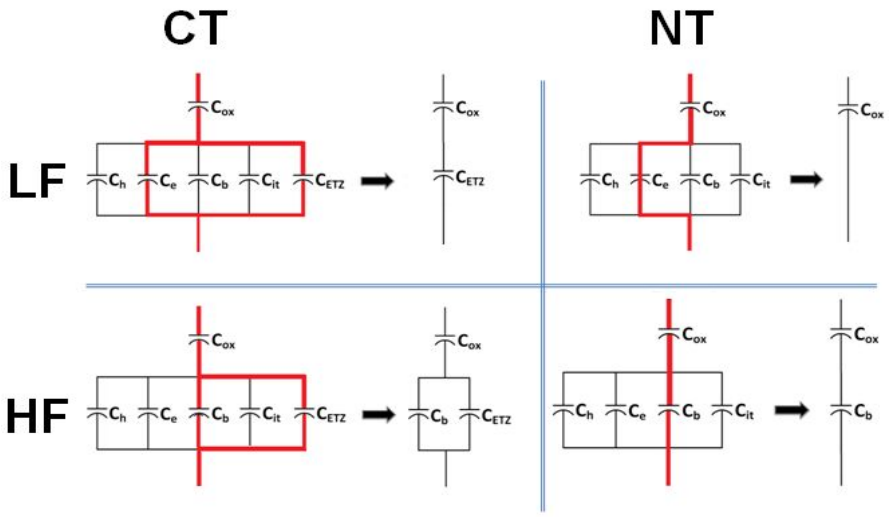


Fig. 5.9 Capacitances during inversion (both high and low frequency) of NT and CT devices. At low frequencies (LF), in NT devices, the low-frequency equivalent circuit becomes the oxide capacitance again since inversion charge is high and they can follow ac signal. However, in CT devices, there is an additional capacitance due to electron trap zone and it is in series with C_{ox}. In high frequency (HF), in NT devices when the inversion charge is unable to follow the ac voltage, the equivalent circuit is C_{ox} in series with C_b , where $C_b = K_s \varepsilon_o / W_{inv}$ with W_{inv} the inversion space-charge region width [5.17]. In CT devices, there is also C_{ETZ} in parallel with C_b.

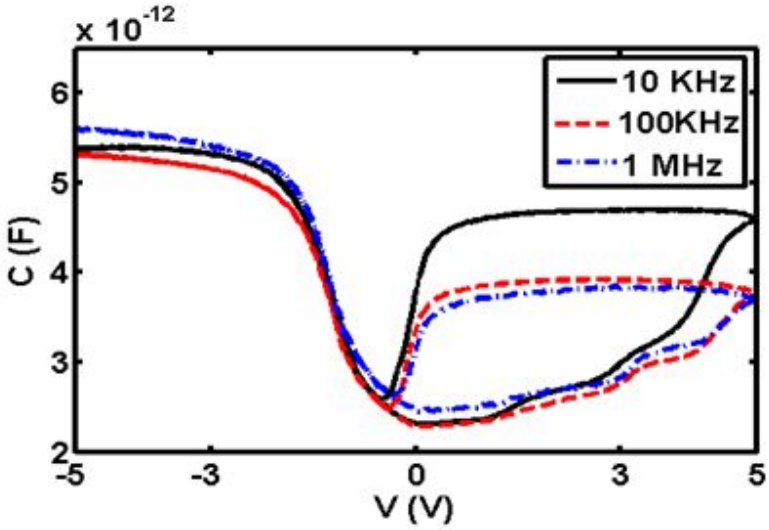


Fig. 5.10 Demonstration of hysteresis in a CT device: First device is swept from accumulation to inversion, and then from inversion to accumulation. During the first sweep, the traps get filled with electrons, and enough inversion charge is built up, hence C_e and C_{ETZ} no longer dominates the inversion capacitance and its equivalent circuit resembles the one of a NT device.

Fig. 5.11 and **5.12** show drain current vs. drain voltage (I_d-V_d) plots of NT and CT devices, respectively. In fully depleted (FD) devices, an electrically induced floating body appears within the depleted body when a negative back gate voltage is applied [5.5]. The NT devices show a kink effect with a reverse back bias, but the CT devices do not show such a prominent kink effect, since the electrons trapped in the ETZ do not let holes accumulate enough (as shown in **Fig. 5.13**) to lower the V_t as recombination at the traps is effective to suppress the kink effect [5.16]. It should be noted that the difference in the current levels in NT and CT devices are due to the required processing step differences in order to create the ETZ, and the resulting difference in transistor parameters.

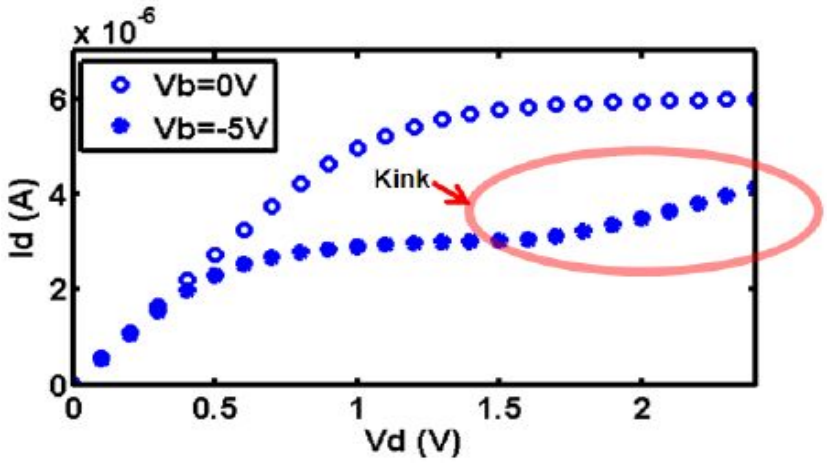


Fig. 5.11 Drain current vs. drain voltage of an NT device with back voltage (V_b) as a parameter. With zero back bias, there is no kink effect and a kink shows up with a reverse back bias, as expected.

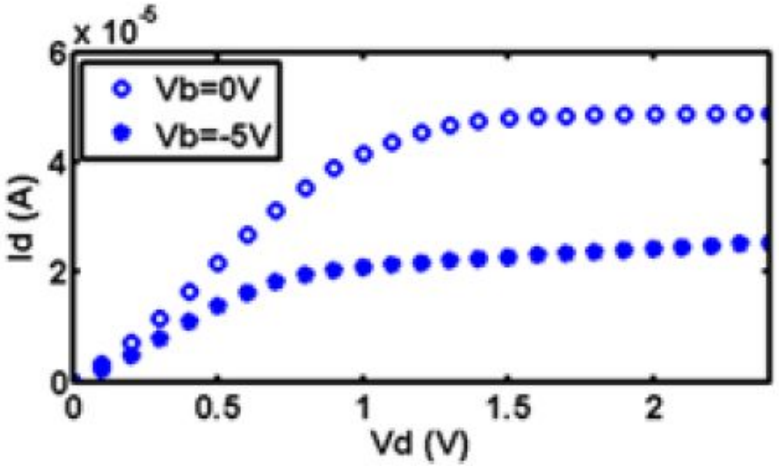


Fig. 5.12 Drain current vs. drain voltage of a CT device with back voltage (V_b) as a parameter. In this case, trap-assisted recombination suppresses kink effect.

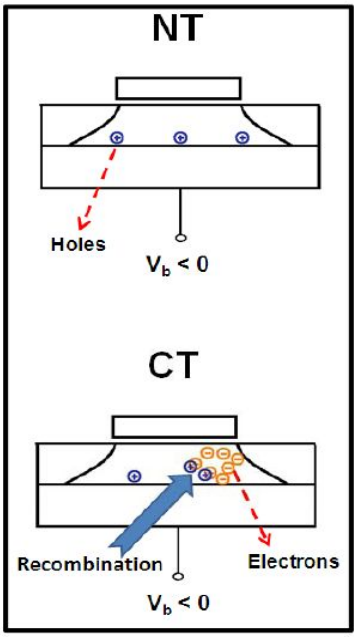


Fig. 5.13 With a reverse back bias holes get accumulated in an NT device and hence kink effect is observed; but in a CT device traps provide the effective recombination for the holes trying to accumulate, and kink effect is suppressed.

Operating voltages of 1T CT DRAM are shown in **Fig. 5.14**. Pulses with 50-ns width are applied to the drain to set the device in “0” or “1” state. Setting a “0” state makes use of the excess electrons created by impact ionization occurring at the drain side, and some of the electrons being trapped in the ETZ. A “1” state is set as the electric field de-traps electrons and drift out of the ETZ toward the source. Also, holes are attracted toward the ETZ to recombine with the electrons. Alternatively, by biasing both drain and gate negatively, electron–hole pairs can be generated by source-side impact ionization, and due to the negative drain bias, holes will be attracted toward the ETZ, which is located at the drain side, and recombine with the trapped electrons. In this work, in the body of an n-channel SOI device, a void structure is intentionally formed by silicon consumption by the salicidation mechanism, which creates the ETZ due to its interface states and dangling bonds. The formation of voids is controllable, scalable, and manufacturable, as silicon-on-nothing (SON) has been proposed as one of the candidates for sub-50 nm MOSFETs [5.6]–[5.8]. Researchers have developed several methods to fabricate voids for SON MOSFETs [5.8]–[5.11].

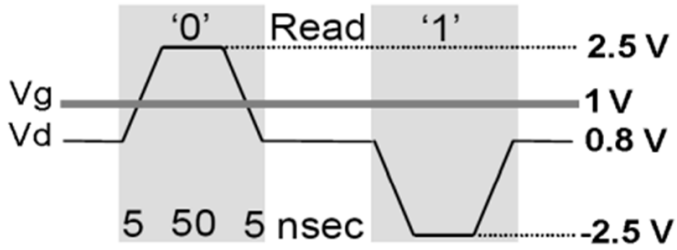


Fig. 5.14 Operating voltages. Pulses with 50-ns width are used to set "1" and "0" states.

Fig. 5.15 shows the DRAM operation experimentally. First, a "0" state is set, and then, after 30 μs, a "1" state is set. The I_d difference between the two states is 53 μA. In state "0," the existence of electrons increases the threshold voltage of the transistor, so a lower drain current is observed, whereas in state "1," the absence of electrons causes a higher drain current. **Fig. 5.16** shows I_d evolution under reading after state '1' and state '0' are set. While the current retention time of hundreds of microseconds of this first proof-of-concept demonstration appears to be insufficient for current DRAM applications, optimizing the device further enhances the retention time by engineering the distribution, concentration, and properties of the traps.

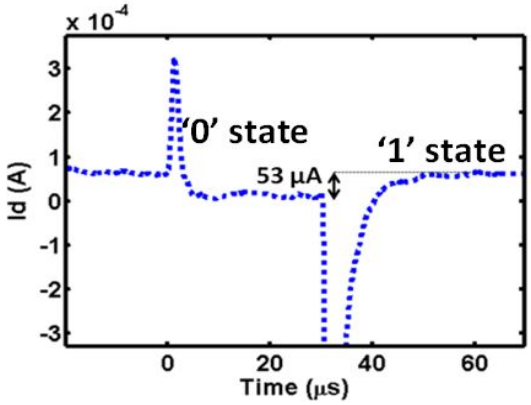


Fig. 5.15 1T CT DRAM device is first programmed to a '0' state and then set to a '1' state. The current difference between these two states is 53 µA.

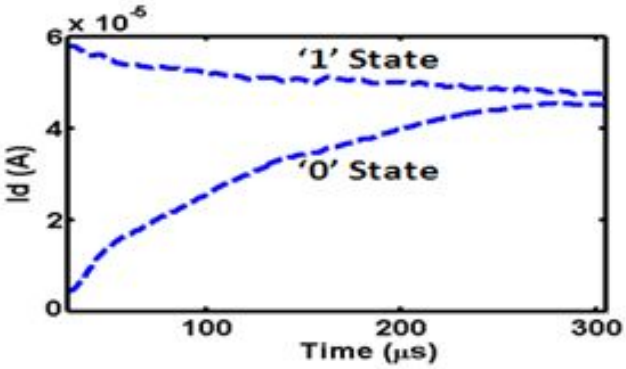


Fig. 5.16 Drain current evolution during read after setting '1' and '0' states.

The ETZ could also be realized in several ways such as embedding nanocrystals in the body of the device. Some studies have used nanocrystals in the gate stack to create charge trapping sites for nonvolatile memories, and some have also suggested them for 1T DRAMs [5.12], [5.13]. However, because these devices rely on tunneling, they are slow and require higher voltages. They also have reliability issues due to hot-carrier injection. Programming using either hot-electron injection or Fowler–Nordheim tunneling would damage the tunnel oxide of the cell. Direct tunneling causes little or no damage to the tunnel oxide, but to achieve the speed (tens of nanoseconds) needed for a DRAM application, the tunnel-oxide thickness must be on the order of 10–12 Å [5.12], and high program/erase voltages will still be needed (around 8–10 V) [5.14], [5.15]. All these issues make these gate-stack-engineered tunneling-based devices less attractive as DRAMs. Lower voltage and faster operation is achieved with our novel 1T CT DRAM device compared to gate-stack-engineered tunneling-based devices.

These charge-trap based DRAMs are more scalable (due to the lack of an external capacitor) than traditional 1T/1C DRAMs and occupy half the area. Also, they are more scalable compared to hole-based (i.e., floating-body) 1T DRAMs since, particularly as the volume of the device decreases, the amount of holes in the body, which are free to diffuse and leak, has difficulty to create enough memory-sensing window compared to electrons in the trapping zone, which are less prone to diffusion. Furthermore, in hole-based 1T DRAMs, there is a limit to how thin the body can be (and, hence, how much it

can be scaled) since it is not desirable to have the holes so close to the channel to mitigate read disturb and to have reasonable retention by decreasing diffusion and leakage to the source/drain. Also, this new memory does not need a reverse back bias for its operation compared to ultimately scaled fully depleted (FD) hole-based (floating-body) 1T DRAMs. Also, as it is shown in Chapter 4, the 1T QW DRAM gives us the opportunity to control the spatial carrier distribution, but only in one dimension. The carriers (i.e. holes in that case) could be moved either closer or further to the front gate in a manner that is parallel to the front gate. On the other hand, in the 1T CT DRAM the spatial distribution of the trapped carriers could be engineered in two dimensions (2D) by engineering the location of the traps within the body. This gives more flexibility in terms of design and engineering.

5.5 Conclusions

We have proposed a novel 1T DRAM [5.18] that relies on the existence and absence of electrons within its body and uses a charge-trapping mechanism in its memory operation, unlike conventional 1T DRAMs that rely on holes and employ floating-body effects in its memory operation. An ETZ has been created within the body of an FD SOI device. We use the traps formed along void structures to create the ETZ, but a 1T CT

DRAM could be realized by forming the necessary ETZ with other methods such as embedding nanocrystals or other appropriate materials.

5.6 References

[5.1] T. Tanaka, E. Yoshida, and T. Miyashita, “Scalability study on a capacitorless 1T-DRAM: From single-gate PD-SOI to double-gate FinDRAM,” in *IEDM Tech. Dig.*, Dec. 13–15, 2004, pp. 919–922.

[5.2] C. Kuo, T.-J. King, and C. Hu, “A capacitorless double-gate DRAM cell,” *IEEE Electron Device Lett.*, vol. 23, no. 6, pp. 345–347, Jun. 2002.

[5.3] M. G. Ertosun, H. Cho, P. Kapur, and K. C. Saraswat, “A nanoscale vertical double-gate single-transistor capacitorless DRAM,” *IEEE Electron Device Lett.*, vol. 29, no. 6, pp. 615–617, Jun. 2008.

[5.4] M. G. Ertosun and K. C. Saraswat, “Investigation of capacitorless double gate single-transistor DRAM: With and without quantum well,” *IEEE Trans. Electron Devices*, vol. 57, no. 3, pp. 608–613, Mar. 2010.

[5.5] C. Kuo, T.-J. King, and C. Hu, “A capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications,” *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2408–2416, Dec. 2003.

[5.6] W. Bu, R. Huang, M. Li, Y. Tian, and Y. Wang, “A novel technique of silicon-on-nothing MOSFETs fabrication by hydrogen and helium co-implantation,” in *Proc. 7th Int. Conf. Solid-State Integr. Circuits Technol.*, Oct. 18–21, 2004, vol. 1, pp. 269–272.

[5.7] R. Koh, “Buried layer engineering to reduce the drain-induced barrier lowering of sub-0.05 μm SOI-MOSFET,” *Jpn. J. Appl. Phys.*, vol. 38, no. 4B, pp. 2294–2299, Apr. 1999.

[5.8] M. Jurczak, T. Skotnicki, M. Paoli, B. Tormen, J. Martins, J. L. Regolini, D. Dutartre, P. Ribot, D. Lenoble, R. Pantel, and S. Monfray, “Siliconon-nothing (SON)—An innovative process for advanced CMOS,” *IEEE Trans. Electron Devices*, vol. 47, no. 11, pp. 2179–2187, Nov. 2000.

[5.9] S. Monfray, T. Skotnicki, Y. Morand, S. Descombes, M. Paoli, P. Ribot, A. Talbot, D. Dutartre, F. Leverd, Y. Lefriec, R. Pantel, M. Haond, D. Renaud, M.-E. Nier, C. Vizioz, and D. Louis, “First 80 nm SON (silicon-on-nothing) MOSFETs with perfect morphology and high electrical performance,” in *IEDM Tech. Dig.*, 2001, pp. 29.7.1–29.7.4.

[5.10] S. Monfray, T. Skotnicki, B. Tavel, Y. Morand, S. Descombes, A. Talbot, D. Dutartre, C. Jenny, P. Mazoyer, R. Palla, F. Leverd, Y. Le Friec, R. Pantel, M. Haond, C. Charbuillet, C. Vizioz, D. Louis, and N. Buffet, “SON (silicon-on-nothing) P-MOSFETs with totally silicided (CoSi2) polysilicon on 5 nm-thick Si-films: The simplest way to integration of metal gates on thin FD channels,” in *IEDM Tech. Dig.*, 2002, pp. 263–266.

[5.11] T. Sato, H. Nii, M. Hatano, K. Takenaka, H. Hayashi, K. Ishigo, T. Hirano, K. Ida, N. Aoki, T. Ohguto, K. Ino, I. Mizushima, and T. Tsunashima, “SON (silicon on nothing) MOSFET using ESS (empty space in silicon) technique for SoC applications,” in *IEDM Tech. Dig.*, 2001, pp. 37.1.1–37.1.4.

[5.12] R. A. Rao, R. F. Steimle, M. Sadd, C. T. Swift, B. Hradsky, S. Straub, T. Merchant, M. Stoker, S. G. H. Anderson, M. Rossow, J. Yater, B. Acred, K. Harber, E. J. Prinz, B. E. White, Jr., and R. Muralidhar, “Silicon nanocrystal based memory devices for NVM and DRAM applications,” *Solid State Electron.*, vol. 48, no. 9, pp. 1463–1473, Sep. 2004.

[5.13] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan, “Volatile and non-volatile memories in silicon with nano-crystal storage,” in *IEDM Tech. Dig.*, Dec. 1995, pp. 521–524.

[5.14] R. F. Steimle, M. Sadd, R. Muralidhar, R. Rajesh, B. Hradsky, S. Straub, and B. E. White, Jr., “Hybrid silicon nanocrystal silicon nitride dynamic random access memory,” *IEEE Trans. Nanotechnol.*, vol. 2, no. 4, pp. 335–340, Dec. 2003.

[5.15] M. She and T.-J. King, “Impact of crystal size and tunnel dielectric on semiconductor nanocrystal memory performance,” *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1934–1940, Sep. 2003

[5.16] Y. Nakajima, K. Sasaki, T. Hanajiri, T. Toyabe, T. Morikawa and T. Sugano, Physica E: Low-dimensional Systems and Nanostructures, Vol:24, Issues:1-2, pp. 92-95, Aug. 2004.

[5.17] Dieter K. Schroder, "Semiconductor Material and Device Characterization", Wiley-IEEE Press

[5.18] M. G. Ertosun, K-Y. Lim, C. Park, J. Oh, P. Kirsch, K. C. Saraswat, "Novel Capacitorless Single-Transistor Charge-Trap DRAM (1T CT DRAM) Utilizing Electrons," IEEE *Electron Device Letters*, vol.31, no.5, pp.405-407, May 2010

Chapter 6

Conclusions and Future Work

6.1 Thesis summary

Capacitorless DRAM, because of its smaller size, has been researched widely recently. We experimentally demonstrated a double-gate (DG) capacitorless single-transistor (1T) DRAM on a bulk silicon wafer [6.1], which is a big advantage in terms of processing and process integration. In this device, one of the MOS gates is used as a conventional switching transistor, whereas the other (back) gate is used to create the floating-body storage node. By reverse biasing the back gate, one can keep the excess holes in the body and obtain memory operation even for the highly scaled devices under fully depleted condition. This vertical source/drain structural configuration does not suffer from width discretization issues as experienced with FINFETs and provides a simple back gate option stemming from electrical isolation between two gates. Also,

since the proposed structure has a low body impurity concentration, it has a lower junction leakage current, hence longer retention time. Finally, a DG structure effectively terminates the field lines from the drain; one can use a relatively thicker gate oxide than what is used for the logic transistors, which helps increase the body coefficient.

We have experimentally demonstrated the DG capacitorless 1T DRAM on a bulk silicon wafer, with the measured retention times of up to 25 ms [6.1]. The scalability limits for conventional 1T/1C DRAM is the capacitor, whereas in the 1T DRAM, this limit is due to the lithography. Our devices do not have capacitors, and their gate lengths are not defined by lithographic processes. In fact, various process innovations enable gate length definition using well-controlled and characterized etch processes. Because this process is capable of much smaller features, it facilitates scaling of the DRAM.

Furthermore we introduced a novel 1-transistor (capacitorless) quantum well DRAM (1T-QW DRAM) [6.2]. First we use extensive simulations to find the optimum parameters for DG 1T DRAM. Subsequently, we do a detailed analysis of the 1T-QW DRAM, one version of which using a SiGe or Ge quantum well in Si. Compared with conventional 1T DRAM, this new memory has superior characteristics, such as the ability to produce higher V_t shifts, hence wider sensing window, and also control over the distribution of the holes within the body. The distribution of stored holes can be successfully moved closer to the front gate, which is impossible in conventional 1T

DRAMs where holes are stored just close to the back gate interface. This property not only yields an increased V_t shift (i.e. higher sense margin), but also a higher retention time [6.3, 6.4, 6.5].

This novel 1T QW DRAM has several advantages in terms of performance and scalability. One of them is introducing a "storage pocket" within the device. It also allows the possibility of engineering the spatial distribution of the holes in the body of the device, which is not possible in the other 1T DRAMs. We studied the effects of the germanium content in the Quantum Well in terms of difference in I_d (ie. V_t shift), sensing current window, '0' state degradation and the erase mechanism. It is found that the use of SiGe instead of pure germanium for the Quantum Well has added advantages such as: Improvement in '0' state degradation, and hence in retention; the reduction in the need of PowerErase; the added ability to be able to form better interfaces with Si/SiGe heterostructure, compared to Si/Ge heterostructure; and ease in fabrication process. This new memory has the ability to achieve higher I_d difference values and retention values and a better candidate for scaled new technology nodes. We successfully quantify these advantages [6.6, 6.7].

In conventional capacitorless single-transistor (1T) DRAMs, excess holes are created in the body of transistors using impact ionization or gate-induced drain leakage mechanisms. The existence of holes in the floating-body reduces the threshold voltage (V_t), so a higher drain current (I_d) is obtained during a read operation compared to the

case where there are no excess holes. A current-sensing circuitry is employed to sense the I_d level, hence the state that is stored within the device. We propose and experimentally demonstrate [6.8] another novel DRAM which utilizes the charge-trap (CT) mechanism in its memory operation. This is achieved by engineering the device to create an intentional electron-trapping zone (ETZ) within the body of the device. Furthermore, lower voltage and faster operation is achieved with this novel body-engineered charge trap (CT) device compared to previously explored gate-stack-engineered tunneling-based CT devices.

This novel 1T CT DRAM [6.6, 6.8] relies on the existence and absence of electrons within its body and uses a charge-trapping mechanism in its memory operation, unlike conventional 1T DRAMs that rely on holes and employ floating-body effects in their memory operation. An ETZ has been created within the body of an FD SOI device. We use the traps formed along void structures to create the ETZ, but a 1T CT DRAM could be realized by forming the necessary ETZ with other methods such as embedding nanocrystals or other appropriate materials.

6.2 Contributions and future work

- For the first time, we have experimentally shown:
    - a vertical double gate (DG) capacitorless single-transistor (1T) DRAM
    - on a bulk silicon wafer
    - fabricated with process innovations

- Double Gate Capacitorless 1T DRAM is analyzed with extensive simulations

- For the first time, energy band engineering approach for 1T DRAM: A novel 1T quantum well DRAM (1T QW DRAM) is introduced:
    - "Storage pocket" within the device.
    - Possibility of engineering the spatial distribution of the holes
    - Ability to achieve higher V_t shift and also retention values

- A novel substrate/body engineered 1T Charge Trap DRAM (1T CT DRAM) is introduced:
    - Employing charge trapping effect and utilizing electrons

    - For the first time, charge trapping mechanism within the body
    - For the first time, demonstration of charge trapping along with fast and low voltage DRAM operation
- More detailed theoretical and simulation analysis of 1T CT DRAM (future work)

6.3 References

[6.1]M. G. Ertosun; H. Cho; P. Kapur; K.C. Saraswat; "A Nanoscale Vertical Double-Gate Single-Transistor Capacitorless DRAM," Electron Device Letters, IEEE , vol.29, no.6, pp.615-617, June 2008

[6.2]M. G. Ertosun; P. Kapur; K.C. Saraswat; "A Highly Scalable Capacitorless Double Gate Quantum Well Single Transistor DRAM: 1T-QW DRAM," *Electron Device Letters, IEEE* , vol.29, no.12, pp.1405-1407, Dec. 2008

[6.3]K. Saraswat, M. G. Ertosun, D. Kim, T. Krishnamohan, D. Kuzum, P. McIntyre, Y. Nishi, A. Okyay, J. Park, H. Yu; "Ge MOSFETs and Single Transistor DRAM" 6th International Symposium on Advanced Gate Stack Technology, August 2009 (Invited)

[6.4]M. G. Ertosun; K. C. Saraswat; “Characteristics Of The Capacitorless Double Gate Quantum Well Single Transistor DRAM” IEEE International Conference on Simulation of Semiconductor Process and Devices 2009, September 2009

[6.5]M. G. Ertosun, K. C. Saraswat; “Novel Memory Devices: Capacitorless Quantum Well 1T DRAM”, Semiconductor Research Corporation (SRC) / The US National Science Foundation (NSF) / The Singapore Agency for Science, Technology and Research (A*STAR) Forum on 2020 Semiconductor Memory Strategies: Processes, Devices, and Architectures, Singapore, October 2009 (Invited)

[6.6]M. G. Ertosun; “Novel Single Transistor DRAM Technologies”, 5th Emerging Technologies in Solid State Devices Workshop, Baltimore, December 2009 (Invited)

[6.7]M. G. Ertosun, K. C. Saraswat; "Investigation of Capacitorless Double-Gate Single-Transistor DRAM: With and Without Quantum Well," IEEE Transactions on Electron Devices, vol.57, no.3, pp.608-613, March 2010

[6.8]M. G. Ertosun, K-Y. Lim, C. Park, J. Oh, P. Kirsch, K. C. Saraswat, "Novel Capacitorless Single-Transistor Charge-Trap DRAM (1T CT DRAM) Utilizing Electrons," IEEE *Electron Device Letters*, vol.31, no.5, pp.405-407, May 2010