Order Number 1355158

A model for estimating allowable transition metal contamination in DRAMs

Schmid, John Robert, M.S.

The University of Arizona, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A MODEL FOR ESTIMATING ALLOWABLE TRANSITION METAL CONTAMINATION IN DRAMs

by

John Robert Schmid

A Thesis Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE
WITH A MAJOR IN ELECTRICAL ENGINEERING

In the Graduate College

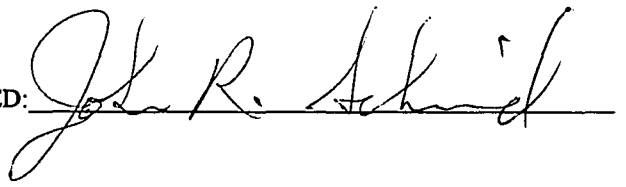
THE UNIVERSITY OF ARIZONA

1 9 9 3

STATEMENT BY AUTHOR

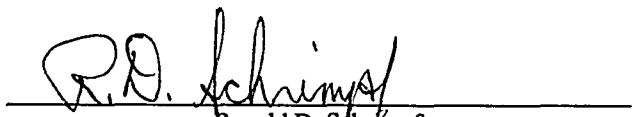
This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: 

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:



Ronald D. Schrimpf
Associate Professor of
Electrical and Computer Engineering

7/29/93
Date

ACKNOWLEDGMENTS

This work was performed at the University of Arizona SEMATECH Center of Excellence for Defect/Contamination Assessment and Control and Sandia National Laboratories under contract to SEMATECH as administered by the Semiconductor Research Corporation under contract # 88MC-501.

I wish to thank all of the people working on SEMATECH Task III (Contamination Induced Electrical Effects, University of Arizona SCOE) for their assistance and support in the work which resulted in this thesis. In particular I wish to thank my advisor, Dr. Ronald Schrimpf, and the members of the Task, Dr. Harold Parks, Dr. Jacob Hohl, Dr. John Ramberg, Mr. Robert Craigin, and Mr. Chris Damianou, who all made valuable contributions to this work. In addition, I wish to extend my thanks to Dr. Ron Jones and Dr. Norm Brown of Sandia National Laboratories, whose labors allowed experimental verification of the model. SEMATECH made this research possible by supporting the project, through not only funding, but also support from its members and mentors. With their guidance, valuable experience and knowledge, many seemingly insurmountable problems were solved.

Finally, I wish to thank my family, Mary and Bryce, for their patience and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES CONTINUED

LIST OF TABLES

ABSTRACT

Due to new memory-cell architectures, the leakage-current requirements for semiconductor memories will become less stringent with increased levels of integration. The implication of these requirements with regard to allowable metallic contamination levels is investigated with a one-dimensional model based on Shockley-Read-Hall generation-recombination. The model was developed to predict leakage-current in carrier-depleted regions as a function of basic process and metallic contaminant parameters. As device dimensions are reduced, transition metal homogeneous contamination in process chemicals can be an important source of generation-recombination centers that result in the dominant generation-current in the space-charge region. The model allows an estimation of an upper bound for transition metal contamination in advanced processes and is applied for DRAM leakage predictions. Using the model, it is demonstrated that the trend toward lower leakage-current density requirements reverses after the 64-Mbit generation DRAM as a result of memory-cell architecture trends which significantly reduce the space-charge volume.

1. INTRODUCTION

This thesis presents a method for estimating the affects of metal contaminants on leakage current in silicon devices. For this model to be useful in a practical sense, it should be applicable to a general type of process. The model depends on basic process parameters such as background doping level and certain characteristic parameters of a contaminant species when incorporated into the bulk silicon. The utility of the model lies in the fact that if process, device and circuit requirements are known, estimations of required process cleanliness can be made. To gage the usefulness of such an approach, the model is applied to one of the main technological drivers of silicon technology, the DRAM memory cell. By identifying leakage current requirements of the DRAM cell, the model can then be applied to determine the bounds for allowable metal contamination in the CMOS process.

To briefly outline the topics covered in this thesis, the next chapter reviews DRAM memory cell structure, function and requirements in order to identify the leakage current limits of advanced cells. In the third chapter, the analytical development of the model is presented, along with examples of application. In the fourth chapter, the details of a controlled contamination experiment are presented, with leakage current results as a function of metal contamination. These results are then compared to the model.

1.1 Leakage Current as a Limiting Factor for Advanced Processes.

Leakage current in a DRAM cell creates a problem since the leakage drains away the stored charge used to represent the binary information. Obviously, the cell can only tolerate a certain amount of leakage after which, at some reduced level of charge, the cell's bit of information no longer represents the same data. Not only can leakage current affect the information in a DRAM cell, as will be discussed in Chapter 2, leakage current affects other critical DRAM parameters such as access time. Previous estimates of junction leakage based on the refresh cycle design requirements for a 0.5 μm geometry 16 Mbit DRAM were reported to be less than $2x10^{-16}$ A/bit or $5x10^{-9}$ A/cm^2 [1, 2, 3]. Based on data

from present 256 Mbit cells and newer estimates on requirements resulting from 1 Gbit cell architecture implementation, this thesis presents an estimate of the 1 Gbit DRAM leakage requirement of less than 1.95×10^{-17} A/bit. This requirement puts an upper limit of approximately 4×10^{-9} A/cm^2 on leakage current density for 1 Gbit DRAM processes. This current density value is not significantly different from the current density requirements for the 16 Mbit DRAM due to cell structure changes that have significantly reduced leakage.

1.2 Overview of the Model and Application

A one-dimensional model based on Shockley-Read-Hall deep-trap theory was developed to calculate generation in the space-charge region and recombination in the undepleted bulk silicon as a function of metallic impurity type and concentration. Generation lifetimes are calculated and these in turn are used to estimate junction leakage current. The model assumes that metallic impurities are homogeneously distributed throughout the bulk silicon. The density of generation-recombination centers is assumed to be equal to the impurity density, and minority-carrier lifetimes are calculated using the trap density and experimental values of capture cross-sections. The model evaluates the affects of multiple trapping levels introduced by a particular metal contaminant, and analyzes leakage resulting from mixtures of metal contaminants at various relative concentration levels. This allows evaluation of junction leakage resulting from typical multi-metal contaminant sources; stainless steel, for example, would include the affects of iron, nickel and chromium. The relative concentrations of the specific contaminants can also be manipulated to match the contaminant source. Using various deposition studies in the literature which relate the amount of contaminant deposited onto the wafer to a sequence of processing steps, it is possible to estimate required process cleanliness based on junction leakage requirements.

For given concentrations of metallic impurities, the model predicts the generation-recombination rate and leakage current for pn junctions. The model can be used to estimate upper limits on allowable

metallic contamination concentration based on lifetime and leakage current degradation. Models for carrier lifetime as a function of G-R center concentration were established by prior work [4]. The present work extends this result to allow mapping of contamination concentration in the processing solutions to leakage currents in devices. Assuming generation in the space-charge region is the only component of reverse biased junction leakage current, first order analysis indicates that transition metal contamination must be kept below $4x10^{12}cm^{-3}$ for 1 Gbit DRAM application. Reverse bias junction leakage, estimated in this manner with the effects of deep-trap levels, enables an upper bound on allowable G-R center concentration, and therefore an allowable transition metal contamination, to be estimated.

2. LEAKAGE CURRENT IN DRAMs

Discussions on physical limiting phenomena of semiconductor device performance often utilize the dynamic random access memory (DRAM) as the technological example. DRAMs drive silicon technology because the memory densities required by tomorrow's computers demand the ultimate in miniaturization from memory circuits, along with rapid access times. Large corporations devote significant resources to developing sub-micron complementary metal-oxide-semiconductor (CMOS) processes required for the next generation of DRAMs. The bottom line is that the immense scientific, economic and social impetus to improve computer performance results in much of the leading edge research in CMOS technologies initially being applied to DRAMs.

There are many parameters that affect DRAM performance, however, the discussion which follows is confined to parameters that affect leakage current in DRAMs, since these parameters are germane to the leakage current model application. Essentially all DRAMs in production and being developed are based on the one transistor memory cell, with the transistor acting as a transfer gate for charge stored on the memory cell capacitor. Since the memory cell suffers a gradual loss of charge due to leakage current, the memory cell must be "refreshed" periodically. Refresh requirements dictate the amount of leakage that can be tolerated from a memory cell. In the development of next generation DRAMs, it is advantageous to increase refresh time (decrease leakage). The primary reason for this is related to memory time required to support the CPU and power dissipation issues from refreshing more word lines, which results in the periodic operation of larger numbers of sense amplifiers (the number of sense amplifiers on the DRAM die increases proportionally with the level of integration). Increases in the number of sense amplifiers affects layout complexity, chip area and memory availability to the operating system, in addition to active power dissipation on the die. Decreasing the memory cell leakage allows increased refresh time and the longer refresh period alleviates some of the problems mentioned above.

To gain a perspective on the leakage current model's utility for evaluating leakage requirements of advanced processes such as DRAMs, a brief overview of DRAM structure and function is presented in the following section. With this background, a qualitative feel for memory-cell-related parameters is

established and a more detailed discussion of leakage components in a DRAM can be developed in the subsequent section. In the final section of this chapter, leakage current limits are examined from the aspect of implications for future DRAM processes. As parameters are introduced, their abbreviations are given; also, a list of parameter abbreviations can be found in Appendix A.

2.1 A Review of General DRAM Circuit Components and Function

To understand the magnitude of leakage current decrease required to allow an increase of refresh time for next-generation DRAMs, a brief tutorial on DRAMs follows. This discussion includes examination of DRAM structure, function and leakage related parameters for present and future DRAMs.

Read/write memory has become known as random-access memory (RAM). However, read-only memory (ROM) is also randomly accessible, so the use of the term random-access memory is somewhat imprecise. Data stored in RAM are not permanent in nature, unlike ROM, the data can be changed. Another difference between ROM and RAM is that RAM is a volatile memory while ROM is not. A volatile memory will lose the data stored when power is removed from the circuit, while ROM will retain its data. There are two basic types of RAM manufactured with MOS technologies, static RAM (SRAM) and dynamic RAM (DRAM). SRAM will maintain data stored in the memory cell for an indefinite period as long as power is not removed. For DRAMs, just maintaining power is not sufficient to retain data in a memory cell; data maintenance additionally requires a periodic rewrite of the data into the memory cell. This added requirement is necessary because the storage elements are capacitors. If the storage cell capacitors are not recharged by rewriting the data, they discharge through leakage from the memory cell node and the data are lost. This recharging process is known as "refreshing" the RAM.

The basic architecture of a DRAM is shown in Figure 1. Components of the DRAM include the address buffers, control buffers, data buffers, row (word) and column (bit) decoders, sense amplifiers, refresh circuitry, output buffers and memory array. The address buffer serves to limit the loading on the circuitry driving the address inputs of the DRAM. The buffer of an address line typically consists of a few

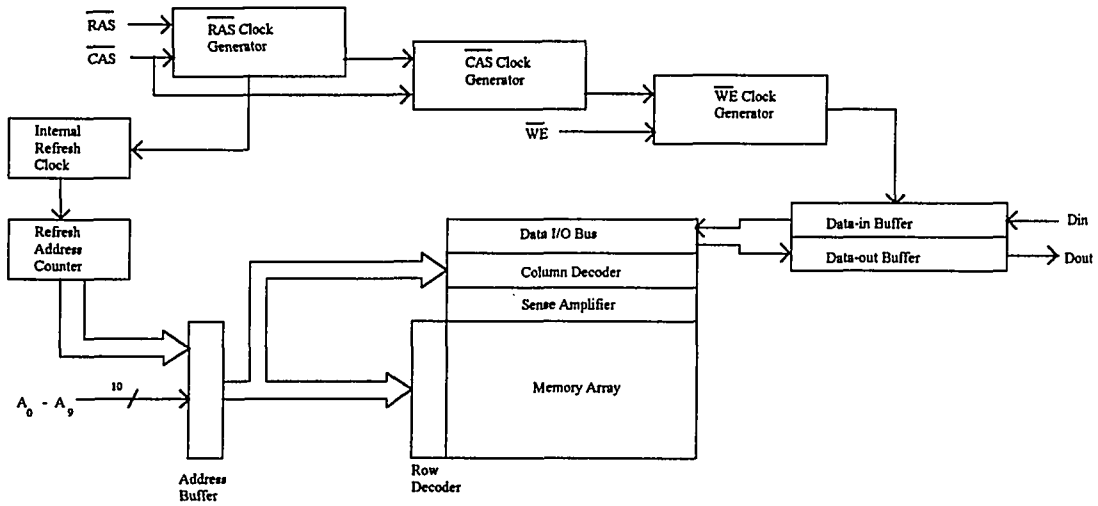


Figure 1. The basic components of a 1 Mbit DRAM.

simple logic gates connected in series; they affect the timing of address transmission to the row and column decoders. Most modern DRAMs lower their pin count by using address sharing. The initial group of address signals are all sent to the row decoder, and a following group of address signals (again using all address bits) is sent to the column decoder. This is accomplished with two control or clock signals, row address strobe (RAS) and column address strobe (CAS). The row and column decoders are simply logic that selects the appropriate word and bits for the write or read. The outputs of the memory array require amplifiers and output buffers to enable interface with external circuitry. The outputs of the memory cells are very low amplitude voltages and currents. The sense amplifiers provide the detection and amplification of these signals. The amplifiers are often cross-coupled flip-flops with a maximum reliable sensitivity in the 100-150 mV range [5]. The refresh circuitry typically consists of a counter which generates refresh addresses for the rows of the memory array and refresh control logic which disables the outputs of the DRAM. The refresh control logic disables the output so that during the refresh, data do not appear at the data-out buffer.

Finally, the DRAM component of primary interest from a leakage perspective is the memory array. This array is composed of a vast number of memory cells, each cell storing a bit of information.

These cells are ordered in various matrix patterns. The level of integration of a DRAM is basically the number of bits of data that a memory array can store. As mentioned previously, the charge storage memory cell consisting of one transistor and one capacitor, shown in Figure 2, is the basis of all DRAMs currently in commercial production, and is expected to remain the dominant architecture through the 1 Gbit level of integration [5].

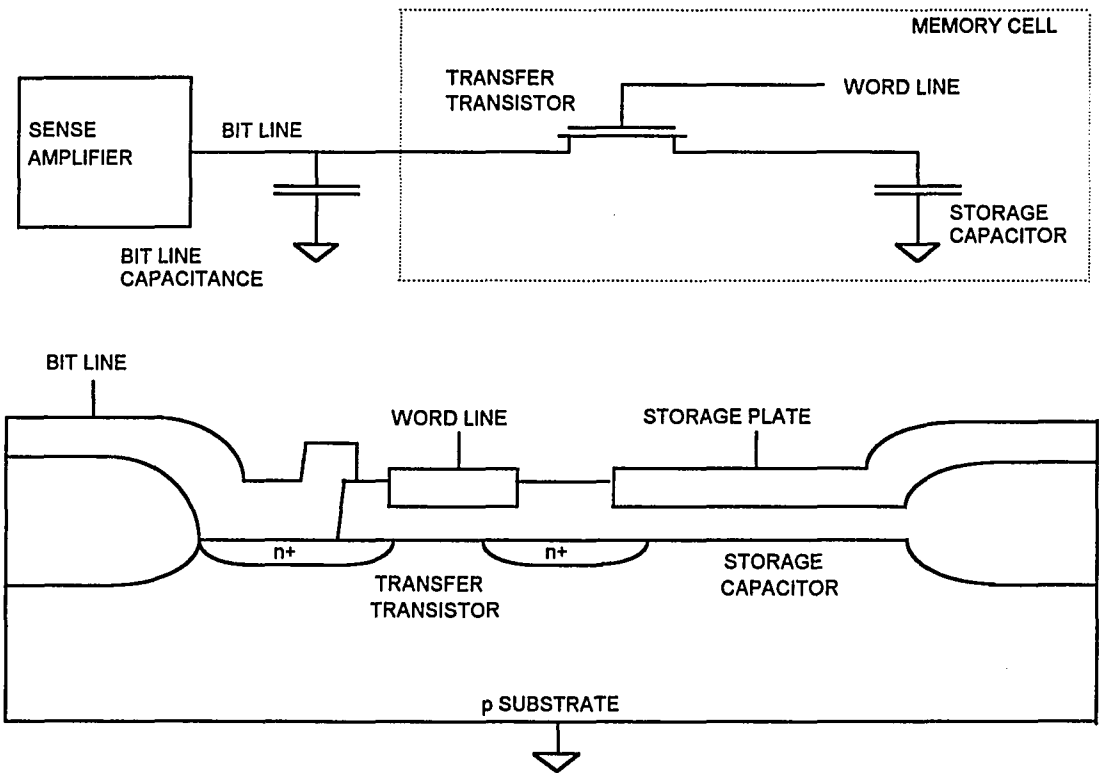


Figure 2. The circuit diagram and cross-section for the basic one-transistor, planar-memory cell.

Binary information, the "0" or "1", is stored on the memory cell capacitor. The charge stored on the capacitor is accessed through the transfer transistor for read or write operations. The transistor is turned off and on by the word line; when turned on, the transistor connects the storage capacitor to the bit line, allowing the charge stored on the capacitor to cause a signal at the sense amplifier. This method of

storing information is based on the charge storage concept and is the method on which all modern DRAMs are based.

Current DRAMs are almost exclusively built with CMOS processes, taking advantage of low power dissipation, simpler designs, and improved noise margins. During development of the modern DRAM, chips with densities up to 1 Mbit typically used planar storage capacitors on the die. DRAMs with higher densities, but with approximately the same chip size, were achieved by developing a capacitor that exploited the vertical dimension on the die. Structures such as trench or stacked capacitors were developed and will be discussed in the following section.

2.2 DRAM Parameters and Memory Cell Structure

In 1971, 1 Kbit DRAMs became commercially available. This first "high-density" DRAM had a three-transistor cell [6]. Since that time, essentially every DRAM commercially manufactured has been based on the one-transistor cell. From the time of this 1 Kbit DRAM, the level of integration has increased four times every three years for the last 22 years, so that today 16 Mbit DRAMs are commercially available. Indeed, NEC recently announced production of a 256 Mbit Chip [7], not just a prototype memory cell; additionally, Toshiba recently built a prototype memory cell with 256 Mbit design rules that is intended to be scaled for a 1 Gbit cell [8].

This significant increase in the level of integration has been accomplished in a number of ways. First process advancements over these years have been dramatic. The early DRAMs were built with NMOS processes; in the 1980s, DRAM processes converted to CMOS, which offered advantages in power dissipation, design simplification, higher cell density, improved noise margins and better soft error protection. Additional process improvements included ion implantation, plasma etching, LOCOS, reflow glass, silicides and optical projection lithography. These process improvements, along with others, have allowed the memory cell area to be reduced by approximately three times for each generation. The NEC 256 Mbit DRAM mentioned above has a memory cell area of only 0.72 μm^2. It is interesting to note that

the 1989 prediction for 256 Mbit DRAM cell size (A_c) by Tasch and Parker [5] of 0.17-0.34 μm^2 was optimistic. Their prediction was based on the following expression:

$$A_c = \frac{A_d \cdot \eta}{N}, \tag{2-1}$$

where N is the number of bits ($N = 2.95 \times 10^8$ for a 256 Mbit DRAM, with a 15% increase in bit number for redundancy) and η is the array efficiency that was assumed to be 50% based on historical consistency over several generations. Additionally, Tasch and Parker predicted die size (A_d) for the 64 Mbit and 256 Mbit DRAM to range between 100 and 200 mm^2. Their 0.34 μm^2 cell area is derived from using the 200 mm^2 die area. The NEC 256 Mbit DRAM has a die area of 333 mm^2 and a 0.72 μm^2 cell area, indicating moderate improvements in efficiency, η, even though the actual die and cell areas are significantly larger than predicted. Trends to increase efficiency are a result of the fact that physical phenomena in the submicron region have restricted the continued application of traditional scaling laws to obtain high memory density. Due to these restrictions, which will be discussed in the following sections, new techniques must be employed to attain density increases and higher efficiency. Innovative memory cell architecture and peripheral circuit design cleverness are becoming the areas where higher levels of integration can be made.

Accompanying this cell area decrease with each DRAM generation has been a somewhat smaller, 1.4 times per generation, increase in the size of the die. The 1970's 1 Kbit DRAM had a die area of 10 mm^2, while the NEC 256 Mbit DRAM has a die area of 333 mm^2. A die this large is somewhat undesirable because, from a system point of view, a package width larger than 500 mils may force hardware changes, and from a wafer process point of view, this die size approaches present reticle limits. One of the ways that designers are trying to surmount this die area restriction and also improve efficiency is through the use of leads-over-chip (LOC) layout [9]. In LOC layout, the bonding pads on the die are arranged down the middle of the chip (see Figure 3). This layout allows a reduction in internal power bus routing and improves the output switching characteristics. Also shown in Figure 3 is the unusual package required for this layout. The top of the die is attached to the bottom of the leadframe; it is believed this will result in a

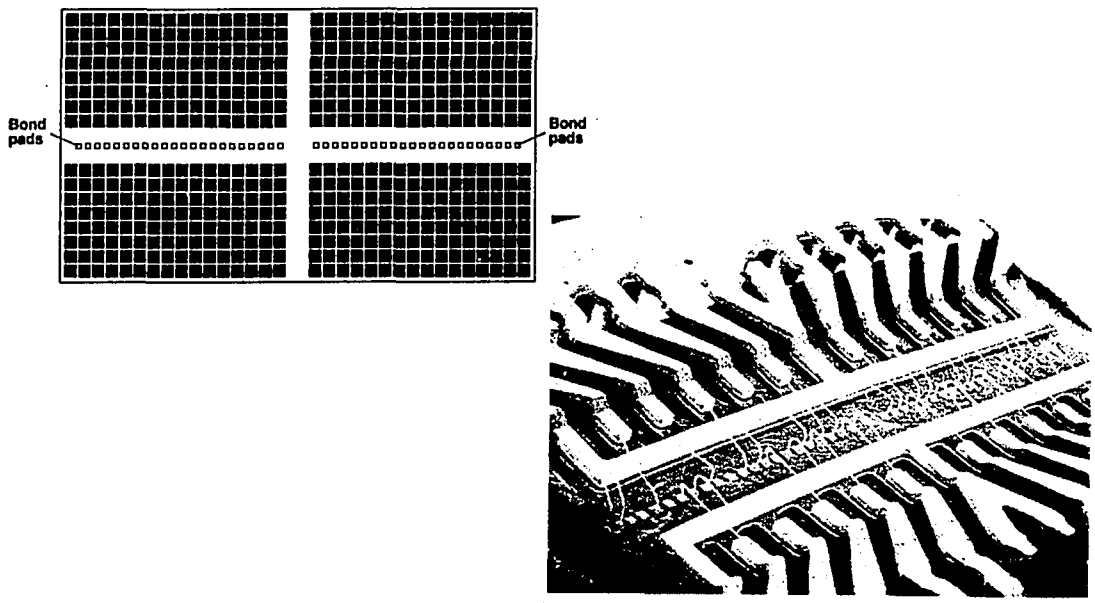


Figure 3. LOC pad layout and assembly allows increased die efficiency, η, and internal power bus routing and switching characteristics are improved (from reference [9]).

stronger and more reliable package. Another simple way to fit a large die in a package of restricted width is to make the die long and narrow; most DRAMs have about a 2:1 ratio of die length to width. An example of this strategy, shown in Figure 4, are a photographs of a Texas Instruments 1 Mbit DRAM die (4.5 x 11.2 mm) and the NEC 256 Mbit DRAM die (13.6 x 24.5 mm).

The historical trends in both memory cell area and die area are shown in Figure 5. One of the main reasons allowing the continued cell size reduction has been a reduction in the minimum feature size of approximately 1.4 times per generation. One of the conventional design rules that relates feature size (F) to cell size (A_C) is

$$A_C = 8F^2. \tag{2-2}$$

Figure 6 plots the minimum feature size versus cell area for cells found in the literature; also shown on

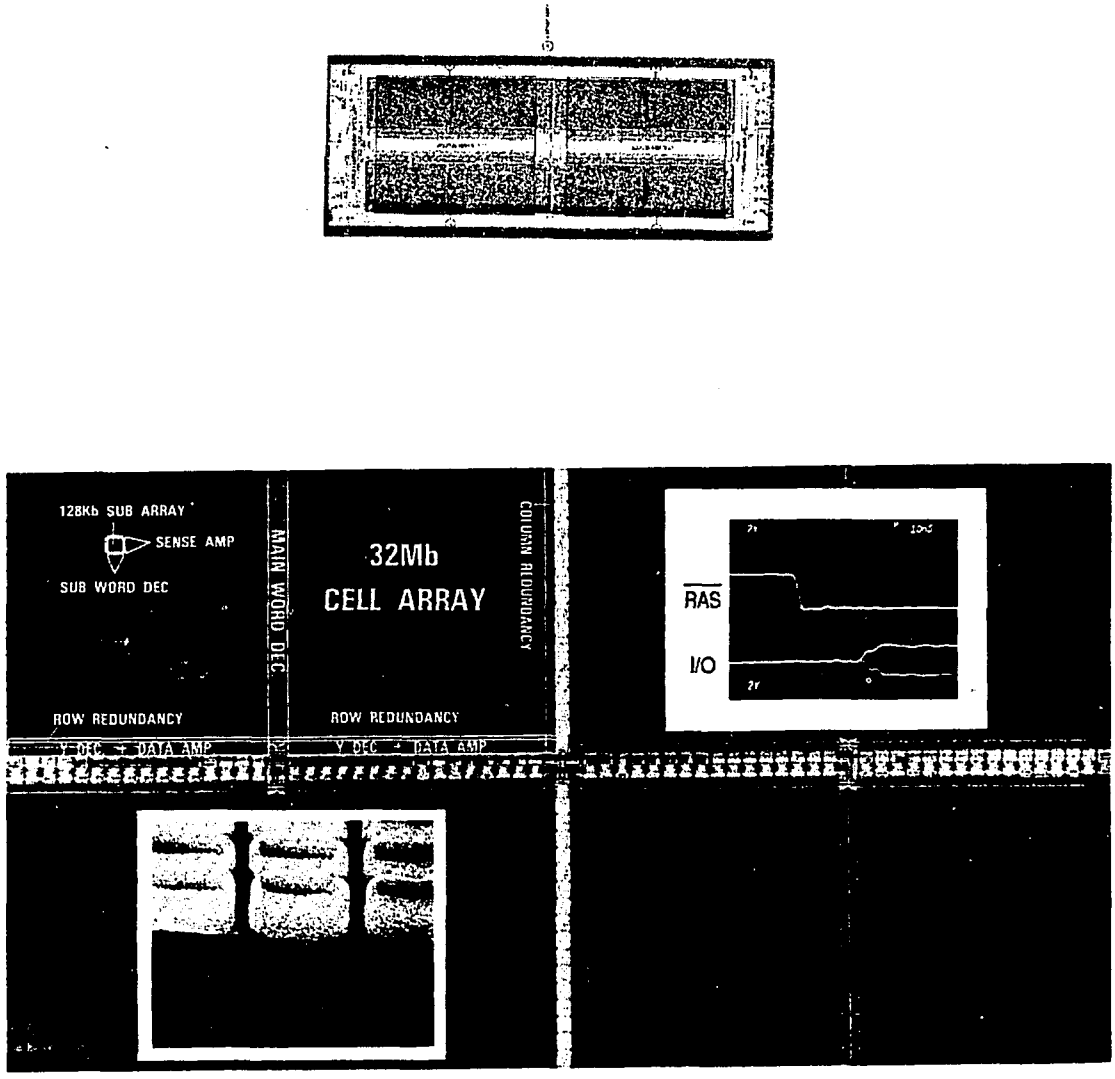


Figure 4. A photograph of a Texas Instrument 1 Mbit DRAM die (4.5 x 11.2 mm) and a NEC 256 Mbit DRAM die (13.6 x 24.5 mm) from reference [7]. The scale is approximately 6X magnification of both die; readily apparent from the comparison is the substantial increase in die size with level of integration. This figure also illustrates the use of a long, narrow die to satisfy packaging requirements.

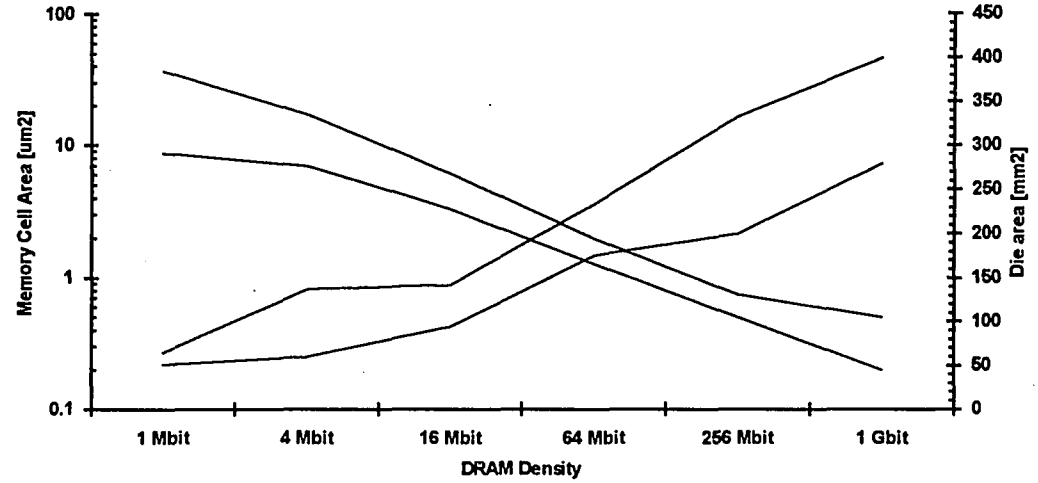


Figure 5. The memory cell area and die area in relation to level of integration. The two lines represent the minimum and maximum found in the literature for actual cells and die (1 Gbit cell and die areas are speculative projections from the literature).

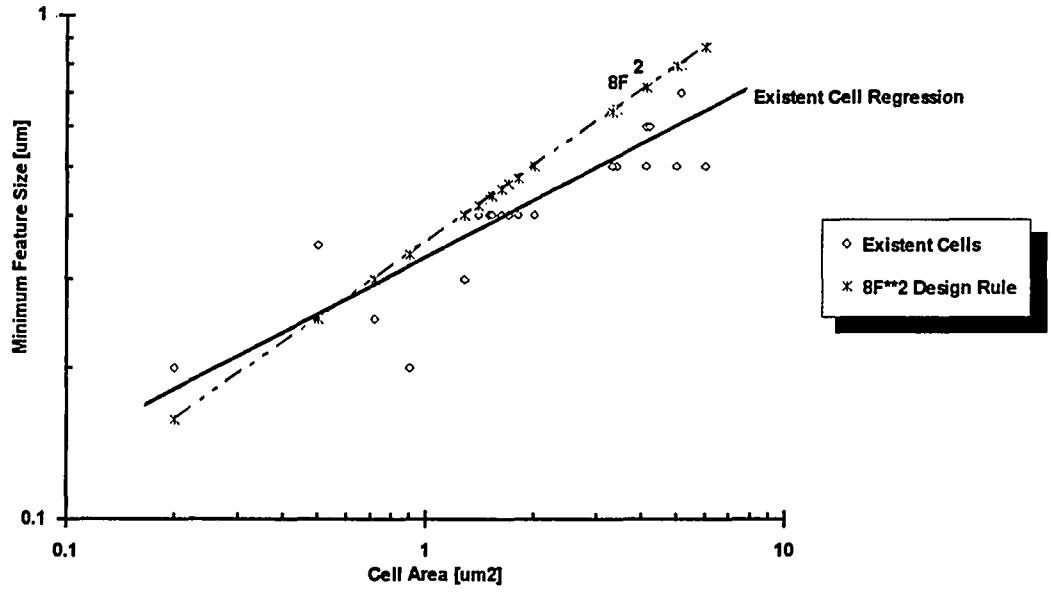


Figure 6. Minimum feature width versus cell area for existent cells; also shown is the conventional design rule, $8F^2$.

this plot, as a basis for comparison, is the $8F^2$ design rule. Tables 1A (1 and 4 Mbit DRAMs), 1B (16 and 64 Mbit DRAMs) and 1C (256 Mbit and 1 Gbit) summarize a number of important DRAM parameters from various sources in the literature.

Obviously, the 1.4X reduction in feature size alone is not going to allow the trend of 3X cell size reduction to continue. The main problem in continued cell size reduction is due to the fact that the amount of charge stored in the cell has remained relatively constant over the years. This is forced by a number of constraints. Noise from background radiation, alpha particle flux from packaging materials and die metalization, can cause charge variations within the storage cell by the generation of minority carriers. Typically, the charge within the modern cell is on the order of 150 fC (see Table 2). Further reduction of charge below this level risks a significant data error rate or soft error rate (SER), due to alpha particle flux. With an increase in circuit density, noise from circuit operation also becomes more of a problem. So noise from both radiation and circuit operation along with, of course, thermal noise requires that the amount of stored charge, which is used as the sensing signal, remain at an adequately high level to allow a reasonable signal to noise ratio. From Figure 2 it can be seen that the voltage at the sense amplifier is a function of both the bit line and storage capacitances. The charge within the cell can be written as:

$$Q_c = \int_{V_{sg}}^{V_{sg}} C_s(V_s)dV_s \approx C_s \Delta V_s, \tag{2-3}$$

where Q_C is the charge stored in the cell, C_S is the storage cell capacitance and V_S is the voltage at the storage node. The voltage at the sense amplifier is approximately:

$$\Delta V_{sen} \cong \frac{C_s \cdot \Delta V_s}{2(C_s + C_{BL})} - \Delta V_{noise}, \tag{2-4}$$

where V_{sen} is the signal at the sense amplifier, C_{BL} is the bit line capacitance, and ΔV_{noise} is the noise component due to SER, leakage, reference cell mismatch, V_{DD} variations, etc. The bit line capacitance has remained relatively constant or has increased with each generation as a result of more bits per bit line, even though the bit line capacitance per bit has been reduced because of cell area reduction. The bit line

Mbit	Ad [mm^2]	Ac [μm^2]	Feature [μm]	Cell Type	Vdd [V]	T refresh [msec]	Access [nsec]	Pack Width [mil]	Manufacturer	Reference
1	65	35.7	1.2	HI-C (P)	5	8	90	300	Mitsubishi	10
1	50.8	28.5	1	P	5	8	65		Intel	11
1		36.8	1.25	HI-C (P)	5				AT&T	12
1		26.5		STC (S)				300	Fujitsu	13
1		8.8	0.8	BSE (T)					NEC	14
1	52.7	26	1	CCC (T)				300	TI	15
4		12.7	0.9	BSC (T)					Fujitsu	16
4	99.2	10.6	0.8	BSE (T)	5	16	95		NEC	17
4		10.9	0.8	FCC (T)	5			300	Mitsubishi	18
4	60	7	0.8	FCC (T)	5			300	NEC	19
4	100	9	1	TTC (T)	5	16	170		TI	20, 21, 22, 23
4		11	0.8	FCC (T)	3.3				Toshiba	24
4		10.6	0.9	CCC (T)					Siemens	25
4	137	17.4	1	SPT (T)	3.5		80	400	Toshiba	26

Table 1A. A summary of literature on important 1 and 4 Mbit DRAM parameters. Under cell type, (P), (T) and (S) refer to planar, trench and stacked respectively. See Appendix C for a detailed description of the individual cell types.

Mbit	Ad [mm^2]	Ac [μm^2]	Feature [μm]	Cell Type	Vdd [V]	T refresh [msec]	Access [nsec]	Pack Width [mil]	Manufacturer	Reference
16		5.12	0.7	BSC (T)					Siemens	27
16	130	4.1	0.6	STC (S)	3.3	32	55		NEC	28
16		3.4	0.5	FCC (T)	3.3				Hitachi	29
16		5	0.5	IVEC (T)	3.3				NTT	30
16		5	0.5	FCC (T)	3.3				Toshiba	24
16		6	0.5	BSC (T)	3.3				Mitsubishi	31
16		4.2	0.6	BSC (T)	3.3				Hitachi	32
16	141.9	4.2	0.6	STC (S)	3.3		60		Hitachi	33, 34
16		6.1		BSC (T)	4		70		Toshiba	35
16		3.4	0.5	STC (S)					Hitachi	36
16	93.9	3.3	0.5	IVEC (T)	5			300	Matsushita	3
16	140.9	4.1	0.5	IVEC (T)	3.3				IBM	37, 38
16	134.8	4.8	0.5	STC (S)	3.3	32	60	400	Mitsubishi	39
16	136.9	4.8	0.6	BSC (T)	4		45		Toshiba	40
64	234.4	2	0.4	STC (S)	3.3		50		Matsushita	41
64	233.8	1.7	0.4	STC (S)	3.3	64	45		Mitsubishi	42
64	224.7	1.8	0.4	STC (S)	3.3		40		Fujitsu	43
64	176.4	1.53	0.4	BSC (T)	3.3		33		Toshiba	44
64	197.5	1.28	0.3	STC (S)					Hitachi	45
64		1.4	0.4	STC (S)	3.3				Toshiba	46
64	193.7	1.5	0.4	BTC (T)	3.3				IBM	47
64		1.53	0.4	STC (S)					OKI	48
64	186	1.62	0.4	STC (S)	3	64	30	500	NEC	49

Table 1B. A summary of literature on important 16 and 64 Mbit DRAM parameters. Under cell type, (P), (T) and (S) refer to planar, trench and stacked respectively. See Appendix C for a detailed description of the individual cell types.

Mbit	Ad [mm^2]	Ac [μm^2]	Feature [μm]	Cell Type	Vdd [V]	T refresh [msec]	Access [nsec]	Pack Width [mil]	Manufacturer	Reference
256	333.2	0.72	0.25	STC (S)	2	128	30		NEC	7, 50
256		0.5	0.35	IVEC (T)					Toshiba	8
256		0.72	0.25	STC (S)					Matsushita	51
256		0.75		STC (S)					SamSung	52
256		0.72	0.25	STC (S)					Hitachi	53
256				STC (S)					NEC	54
256		0.9	0.2	STC (S)					OKI	48
256				STC (S)					NEC	55
1024		0.15							IBM	56
1024		0.05			2				IMEC	57
1024		0.2	0.2						Toshiba	8

Table 1C. A summary of literature on important 256 Mbit and 1 Gbit DRAM parameters. Under cell type, (P), (T) and (S) refer to planar, trench and stacked respectively. See Appendix C for a detailed description of the individual cell types.

capacitance is usually approximately 7 to 20 times larger than the storage capacitance [5]; this means that Q_C, the storage charge, has to be large enough to compensate for the capacitive divider and to allow adequate voltage change to ensure detection at the sense amplifier. As mentioned previously, the minimum detectable voltage differences at the sense amplifier are in the 100 to 150 mV range. Also, reducing the stored charge causes an exponential increase in the SER [58]. Due to noise margins and the bit line capacitance, it is expected that there will be no significant reduction in the required charge capacity for future DRAMs.

Previous generations of DRAMs accomplished maintenance of charge capacity in spite of cell area reduction with a decrease in the memory cell's capacitor dielectric thickness. However, the present 5 nm SiO_2 or SiO_2-Si_3N_4 sandwich dielectrics in 16 Mbit DRAMs are very close to the physical limit for reliable dielectric thickness for these materials. In design criteria for advanced cells, one group, as part of their design method, reduced the oxide thickness until the maximum electric field across the storage capacitor dielectric was 4.5 MV/cm or the tunneling current through the capacitor oxide reached 1% of the leakage from the cell due to carrier generation and bulk diffusion [59]. Using the 4.5 MV/cm criterion, the limit on oxide thickness would be about 7 nm with a 3.3V V_S, depending on the process. The storage capacitor's oxide thickness is limited to about 4 nm by direct tunneling of electrons from the

silicon conduction band to the gate electrode for charge densities below 3.1×10^{-6} C/cm^2 and by Fowler-Nordheim tunneling for greater charge densities [60, 61] (see Figure 7). However, this physical limit does

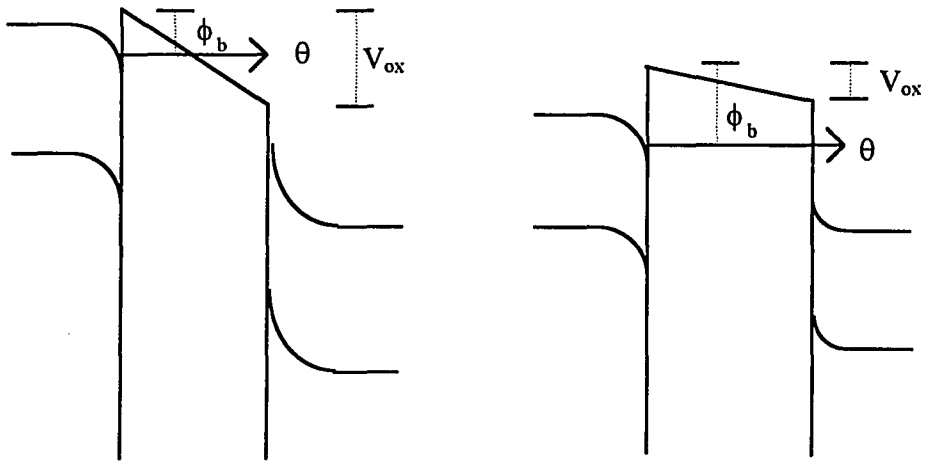


Figure 7. (a) Fowler-Nordheim tunneling is associated with a triangular barrier. (b) Direct tunneling is associated with a trapezoidal barrier, for example when $V_{ox} < \phi_b$.

not take into consideration thickness variations for a capacitor in commercial production. At present, reliable production seems to have limited SiO_2 thickness to approximately 10 nm.

Being at the physical limits of SiO_2 thickness dictates that to maintain the same charge capacity for the cell, the area of the capacitor has also reached a minimum value. This represented a major hurdle during the transition from 1 Mbit DRAMs to 4 Mbit DRAMs, in which continued use of planar architecture (see Figure 2) for the cell capacitor would have resulted is an unacceptable increase in die area. For a charge capacity of 150 fC using 3.3V supplies, a 45 fF capacitor is needed. State-of-the-art oxide-nitride sandwich dielectrics provide about 7-10 fF/cm^2 [9]. These values imply that a capacitor with an area of 4 to 7 μm^2 is needed. This represents an interesting design challenge for the 64 Mbit DRAM with a cell size of only 1.5 μm^2. The ratio of capacitor area to cell area has become a key figure of merit for the DRAM cell; memory cells have had to resort to exploiting the third dimension to maintain this ratio at higher levels of integration. For 3D cells, the A_S/A_C ratios are typically in the range of 1.3 to 5.3 [5].

Table 2 summarizes reported values for these ratios and other capacitor related parameters. Another way to boost charge capacity with less area is to use alternative dielectrics with higher dielectric constants. This move to new dielectrics is happening at a rapid pace since it is really being forced by the fact that present processes have reached the physical limits of the SiO_2 capacitor dielectric. Table 2 shows that a number of advanced cells are using alternate dielectrics; these dielectrics will be discussed in more detail in the final section of this chapter.

To maintain memory cell charge capacity, inspite of a reduction in cell area, the area of the storage capacitor has been increased with an assortment of 3D or vertical cell types. Cell architecture innovations seem to have fallen into two general categories of vertical capacitor structures, the stacked capacitor cell and the trench capacitor cell, each with multiple variations. Figure 8 shows these basic vertical capacitor structures, along with a planar cell. The following discussion on cell types is not intended to cover every architectural variant, it is rather intended to give the reader a feel of the basic cell types and the cell trends which have important implications for DRAM cell leakage current.

Mbit	Ac [um2]	As/Ac	As [um2]	Tox [nm]	Dielectric Material	Cs [fF]	Vdd [V]	Manufacturer	Reference
1	35.7	0.4	14.28	10	SiO_2	45	5	Mitsubishi	10
1	28.5	0.3	8.55	25			5	Intel	11
1	36.8	0.5	18.4	15	SiO_2/Si_3N_4	38	5	AT&T	12
1	26.5	0.5	13.25	10		45		Fujitsu	13
1	8.8	2.3	20.24	18	SiO_2	35		NEC	14
1	26			30				TI	15
4	12.7	3.7	46.99			40		Fujitsu	16
4	10.6					50	5	NEC	17
4	10.9	1.3	14.17	10		50	5	Mitsubishi	18
4	7	2.1	14.7	10		50	5	NEC	19
4	9	2.9	26.1	15	SiO_2	25	5	TI	20, 21, 22, 23
4	11	1.6	17.6	10	SiO_2	60	3.3	Toshiba	24
4	10.6			13		40		Siemens	25
4	17.4					40	3.5	Toshiba	26

Table 2A. Parameters related to charge capacity of the memory cell capacitor for 1 and 4 Mbit DRAMs.

Mbit	Ac [um2]	As/Ac	As [um2]	Tox [nm]	Dielectric Material	Cs [fF]	Vdd [V]	Manufacturer	Reference
16	5.12			9	$SiO_2/Si_3N_4/SiO_2$	50		Siemens	27
16	4.1						3.3	NEC	28
16	3.4	5.3	18.02	10	SiO_2	60	3.3	Hitachi	29
16	5	1.7	8.5	10	SiO_2	30	3.3	NTT	30
16	5	3.5	17.5	10	SiO_2	60	3.3	Toshiba	24
16	6	3.2	19.2	10		50	3.3	Mitsubishi	31
16	4.2	2	8.4	5.5	SiO_2/Si_3N_4	51	3.3	Hitachi	32
16	4.2	1	4.2	5	SiO_2/Si_3N_4	30	3.3	Hitachi	33, 34
16	6.1	1.4	8.54			30	4	Toshiba	35
16	3.4			5	SiO_2/Si_3N_4	35		Hitachi	36
16	3.3	4.3	14.19	10	SiO_2	63	5	Matsushita	3
16	4.1				$SiO_2/Si_3N_4/SiO_2$	100	3.3	IBM	37, 38
16	4.8						3.3	Mitsubishi	39
16	4.8						4	Toshiba	40
64	2					35	3.3	Matsushita	41
64	1.7			5	SiO_2/Si_3N_4	30	3.3	Mitsubishi	42
64	1.8						3.3	Fujitsu	43
64	1.53			5		34	3.3	Toshiba	44
64	1.28			3	Ta_2O_5			Hitachi	45
64	1.4					30	3.3	Toshiba	46
64	1.5			10	$SiO_2/Si_3N_4/SiO_2$	47	3.3	IBM	47
64	1.53		3.6	5	$SiO_2/Si_3N_4/SiO_2$	25		OKI	48
64	1.62					40	3	NEC	49

Table 2B. Parameters related to charge capacity of the memory cell capacitor for 16 and 64 Mbit DRAMs.

Mbit	Ac [um2]	As/Ac	As [um2]	Tox [nm]	Dielectric Material	Cs [fF]	Vdd [V]	Manufacturer	Reference
256	0.72			5	SiO_2/Si_3N_4	30	2	NEC	7, 50
256	0.5			5				Toshiba	8
256	0.72			5.3	$SiO_2/Si_3N_4/SiO_2$	25		Matsushita	51
256	0.75					32		SamSung	52
256	0.72		0.29	3	Ta_2O_5	25		Hitachi	53
256				2.5	Ta_2O_5	40		NEC	54
256	0.9			5	$SiO_2/Si_3N_4/SiO_2$	20		OKI	48
256			0.5	0.8	$(Ba_{0.5}Sr_{0.5})TiO_3$	20		NEC	55
1024	0.15							IBM	56
1024	0.05			5		15	2	IMEC	57
1024	0.2							Toshiba	8

Table 2C. Parameters related to charge capacity of the memory cell capacitor for 256 Mbit and 1 Gbit DRAMs.

A method to boost the storage capacity of the planar cell, prior to advent of trench cells, was the planar high capacitance (Hi-C) cell. An increased capacitance was obtained with the planar capacitor by implanting the storage region of a p substrate with a deep p-type implant and a shallow n-type implant [62]. This resulted in an increased depletion capacitance in the storage region, giving an increase in charge capacity of about 0.5 times and a leakage current reduction of about 3 times.

In the stacked capacitor cell (STC), the charge is stored on a stacked poly-poly or poly-metal capacitor which is formed on top of the field oxide or transfer transistor. Advantages of this cell, in addition to its reduced area, include a large reduction in leakage currents from the storage node due to the elimination of the depletion and inversion layer in the storage capacitor. This fact also increases the cell's immunity to SER. Additionally, this process is relatively easy to adapt for use with alternate dielectrics [56]. Disadvantages of this cell include a rather poor surface topography and a low dielectric breakdown [57].

As can be seen by a review of Table 1B the dominant type of vertical capacitor is the trench cell for 4 and 16 Mbit DRAMs. The trench capacitor or corrugated capacitor cell (CCC) shown in Figure 8 is an early version. Figure 9 shows a photograph of a cross-section through a Texas Instrument 1 Mbit trench cell of this type. The charge in this cell is stored in the bulk silicon surrounding the trench which results in the main disadvantages of this cell: high susceptibility to minority carrier leakage and a higher SER. Another problem with this cell is that the minimum distance between the trench storage node and the storage node or active devices in adjacent cells has to be relatively large due to the possibility of junction punch through. To alleviate these problems, an inverted trench cell or substrate-plate trench (SPT) cell was developed. This cell stores charge on the polysilicon inside the trench and uses the bulk silicon as the other capacitor plate electrode. Since the storage area is inside the isolated trench, the distance between adjacent trenches can be reduced. Also, the minority carrier collection area is significantly reduced.

There are a large number of variations on the SPT cell. One of the trends in the continuing quest to reduce cell area has been the buried trench cell, which is based on the SPT concept. This cell is shown in Figure 10; it has a poly-filled trench that is located underneath the transfer transistor. Fabrication of this

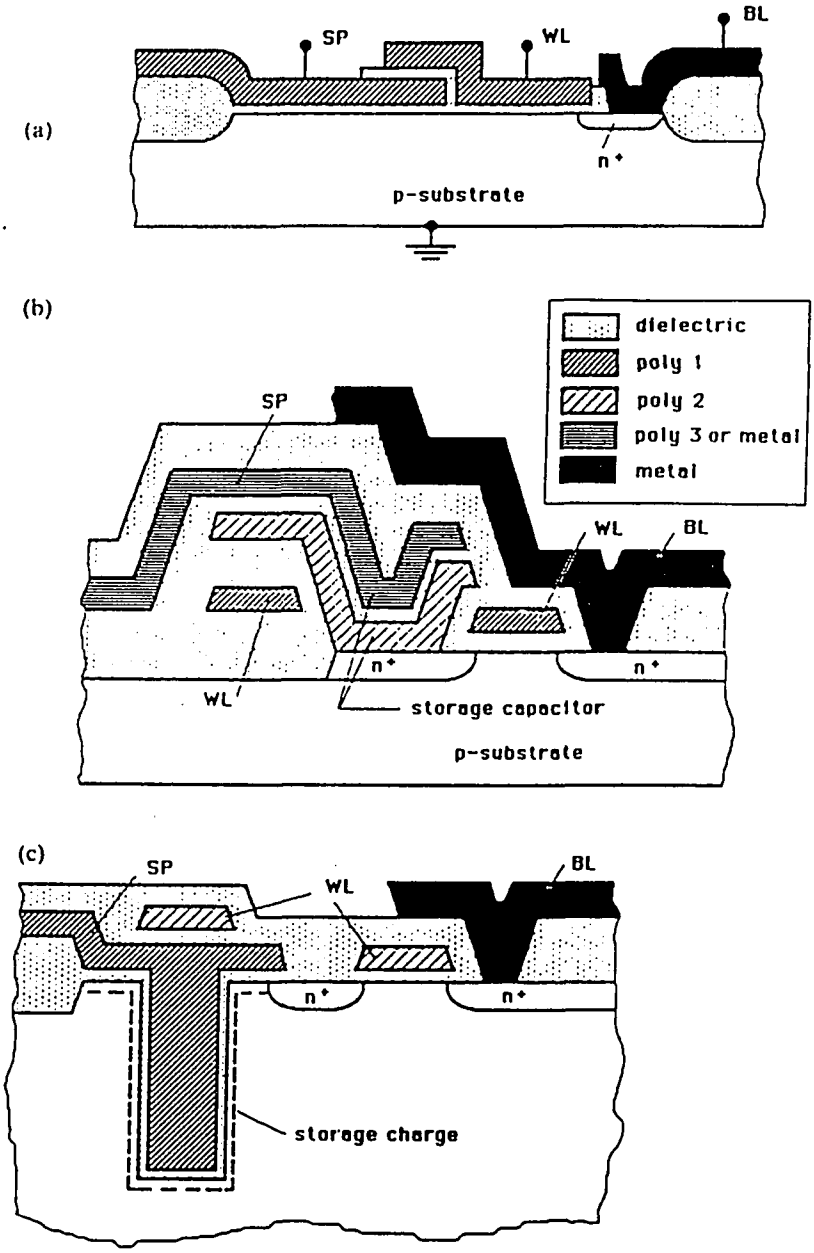


Figure 8. A comparison of (a) planar, (b) stacked capacitor cell (STC), and (c) trench capacitor memory cell (CCC) cross-sections (from reference [57]).

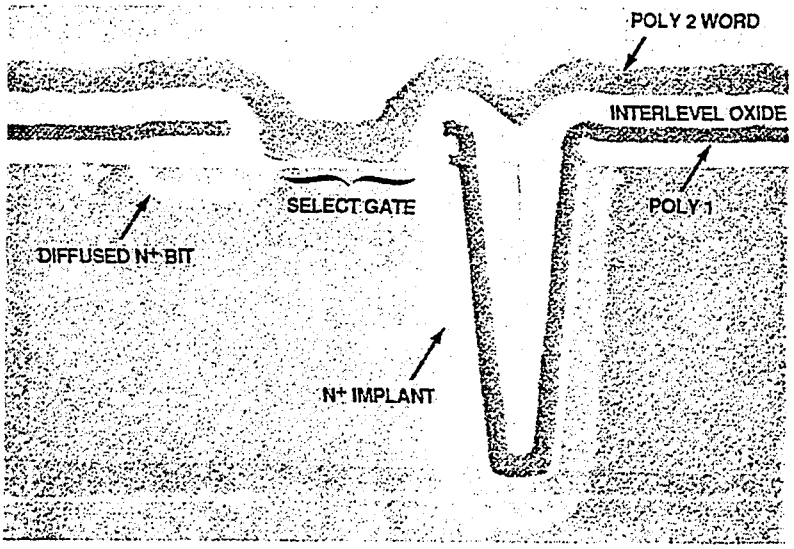


Figure 9. Photograph of a cross-section through a Texas Instruments 1 Mbit CCC cell.

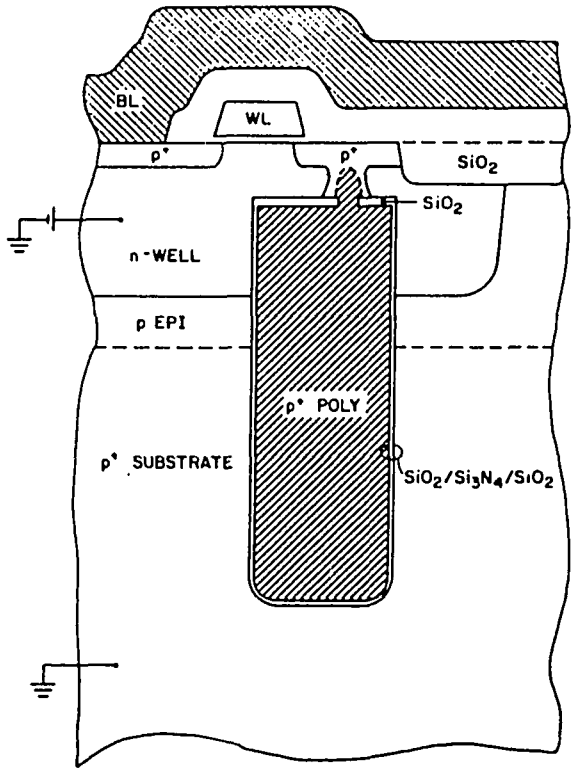


Figure 10. This cross-section is of a buried trench capacitor, a variation of the SPT cell from reference [56].

type of structure was made possible by a self-aligned epitaxy over trench (EOT) technology. A similar type of cell goes to the extent of building a vertical transfer transistor on the trench wall in a further attempt to save cell area; this cell is called the trench transistor cell (TTC). Figure 11 summarizes the basic architectural trends in types of trench cells discussed above and shows at what level of integration they have been employed.

Although trench cells were very popular for 4 and 16 Mbit DRAMs, almost all new types of memory cells now being reported are a variant of the stacked capacitor cell. Most researchers anticipate that the stacked capacitor cell architecture will be used in DRAM designs of 1 Gbit density [63]. A number of different types of advanced stacked capacitor cells are illustrated in Figure 12. As mentioned previously, the stacked capacitor cell has excellent SER immunity and significantly less leakage. Additionally, at higher levels of integration and the resulting reduced cell area, one serious problem with the trench type cell is that the area for the opening of the trench is also scaled down. This means the trench must be deeper to minimize the loss of trench capacitor surface area, and the smaller opening makes etching slower (decreasing throughput of processing of trench capacitors, since trenches are typically etched in single wafer tools).

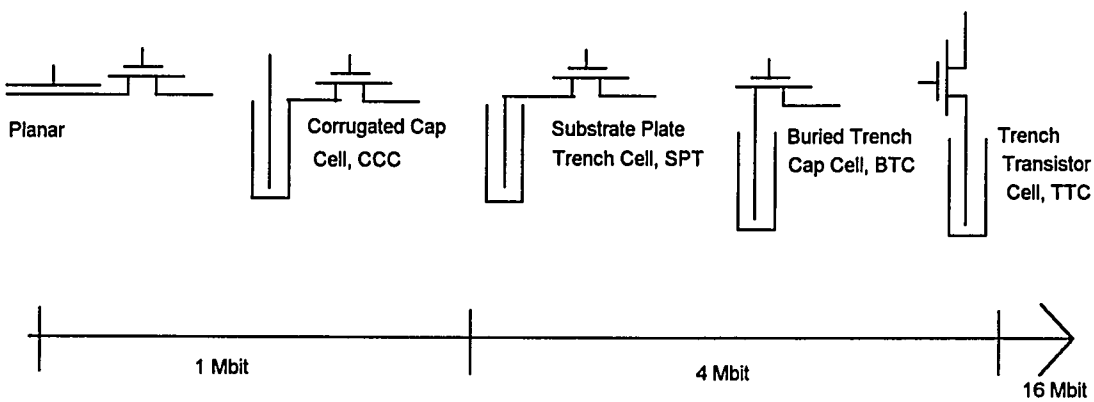


Figure 11. Trench cell architectural trends and the level of integration in which they were employed.

Another reason the stacked capacitor cell is becoming the dominant architecture is that 1 Gbit DRAMs will most probably have an alternate storage capacitor dielectric such as tantalum pentaoxide, Ta_2O_5; the stacked capacitor cell is particularly suitable for this type of process since the dielectric can be deposited between two poly layers without the dielectric contacting the bulk silicon. Based on these observations, the leakage requirements for the 1 Gbit DRAM can best be estimated by examining the leakage aspects of the stacked capacitor cell.

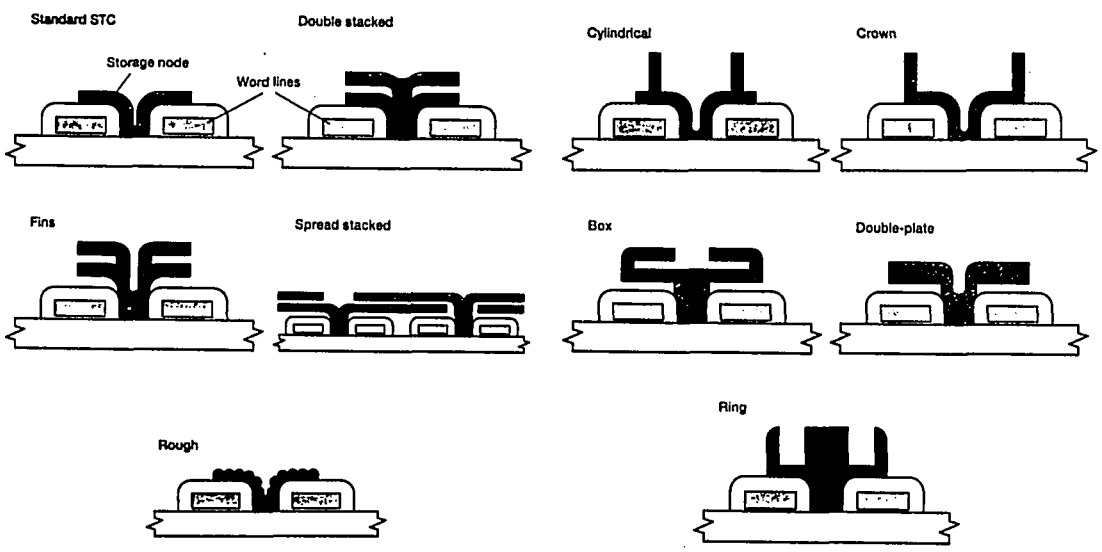


Figure 12. Various types of advanced stacked capacitor cells from reference [63]. The structure of the storage capacitor's bottom plate is illustrated; structural details of the capacitor's dielectric, upper plate, and passivation are omitted.

2.3 Leakage Current in DRAMs

As the devices have been scaled down for higher levels of integration, there have been two countering trends that have dramatic impact on the leakage from a DRAM cell. As was shown during the discussion of types of memory cells, there has been a definite inclination toward the stacked capacitor type cell as the 1 Gbit level of integration is approached, for reasons previously discussed. Inspection of Table 1 shows that all but one 256 Mbit cell is based on the stacked capacitor cell. Moving the capacitor away from the bulk silicon and into a dielectrically isolated position above the cell has significantly reduced the depletion area in the bulk silicon and resulted in a corresponding decrease in cell leakage. On the other hand, inspection of Table 2 shows that the capacitance of the cell has been reduced for the higher levels of integration, making the cell much less tolerant of leakage.

One of the intentions of the model to be presented in Chapter 3 is to estimate the amount of leakage that a DRAM cell can endure. The end product of the model is a value of leakage current based on various process and contamination parameters. The question that will be addressed here is how to relate the leakage value, produced by the model, to leakage constraints of advanced DRAM cells. To answer the question of how much leakage a DRAM cell can tolerate, and to understand the implications of cell architecture and memory cell charge storage capacity on leakage, it is first necessary to review what leakage occurs from a DRAM memory cell.

The leakage current from a storage node in a DRAM cell has a number of components. The total leakage from the cell can be expressed as:

$$I_{leak} = I_{gen} + I_{diff} + I_{sub} + I_{ox}, \tag{2-5}$$

where I_{gen} is the generation current in the depletion region in the silicon substrate under the capacitor plate, I_{diff} is the diffusion current in the silicon substrate, I_{sub} is the subthreshold current of the transfer gate MOSFET, and I_{ox} is the tunneling current through the capacitor dielectric [59]. These components of leakage are illustrated in Figure 13 for a planar cell.

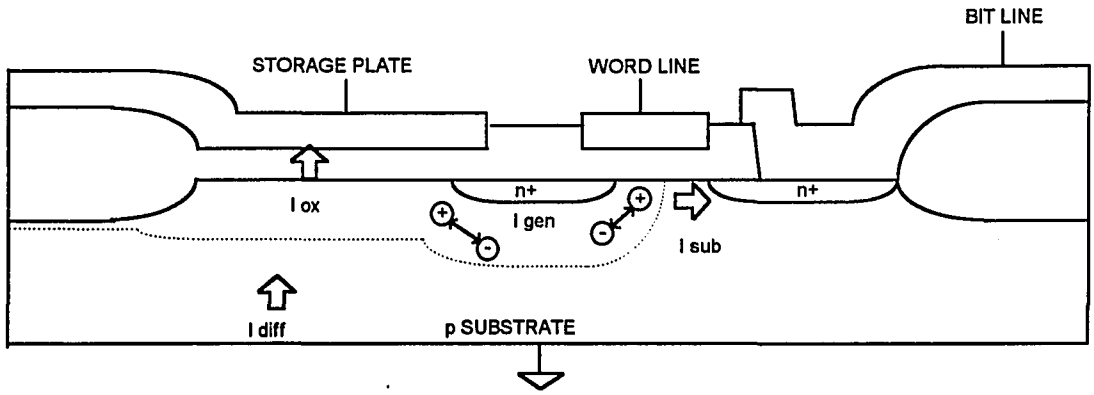


Figure 13. The components of leakage current from the memory cell storage node.

From reference [59], the generation current can be expressed as:

$$I_{gen} = \frac{q \cdot n_i}{2}\left(\frac{W}{\tau_{no}} A_s + S_o A_s + S_p A_p\right). \tag{2-6}$$

S_p is the surface recombination velocity at the field oxide region around the memory cell; A_p is the peripheral area around the memory capacitor; S_o is the surface recombination velocity at the memory node; A_s is the memory capacitor area; W is the depletion region width; τ_{no} is the minority carrier lifetime; n_i is the intrinsic carrier density; and q is the unit charge for a carrier. The diffusion leakage from the storage node can be written as:

$$I_{diff} = q\frac{n_i^2}{N_{ch}}\frac{L_n}{\tau_{no}}\left(A_s + \frac{\pi}{2}A_p\right), \tag{2-7}$$

where L_n is the electron diffusion length and N_{ch} is the channel dopant concentration. The peripheral area around the memory capacitor is defined as:

$$A_p = 4W_d \cdot A_s^{0.5}, \tag{2-8}$$

and the minority carrier lifetime as:

$$\tau_{no} = \frac{1}{v_n \sigma_n N_t}, \tag{2-9}$$

where σ_n is the electron capture cross-section, v_n is the average thermal velocity of an electron in the conduction band, and N_t is the trap density in silicon. The surface recombination velocity is:

$$S_o = \sigma_n v_n N_{ss}, \tag{2-10}$$

where N_{SS} is the surface trap density.

The second component of leakage from the memory cell, I_{sub}, the subthreshold current of the transfer gate transistor, can be expressed as:

$$I_{sub} = \frac{\mu_n Z C_o V_T^2 S}{L_{eff}} \exp\left[\frac{S}{V_T}(V_{GS} - V_{th}) + \frac{B}{V_T} V_{DS}\right], \tag{2-11}$$

where the effective channel length is:

$$L_{eff} = L_j - W_s - W_d, \tag{2-12}$$

the subthreshold constant S is:

$$S = \frac{C_o}{C_o + C_D + C_{it}}, \tag{2-13}$$

the drain-induced barrier lowering coefficient is:

$$B = \frac{\varepsilon_{si}}{\pi \cdot C_o L_j}, \tag{2-14}$$

the depletion layer capacitance per unit area is

$$C_D = \left[\frac{qN_{ch}\varepsilon_{si}}{2(\phi_s - V_{sub})}\right]^{\frac{1}{2}}, \tag{2-15}$$

and the fast state interface trap capacitance per unit area is:

$$C_{it} = qN_{it}, \tag{2-16}$$

where μ_n is the electron mobility, C_o is the gate oxide capacitance per unit area, V_T is the thermal voltage, Z is the channel width, L_j is the metallurgical channel length, W_s and W_d are the depletion widths at the source and drain, and ϕ_s is the surface potential of the transistor [59].

The final component of cell leakage is the tunneling current through the storage capacitor oxide film, I_{ox}. Assuming Fowler-Nordheim tunneling, this leakage is given by:

$$I_{ox} = v_{nox} \frac{\varepsilon_o}{V_T} \frac{\varepsilon_{ox}}{\varepsilon_{si}} E_{ox}^2 A_c \exp\left[-\frac{8\pi}{3h} (2m^* q)^{\frac{1}{2}} \frac{\phi_B^{\frac{3}{2}}}{E_{ox}} \left(1 - \left(1 - \frac{T_{ox} E_{ox}}{\phi_B} \right)^{\frac{3}{2}} \right) \right], \quad (2\text{-}17)$$

where E_{ox}, the electric field across the storage capacitor dielectric, is:

$$E_{ox} = \frac{V_s}{T_{ox}} \quad (2\text{-}18)$$

and the potential barrier between Si and SiO_2 is:

$$\phi_b = 3.2 - 2.59x10^{-4} E_{ox}^{\frac{1}{2}}, \quad (2\text{-}19)$$

where v_{nox} is the average thermal velocity of the electron in SiO_2, m* is the effective mass of the electron in SiO_2, and h is Planck's constant [59].

The question of interest is which of the above components of DRAM leakage is dominant. One group establishing a design methodology for submicron MOSFETs in DRAM applications [59] gave an indication of the relative importance of the components through their design rules. For the memory cell, the design rule reduced the storage capacitor oxide thickness until I_{ox} reached 1 percent of I_{cell}, the component due to generation and diffusion from the substrate. Another criterion which was used to find the ideal threshold voltage of the transfer MOSFET used the design rule of I_{sub}=0.01 I_{cell}. Use of these design rules would indicate that the dominant leakage in a DRAM cell is that due to generation in the depletion region and diffusion current from the substrate.

In support of the conclusion that generation within bulk silicon depletion region of a memory cell contributes the dominant component of leakage, a number of independent sources discovered this fact in their research. These sources examined leakage components in both trench and stacked capacitor cells. A group from Hitachi [33] examining leakage in a 16 Mbit BSC (trench) cell as a function of temperature found an activation energy of 0.57 eV. This value is very close to half the silicon band gap energy,

indicating that the leakage in this cell was primarily due to generation in the bulk silicon depletion region. In a similar type of study, another group of researchers [36] found an activation energy for a 16 Mbit stacked capacitor cell to be 0.6 eV, again indicating generation as the dominant leakage. A group from Texas Instruments [23] examining leakage in a TTC (trench) cell reported that leakage through the oxide from Fowler-Nordheim tunneling was insignificant, and that at normal storage node voltages, the main leakage is trap-assisted generation-recombination leakage from the space-charge region. Finally, as a more interesting offering of independent support, an engineer from IBM [64] performing a failure analysis on a DRAM which was exhibiting an unusual leakage phenomenon found that the charge from mobile ions in the isolation oxide between cells (at the Si-SiO_2 interface) was widening the space-charge region in the cell. The result was increased generation current and failure from excess leakage current.

The research presented above seems to indicate, even for advanced architecture cells, that the dominant leakage is due to generation in the bulk silicon. Therefore, a model that determines generation current in the bulk silicon as a function of process and contaminant parameters would be a useful tool for estimating leakage from DRAM memory cells.

Previously, it was mentioned that the DRAM memory cell must periodically be recharged or refreshed as a result of leakage from the cell. The refresh time interval is defined as the time interval between the beginnings of successive signals that are intended to restore the charge in the memory cell to its original level. This is the actual time interval between two refresh operations and is determined by the system in which the digital circuit operates. A maximum value is specified that is the longest interval for which correct operation of the memory cell is guaranteed. As stated previously, refresh time is a function of a number of DRAM parameters, including the number of wordlines (which affects the time to refresh all cells), the number of sense amplifiers (which has increased as the level of integration has increased), and most importantly, the leakage from the cell which is dependent on the process and cell architecture. It is desirable to keep the time needed to refresh all cells a small fraction of the refresh interval, allowing maximum time for the memory to respond to the CPU. The following example illustrates how long it takes to refresh every memory cell of a DRAM compared to the refresh time. Timing wave forms from

the data book of a NEC 1 Mbit DRAM (Model # PD421000) are shown in Figure 14. The time for a random read/write cycle , t_{RC}, is shown as the time from the row address strobe (RAS) going low to the next RAS high to low transition. In Figure 15, the refresh timing waveform for this same part is shown; the time interval required is again t_{RC}; also note that the data output is held in a high impedance state, preventing the data in the cell from appearing to the system. In a typical refresh scheme, the refresh counter increments through all the word line addresses, refreshing all bits. With this type of refresh, the time to refresh all cells is simply the product of t_{RC} and the number of wordlines. For the NEC 1 Mbit DRAM with a t_{RC} of 160 ns and 1024 word lines, the time to refresh all cells is 0.164 ms. This DRAM has an 8 ms refresh time, so the time to refresh all cells is approximately 2% of the refresh interval. This is fairly typical for this type of refresh scheme. There are a number of other refresh schemes in use (most DRAMs use multiple refresh schemes) that reduce this percentage further. So, from this point of view, the refresh interval is affected by the system constraint, to allow adequate memory performance to support the CPU. But at the device level, the actual limit is still determined by leakage from the cell.

From Table 1, it can be seen that the refresh time has increased by a factor of 2 for each generation; the 256 Mbit DRAM has a refresh time of 128 ms, and the 1 Gbit DRAM is expected to have a refresh period of 256 ms. Leakage from the cell is process and cell type dependent, but general design rules of thumb have been established to ensure an adequate safety margin that prevents the cell's charge degradation from causing a change of data state. Two examples are :

$$I_L = \frac{Q_c}{5 \cdot \tau_R} \tag{2-20}$$

by Tasch and Parker [5] and,

$$I_L = \frac{Q_c}{2 \cdot \tau_R} \tag{2-21}$$

by Lee et al [59]. Based on the refresh time, using the minimum reported charge capacity at each level of integration (from Tables 1 and 2), and correcting for room temperature operation, the leakage per cell can

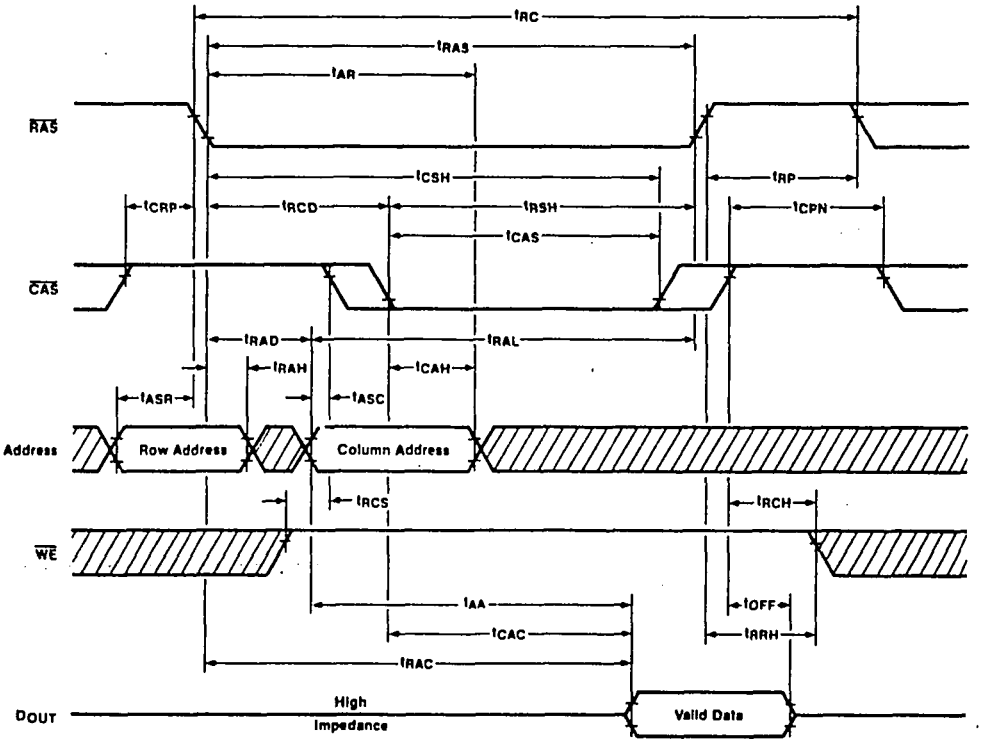


Figure 14. Read cycle waveforms from the NEC data book on 1 Mbit DRAM PD421000.

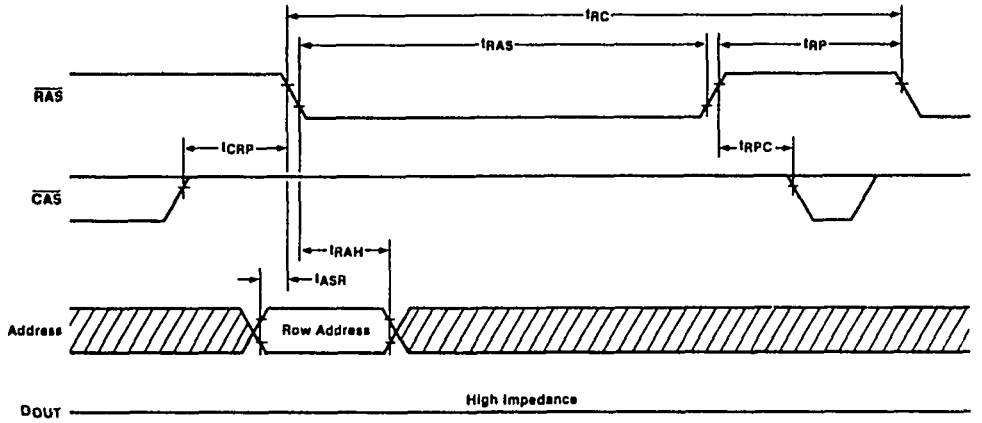


Figure 15. RAS-Only refresh cycle waveforms from the NEC data book on 1 Mbit DRAM PD421000.

be calculated. The leakage values from these calculations are shown in Figure 16. In order to obtain a

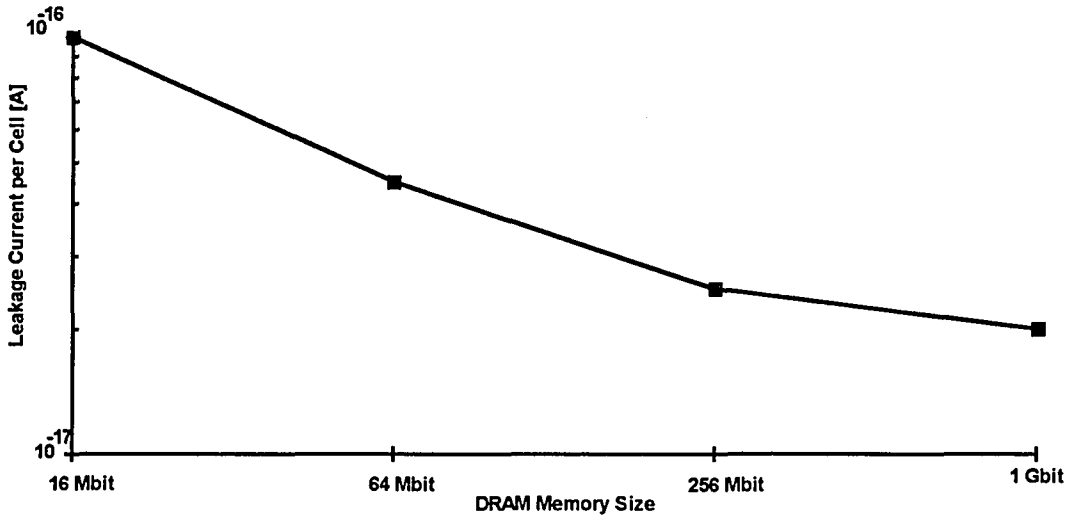


Figure 16. Leakage current per cell requirements for 16 Mbit through 1 Gbit.

current density value from leakage per cell, the leakage area of the cell must be estimated based on the cell architecture. Since most 64 and 256 Mbit DRAMs are stacked capacitor cells, and this is expected to be the architecture for all 1 Gbit cells, leakage from generation will be primarily related to the transfer transistor (the capacitor is above the transistor and isolated from the bulk). Leakage current density can be estimated simply by dividing the cell leakage by the cell area (which approximates the area of the transfer transistor). Trench cells, the dominant architecture for 16 Mbit and lower levels of integration, will have a leakage area that is approximated by the area of the storage capacitor. Estimates of the largest leakage area for each level of integration are shown in Figure 17. The worst case leakage current density values were calculated by dividing the minimum leakage per cell by the maximum cell area at that level of integration; these values are plotted in Figure 18. The startling observation that can be made from Figure 18 is that the leakage requirements after the 64 Mbit levels of integration become less stringent as a result of memory cell architecture trends which significantly reduce the space-charge volume.

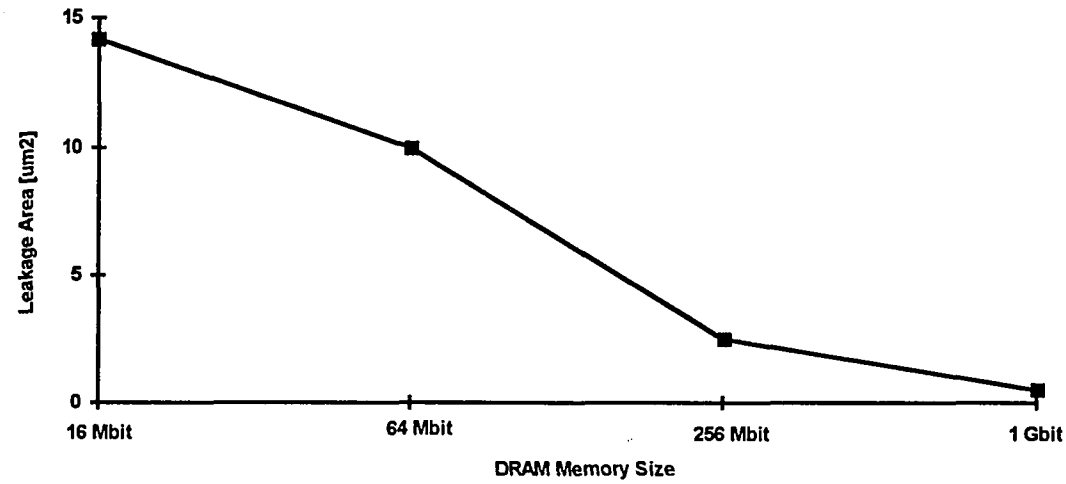


Figure 17. Estimates of leakage area for the memory cell based on dominant architecture at the respective levels of integration.

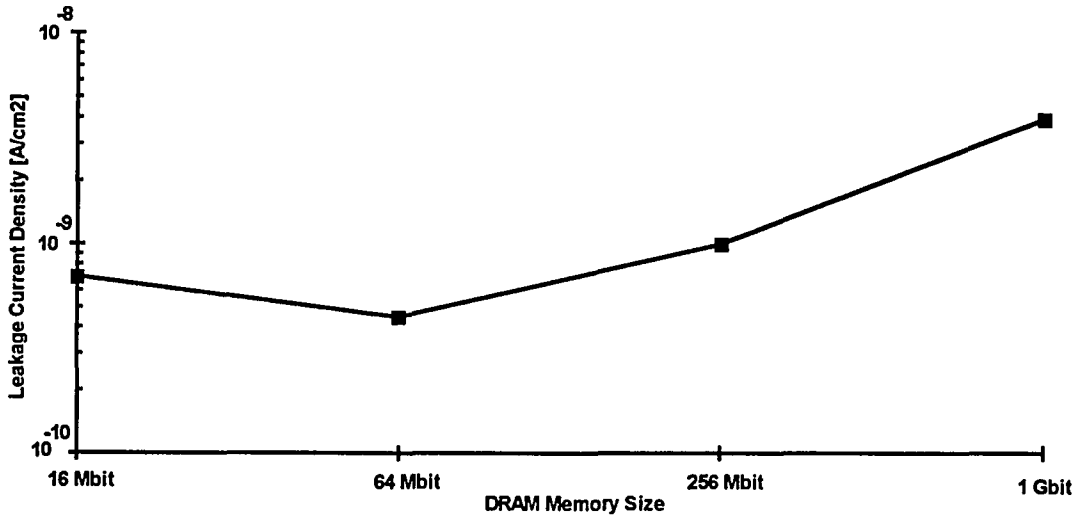


Figure 18. Leakage current density calculations for 16 Mbit through 1 Gbit.

One of the key performance parameters for advanced DRAMs is access time. There is continual design emphasis to reduce this parameter. For reduced access times, the transfer transistor has to have sufficiently high current drive capability to charge or discharge the memory cell capacitor in a length of time, τ_{on}, much shorter than the access time. If a cell has charge capacity Q_c, the current I_{on} required to pass through the transistor within time τ_{on} is:

$$I_{on} = \frac{Q_c}{\tau_{on}}, \tag{2-22}$$

where the time τ_{on} is a small fraction (~0.1) of the access time and represents the time in which the cell must charge or be discharged. Using equations 2-21 and 2-22, an expression can be written that relates leakage current to τ_{on}, the transistor on time:

$$\frac{I_{on}}{I_L} = \frac{5 \cdot \tau_R}{\tau_{on}}. \tag{2-23}$$

This relationship is of concern during design since, in efforts to minimize the leakage current across the transfer transistor, the required ratio of on-current to leakage current must be considered.

So, as was shown, leakage current is a very important parameter to the DRAM cell. It basically determines how long a memory cell capacitor will hold its charge and, for a multitude of reasons, it is desirable to have the cell hold its charge as long as possible.

2.4 Trends for Future DRAM Processes

There are certain characteristics that are highly desirable in a DRAM. These include minimum manufacturing complexity, high density, minimum SER, maximum charge capacity, maximum ratio of storage area to cell area, maximum dielectric constant of the storage capacitor, minimum leakage current, minimum noise, maximum speed, and low parasitic capacitance and resistance in the bit line. There are a variety of design efforts aimed at each of these areas. Trends in some of these areas have significant implications for leakage current from the cell.

One of the primary concerns from a leakage current point of view is the effort to ensure adequate charge capacity of the memory cell. With oxide thickness reaching a minimum thickness due to the physical barrier of tunneling, the capacitor area is also reaching a minimum to maintain the same charge capacity. One way to get around the area minimum is the use of alternate dielectrics with high dielectric constants. As can be seen in Table 2, barium strontium titanium trioxide $(Ba_{0.5}Sr_{0.5})TiO_3$, which has a dielectric constant of approximately 300, was used in a 256 Mbit cell to achieve an equivalent oxide thickness of 0.8 nm [55]. This means that the capacitor area can be reduced substantially and still obtain the necessary charge capacity. Another popular alternate dielectric, which has to this point received the most attention, is tantalum pentaoxide, Ta_2O_5. With a dielectric constant of approximately 25, 256 Mbit cells have been fabricated with an equivalent oxide thickness of 2.5 nm [54]. It has been preferable to use the stacked capacitor cell architecture with alternate dielectrics since, during the fabrication process, this dielectric deposition step is closer to the end of the process and the deposition is made between two polysilicon layers with no alternate dielectric contact to bulk silicon. One concern with alternate dielectrics has been a relatively high leakage current as compared to SiO_2. The $(Ba_{0.5}Sr_{0.5})TiO_3$ 256 Mbit DRAM cell was reported to have a leakage current of $5x10^{-16}$ A/bit; according to the preceding minimum leakage estimates for a 256 Mbit DRAM, this is slightly more than is acceptable.

One interesting way that has been used to enhance the area of the SiO_2 dielectric is to use a "rough capacitor". Hemispherical-grained-Si (HSG-Si) which, through a seeding technique, has its capacitor electrode surfaces covered with bumps or hemispheres, effectively increases the surface area of the electrode [50]. Using this technique has resulted in a 150-225% increase in the per-unit-area capacitance of conventional dielectrics [9]. Apparently, the leakage performance of this type of capacitor is comparable to more conventional capacitor designs.

One group from Toshiba has found a novel way to circumvent the minimum charge capacity requirements of the DRAM cell. They have added a very small area bipolar transistor to the cell to create a self-amplifying cell. A cell has been built that has a capacitance of only 15 fF and still provides adequate signal-to-noise ratio [65].

Further scaling of cell size in the submicron region is becoming more difficult because it is restricted by several physical phenomena. Avalanche breakdown of the drain junction, as a result of electric field crowding at the drain junction corners within the substrate and at the Si-SiO_2 interface (due to gate-proximity effect), degrades the transfer transistor. Bulk punchthrough between the depletion regions around the source and drain can occur; with short channel lengths, an uncontrolled space-charge limited current flows from source to drain. Short channel effects cause a degradation of the subthreshold characteristics by drain-induced barrier lowering. Finally, the hot electron effect results from a high electric field near the drain when the device is in saturation; this causes device degradation. These factors represent a significant barrier to further scaling of the cell size [59].

Based on these types of problems, in 1989, Tasch and Parker stated that the 256 Mbit DRAM may well represent the limit to DRAMs built using the capacitor charge storage concept. However, with significant process innovations, new cell architecture, and clever peripheral circuit design, it appears to a number of authors that a 1 Gbit DRAM based on capacitor charge storage will be realized.

2.5 Summary of DRAM Leakage Tolerances

It has been shown that new DRAM memory cell architectures are having dramatic effects on the leakage from the cell. The trend toward the use of variants of the stacked capacitor cell had significantly reduced the space charge volume of the cell and the associated generation current. Estimates of leakage current density tolerances at different levels of integration were made by first estimating leakage per cell which is a function of charge capacity and refresh interval. Next estimates of leakage area were made for each level of integration by identifying the dominant architecture at that integration level and then approximating the leakage area with either the storage capacitor area or the cell area for trench and stacked capacitor architectures respectively. Finally leakage current density for each level of integration is calculated using these estimates. These values are summarized in Table 3.

DRAM	Estimated Leakage per Cell [A/cell]	Estimated Leakage Area [μm^2]	Required Leakage Current Density [A/cm^2]
16 Mbit	1.0×10^{-16}	14.2	0.7×10^{-9}
64 Mbit	4.5×10^{-17}	10.0	0.5×10^{-9}
256 Mbit	2.5×10^{-17}	2.5	1.0×10^{-9}
1 Gbit	2.0×10^{-17}	0.5	3.9×10^{-9}

Table 3. Summary of estimated parameters used to calculate leakage current density tolerances for 16 Mbit through 1 Gbit DRAMs.

3. THE MODEL

A one-dimensional leakage current model based on Shockley-Read-Hall (SRH) deep-trap theory will be presented in this chapter. With this model, generation in the space-charge region and recombination in the undepleted bulk silicon is calculated as a function of metallic impurity species and concentration. The calculated generation is in turn used to estimate junction leakage current. The model assumes that metallic impurities are homogeneously distributed throughout the bulk silicon. The density of generation-recombination centers is assumed to be equal to the impurity density, and minority-carrier lifetimes are calculated using the trap density and experimental values of capture cross-sections. The model allows multiple contaminants and each contaminant may have multiple trapping levels in the silicon bandgap. In the absence of measurable contamination, minority carrier lifetime is limited by the process nominal lifetime due to impurities such as carbon, oxygen or trace metallics present in the starting materials or normally introduced during processing. The nominal process lifetime is reciprocally added to the contamination generation lifetime to obtain a total generation lifetime. As a result of the reciprocal summation, the shortest lifetime dominates; SRH lifetime dominates when metallic impurities are higher than the nominal process level. The nominal lifetime is the longest lifetime that can be expected for the material and process under normal conditions. The end product of the model is a leakage current density curve versus a metallic impurity concentration in the bulk silicon.

In the previous chapter the leakage current requirements for advanced DRAM processes were developed. With these requirements established, it can be seen that an easy method of determining actual leakage values in the bulk silicon as a function of basic process parameters would be useful for gauging process requirements for advanced MOSFET applications. The principle contributor to leakage in the bulk silicon is the generation of carriers within the space-charge region at trapping levels introduced in the silicon bandgap by metal contaminants. The one-dimensional model that calculates generation current is developed by first reviewing carrier interactions at these generation-recombination (G-R) centers and making use of elements of the Shockley-Read-Hall (SRH) deep-trap theory. Empirical capture cross-section values from the literature are applied in the model; where no empirical data are available,

estimates of capture cross-sections can be made based on certain experimentally derived models for these values. To better facilitate the translation of numerical leakage values to process concerns, a review of some of the basic processes and variables that determine the metal deposition and incorporation into the bulk silicon of the DRAM cell is presented. A demonstration of model use is given by estimating the metal deposition onto a wafer and subsequent distribution within the silicon; this example proceeds from a metal contaminant concentration in a process wet etch to the end product of the model, calculating leakage current density. Certainly a useful application of this model would be for determining what level of process contamination can be allowed, based on the constraints of the resulting device's desired leakage current performance.

3.1 Carrier Interactions at Generation-Recombination (G-R) Centers

The space-charge region in a reverse biased junction typically has the dimensions of on the order of 10^{-4} cm^2. This region contains generation-recombination (G-R) centers and rapidly changing populations of holes and electrons. With reverse bias, generation of carriers in the region leads to excess current above the saturation value predicted by the ideal-diode equations. This component of current in the space-charge region of a reverse-biased junction or a MOS capacitor in the deep-depletion region is dominant over the current component from the quasi-neutral region. Leakage current, such as that of concern in a DRAM memory cell, is primarily a result of this G-R center-dominated component in the space-charge region.

Certain impurity atoms, often trace transition metals (such as gold, copper or iron) found in the starting material or process chemicals, are incorporated into the silicon and introduce energy levels, E_T, into the silicon band gap (see Figure 19). Often this formation of a G-R center involves the decoration of crystal defects with the metal contaminant atom. It will be shown that G-R center densities, N_T, as low as 10^{12} cm^{-3} in silicon can significantly affect reverse biased junction current. These energy levels introduced within the silicon bandgap act as intermediate steps for carriers crossing the bandgap, and greatly increase

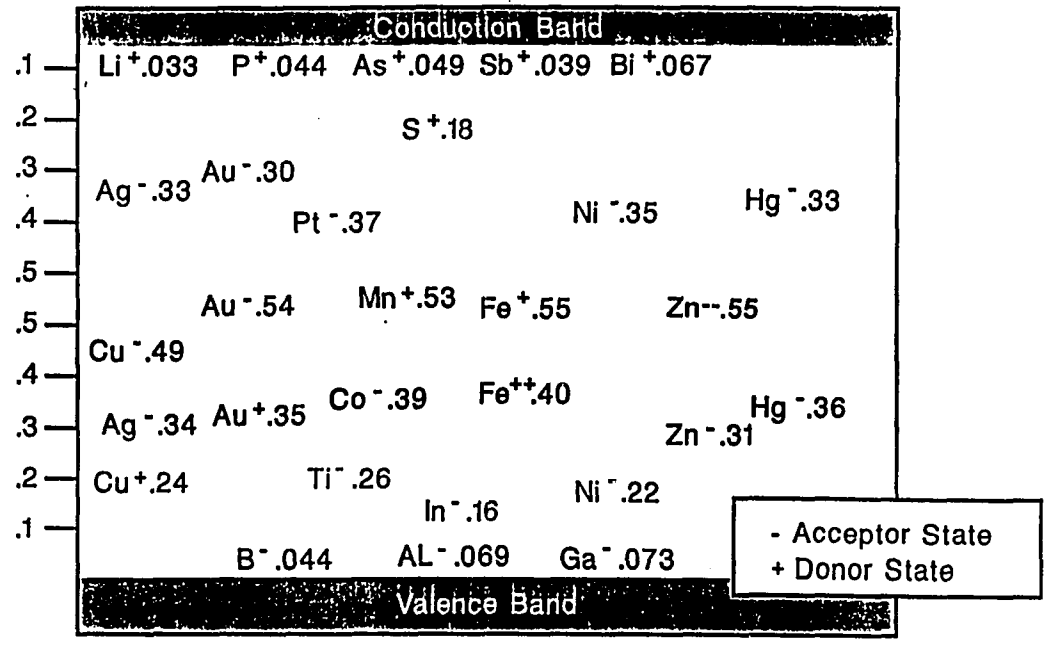


Figure 19. Energy levels introduced within the silicon bandgap by various metallic contaminants.

the probability of a generation or recombination event. Figure 20 illustrates the types of carrier interactions that occur at these G-R centers.

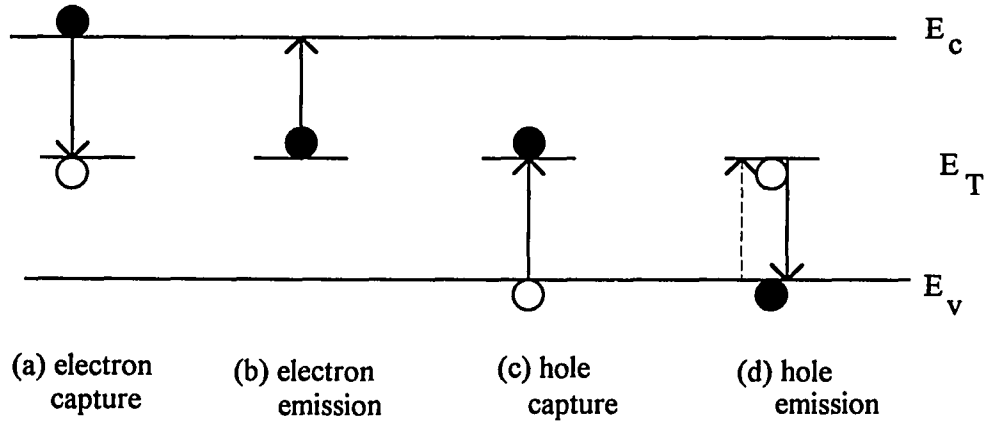


Figure 20. The types of carrier interactions with generation-recombination centers within the silicon energy bandgap. As shown in the figure, hole interactions can be interpreted as electron events. For example, a hole emission is equivalent to electron capture from the valence band.

When a carrier makes a transition from the valence or conduction band to the G-R center energy level, the event is called a capture; and when the carrier moves from the center level to the valence or conduction band, the event is called an emission. The capture of an electron is shown in Figure 20.a; this electron can then be emitted back to the conduction band (Figure 20.b). Also, a hole can be captured from the valence band (Figure 20.c) or emitted back to the conduction band (Figure 20.d). Recombination is usually followed by events (a) and (c), and generation by events (b) and (d). Trapping is a different event than G-R; trapping simply involves a capture of a hole or electron by the intermediate energy level and subsequent emission of that carrier. A G-R event involves the G-R center energy level and both the conduction and valence levels in the event.

The probability of a band-to-band G-R event is low in indirect-band semiconductors such as silicon. Indirect-band refers to the energy-momentum (E-k) plot and the relative position along the k axis of the valence band maximum energy to the conduction band minimum energy (see Figure 21). The fact that these points do not correspond in momenta means that a phonon (lattice vibration) assisted transition,

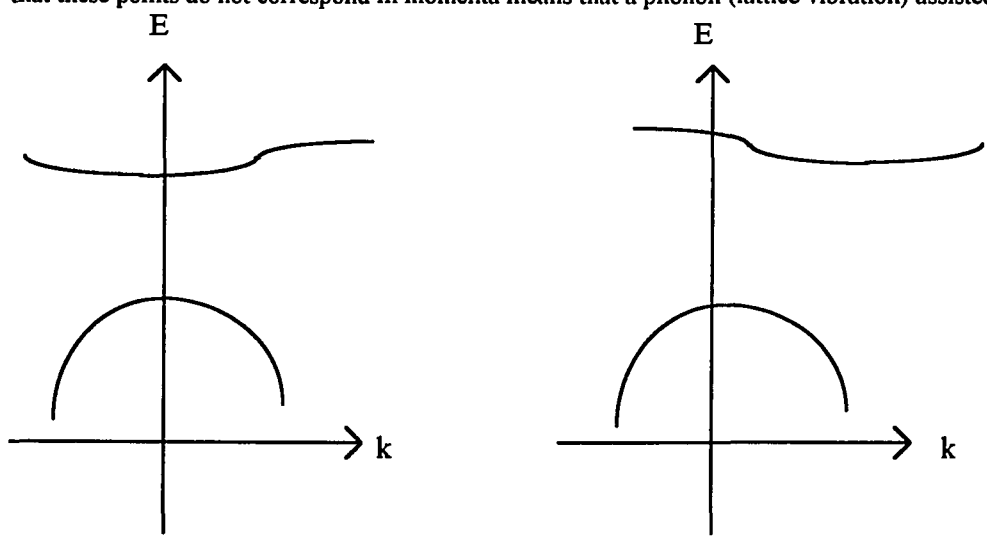


Figure 21. Energy-momentum (E-k) plots of direct and indirect semiconductors.

which is essentially horizontal, needs to occur before a carrier can make a band-to-band transition. Therefore, for a band-to-band generation-recombination event to proceed in an indirect semiconductor, a phonon must be emitted or absorbed. The additional need for a phonon interaction makes the probability of this type of band-to-band G-R event very low for indirect semiconductors; so, in silicon, generation-recombination is dominated by the G-R center-assisted events.

3.2 Analytical Development Involving SRH Theory

Now that the basic interactions of the carriers with G-R centers are defined, it is possible to develop the basic elements of the Shockley-Read-Hall (SRH) theory used to quantify generation current in the space charge-region. Using the continuity equations from the basic carrier interactions with G-R centers it can be seen that the rate of change of electrons and holes can be written as:

$$\frac{\partial n}{\partial t}\Big|_{G-R} = \frac{\partial n}{\partial t}\Big|_{capture} + \frac{\partial n}{\partial t}\Big|_{emission} \qquad \text{and} \qquad (3\text{-}1)$$

$$\frac{\partial p}{\partial t}\Big|_{G-R} = \frac{\partial p}{\partial t}\Big|_{capture} + \frac{\partial p}{\partial t}\Big|_{emission} \qquad (3\text{-}2)$$

The density of empty G-R centers is given by the total density of centers, N_T, times one minus the probability , $f(E_T)$, that the centers are occupied. Similarly, the density of full G-R states is the product of N_T and $f(E_T)$. Using these expressions to represent the density of G-R centers which are filled or empty, the above expressions can be rewritten as:

$$-\frac{\partial n}{\partial t}\Big|_{G-R} = n \cdot c_n[N_T(1-f(E_T))] - e_n[N_T(f(E_T))] \qquad \text{and} \qquad (3\text{-}3)$$

$$-\frac{\partial p}{\partial t}\Big|_{G-R} = p \cdot c_p[N_T(f(E_T))] - e_p[N_T(1-f(E_T))], \qquad (3\text{-}4)$$

where c_n, c_p, e_n and e_p are constants of proportionality, the capture and emission rates (with units of cm^3/sec) and whose physical significance will be discussed shortly. It is convention to write the net electron and hole recombination rates as positive if recombination is dominant, and negative if generation

is dominant. Equation 3-3 shows that the rate at which electrons will be captured by a G-R center is proportional to the number of electrons in the conduction band, the number of empty G-R centers, and the probability of capture of the electron by the center. At thermal equilibrium, the probability function can be replaced with the Fermi function $f_D(E)$; the rate of capture and emission can be set equal allowing one to write:

$$e_n = c_n \frac{n \cdot N_T(1-f(E_T))}{N_T(f(E_T))} = c_n n_i \exp\left(\frac{E_T - E_i}{kT}\right). \tag{3-5}$$

A similar expression can be obtain for the hole emission coefficient:

$$e_p = c_p \frac{p \cdot N_T(f(E_T))}{N_T(1-f(E_T))} = c_p n_i \exp\left(\frac{E_i - E_T}{kT}\right). \tag{3-6}$$

Equations 3-5 and 3-6 are simplified by substituting these commonly-used expressions for the electron and hole concentrations resulting from the G-R energy level:

$$n_t = n_i \exp\left(\frac{E_t - E_i}{kT}\right) \quad and \quad p_t = n_i \exp\left(\frac{E_i - E_t}{kT}\right). \tag{3-7}$$

Under low level injection in which, for instance, the hole population increases greatly while the electron population hardly changes, the rate of electron capture will have to exceed the rate of electron emission by only a very slight amount to increase the rate of hole capture. The net rate of electron capture (electron capture / electron emission) equals the net rate of hole capture. These net rates can simply be interpreted as the rate of recombination that is defined with the symbol U:

$$\begin{aligned} U &\equiv n \cdot c_n[N_T(1-f(E_T))] - e_n[N_T(f(E_T))] \\ &= p \cdot c_p[N_T(f(E_T))] - e_p[N_T(1-f(E_T))]. \end{aligned} \tag{3-8}$$

The capture coefficient is defined by:

$$c_n = v_n \sigma_n \quad and \quad c_p = v_p \sigma_p, \tag{3.9}$$

where v_n and v_p are the average thermal velocity of the carriers in their respective bands, and σ_n and σ_p are the capture cross-sections for electrons and holes of the G-R center. The capture cross-section concept will be discussed in more detail in the next section of this chapter. Solving equation 3.8 using equation 3.5, 3.6, 3.7 and 3.9 one can write:

$$U = \frac{N_T v_t \sigma_n \sigma_p (pn - n_i^2)}{\sigma_p \left[p + n_i \exp\left(\frac{E_i - E_T}{kT} \right) \right] + \sigma_n \left[n + n_i \exp\left(\frac{E_T - E_i}{kT} \right) \right]}. \tag{3-10}$$

The above expression can be simplified to the familiar SRH recombination equation by using the common definition for minority carrier lifetime or mean time to capture of the carriers, which is expressed as:

$$\tau_{no} = \frac{1}{v_n \sigma_n N_T} \quad and \quad \tau_{po} = \frac{1}{v_p \sigma_p N_T}. \tag{3-11}$$

The important SRH recombination equation is then written as:

$$U = \frac{pn - n_i^2}{\tau_{no}(p + p_t) + \tau_{po}(n + n_t)}. \tag{3-12}$$

The SRH recombination is further simplified for the case of primary interest, generation. In the space-charge region of a depleted MOS capacitor or a reverse biased junction, the carrier concentration is reduced so that $n << n_t$, $p << p_t$ and $np << n_t p_t = n_i^2$. The negligible terms are dropped from the recombination equation and the generation equation is obtained:

$$G = -U = \frac{n_i^2}{\tau_{po} n_t + \tau_{no} p_t}. \tag{3-13}$$

Recombination and generation lifetimes can be defined using equations 3-12 and 3-13. The time for one electron-hole pair (ehp) to recombine is the recombination lifetime, $\tau_R = \Delta n/U$, where Δn are the excess electrons. During recombination, for example when forward biased, there are excess carriers in the device, and $pn > n_i^2$. This results in the recombination lifetime being written as:

$$\tau_R = \frac{\tau_{po}(n_o + n_t) + \tau_{no}(p_o + p_t)}{n_o + p_o}, \tag{3-14}$$

where n_0 and p_0 are the equilibrium electron and hole concentrations respectively.

The generation lifetime represents the time required to generate one ehp in conditions where there is a shortage of carriers, i.e., in the space-charge region of a reverse-biased device. Using equation 3-13, generation lifetime is defined as:

$$\tau_G = \frac{n_i}{G} = \frac{\tau_{po} n_t + \tau_{no} p_t}{n_i}. \tag{3-15}$$

The relationship between recombination lifetime, τ_R, and the generation lifetime, τ_G, depends mainly on the ratio of capture cross-sections [66], as can be seen in equation 3-16:

$$\frac{\tau_G}{\tau_R} \approx 2\sqrt{\frac{\sigma_n}{\sigma_p}} \cosh\left[\left(E_T - E_i\right)/kT\right]. \tag{3-16}$$

So typically, τ_G is 50 to 1000 times larger than τ_R. Figures 20 and 21 show the relationship between τ_G and τ_R for different iron and copper impurity levels in the bulk silicon. Theoretically, an unknown contaminant could be identified by measuring the unique ratio of τ_G/τ_R. Some of these ratios were compiled by researchers from this group [67] (see Table 4).

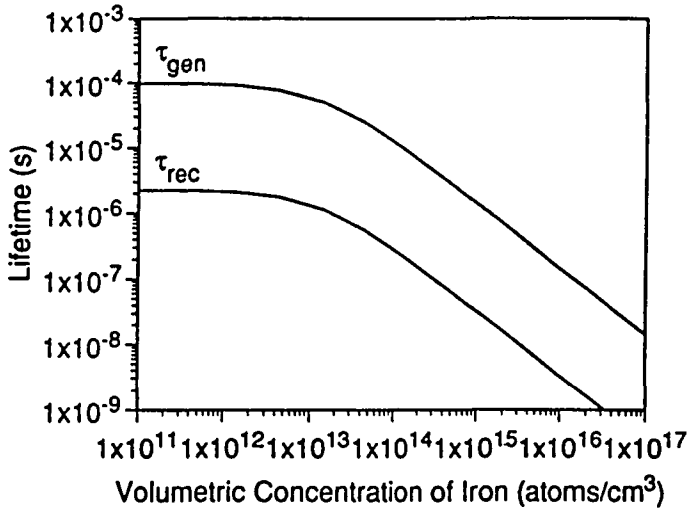


Figure 22. Generation lifetime, τ_G, and recombination lifetime, τ_R, as a function of iron concentration in the bulk silicon.

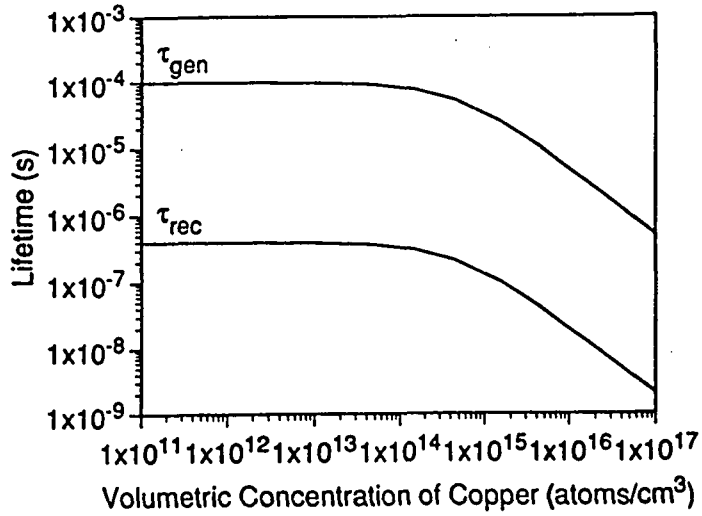


Figure 23. Generation lifetime, τ_G, and recombination lifetime, τ_R, as a function of copper concentration in the bulk silicon.

CONTAMINANT	τ_G / τ_R RATIO
Chromium	1750
Copper	210
Gold	15
Iron	43
Lead	937
Molybdenum	5930
Nickel	27
Silver	5950
Tin	33500
Titanium	97500
Tungsten	679
Zinc	1

Table 4. Characteristic generation-to-recombination lifetime ratios identify contaminants.

For a reverse-biased device, assuming that the generation current is the primary contributor to current across the reverse biased junction, the leakage current equation is:

$$J_L = -qG_{total}W_{gen}. \tag{3-17}$$

This is the desired end: leakage derived from process parameters and densities of contaminant metals in the bulk silicon. The following section explains how experimentally-determined metallic impurity parameters are used with the SRH theory just presented to arrive at leakage values as a function of

contaminant concentration.

3.3 Capture Cross-Section

Capture cross-section is a parameter that is better understood intuitively if initially approached from a real-space, conceptual point of view. There are a variety of complex physical models for capture cross-section in the literature [68, 69] that are beyond the scope of this thesis; however, the conceptual model illustrates the fundamental characteristics of this parameter, which is sufficient for this discussion. In this idealized view of the capture process, the unoccupied traps or empty G-R centers are modeled as fixed spheres randomly distributed throughout the semiconductor volume. The carrier, which is also represented to have a certain volume, moves through the unit volume of the semiconductor at its average thermal velocity. During the carrier's travel, if any of the volume of the carrier enters any of the volume of the fixed center, the carrier is captured. From Figure 24, the value of the capture cross-section can be represented as the area of $\sigma_n = \pi(r_n + r_t)^2$, where r_n is the radius of the carrier and r_t is the radius of the fixed center.

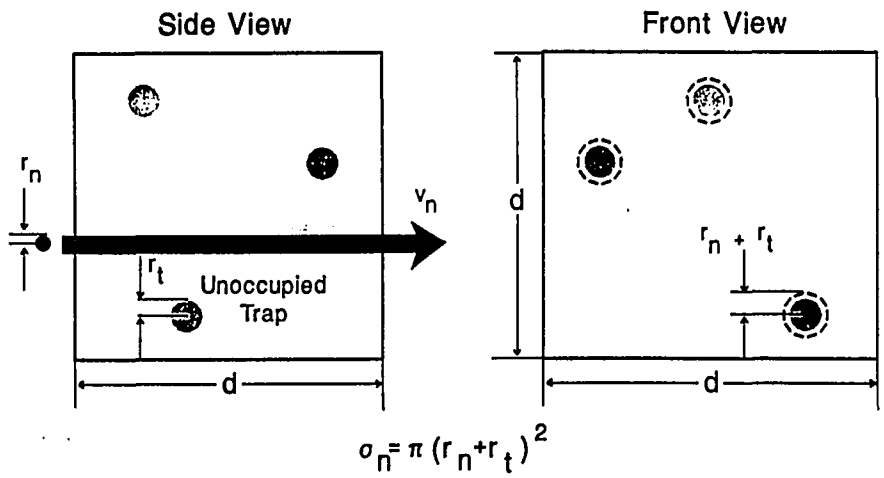


Figure 24. An idealized real-space conceptual model for capture cross-section.

Capture cross-section values have been obtained from experimental data and are compiled in Appendix D. For elements where experimental data were not available, capture cross-sections were estimated using gold as a reference element in Yassievich's model, which shows capture cross-section proportional to the trapping element's atomic number cubed, Z^3 [68]. Gold is used as a reference since this element's capture cross-section is the most studied G-R center in silicon and therefore there is considerable confidence is accuracy of this value.

3.4 Lifetime Summation

In a reverse biased device, generation lifetime can be obtained from Shockley-Read-Hall theory when it is assumed that contamination concentration is much less than the equilibrium majority carrier concentration, and that minority carrier injection is at a low level. The generation lifetimes are evaluated for each trapping level introduced by each contaminant, and summed reciprocally. The generation lifetimes equation is repeated here for convenience:

$$\tau_G = \frac{\tau_{po} n_t + \tau_{no} p_t}{n_i}. \tag{3-18}$$

A particular impurity may have multiple trapping levels in the silicon bandgap. Centers near mid-gap tend to be the most effective G-R locations [70, 71]. In order to maintain generality of the computational method, lifetimes are evaluated for each trapping energy level introduced by an impurity and then these lifetimes are summed reciprocally:

$$\frac{1}{\tau_{G_{sum}}} = \sum_{j=1}^{k} \frac{1}{\tau_{G_j}}, \tag{3-19}$$

where k is the number of trapping levels. The lifetime obtained in this manner represents a specific contaminant's effect at a certain background doping level in the bulk.

Depending on the starting materials and the cleanliness of the process, every process has a nominal lifetime. This lifetime can be thought of as resulting from normal process conditions, and is a

consequence of trace impurities in the starting wafer and the additional contamination encountered during processing. This lifetime is usually well known for an established process, and is defined here as τ_{nom}. As an example of a nominal process lifetime, one researcher [72] developed the empirical expressions given in equations 3-20 and 3-21 for lifetime in his process. The lifetime was found experimentally to be inversely proportional to the background doping level until doping concentration dropped below 10^{17} cm^{-3}, where the lifetime became essentially constant. The constant lifetime is considered the nominal lifetime or background minority carrier lifetime.

$$\text{For} \quad N \geq 3x10^{16} \quad : \quad \tau_G = 10^{(18.257 - 1.34 \log N)} \tag{3-20}$$

$$\text{For} \quad N < 3x10^{16} \quad : \quad \tau_{nom} = 1.5x10^{-4} \tag{3-21}$$

Other researchers have indirectly indicated their nominal lifetime by expressing minimum attainable leakage current density for a reverse biased junction. For example, one group attained leakage current densities of $2x10^{-10}$ A/cm^2 [73]. This nominal lifetime is added reciprocally to the generation lifetime calculated from contaminant deep traps, τ_{Gtot}, as:

$$\frac{1}{\tau_{G_{total}}} = \frac{1}{\tau_{G_{sum}}} + \frac{1}{\tau_{nom}}. \tag{3-22}$$

This result satisfies the empirical constraints that leakage current density increases with increasing background doping level, and that leakage current density tends toward a minimum for lightly doped substrates.

3.5 Example of Model Application

Based on the estimate that the next-generation DRAM requires a leakage current density of less than $3.9x10^{-9}$ A/cm^2, the simple bulk leakage model described above predicts that transition metal contamination must be below $1x10^{12}$ cm^{-3}. This is shown in Figure 25, which is a logarithmic plot of the leakage current density at 3.3V reverse bias (DRAM supply value for present advanced processes) as a

function of concentration of various metal contaminants. The elements used for these calculations were selected on the basis of their probability of incorporation into the Si wafer.

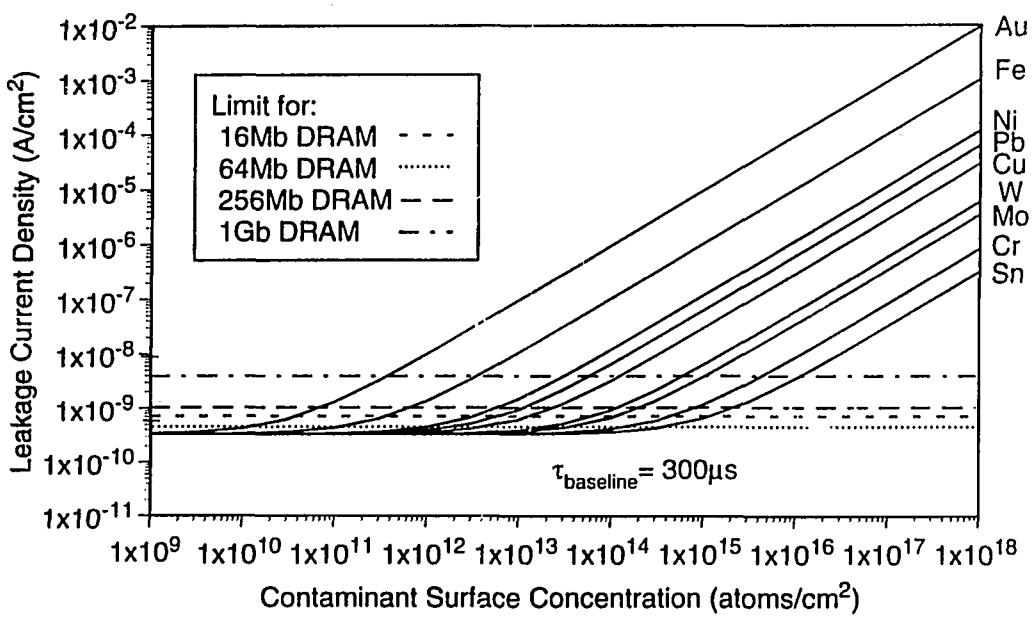


Figure 25. Leakage current density at 3.3V reverse bias versus contaminant concentration in bulk extrinsic silicon with N_A=1x10^{16} cm^{-3}.

Sources of transition metal contamination include cleaning solutions, etching solutions, photoresists, die metallization alloys and erosion of mechanical parts [74]. Photoresist ashing may result in photoresist contaminants being incorporated into the silicon wafer [75,76,77]. Transition metals such as iron, copper and lead appear at significant trace levels in both photoresist and $NH_4F/HF/H_2O$ (BOE) solutions. Other significant BOE contaminants are included in Table 5, which lists values from chemical supplier data sheets for semi-grade and ultra-pure HF and NH_4F. Molybdenum, tungsten and titanium are used for metallizations and silicides in advanced processes [78], and could potentially contaminate the bulk silicon.

Element	Semi-Grade HF [ppb]	Ultra-Pure HF [ppb]	Semi-Grade NH_4F [ppb]	Ultra-Pure NH_4F [ppb]
Aluminum	50	15	200	20
Calcium	200	30	200	30
Chromium	10	10	50	10
Copper	50	10	50	10
Gold	50	5	50	10
Iron	200	40	500	50
Lead	100	10	300	10
Nickel	100	10	100	10
Potassium	200	30	200	25
Sodium	300	40	300	50
Tin	100	10	200	20
Titanium	100	10	200	20
Zinc	100	10	100	10

Table 5. Trace contaminants in semi-grade and ultra-pure HF and NH_4F.

There are basically three possible mechanisms through which metal contaminants can be deposited from solution onto a silicon wafer:

1. Adsorption of the metal species without chemical change.
2. Adsorption of the metal species with chemical decomposition or precipitation.
3. Adsorption of the metal species with electrochemical reduction to either metal or to a partially reduced state.

Kern et al [79] have investigated the possibility of these deposition mechanisms with fundamental electrochemistry considerations. Their conclusions are that mechanism 1 is only probable in an oxidizing environment such as H_2O_2, or in a high concentration of dissolved O_2. Mechanism 2, although possible, is the most easily avoided. If a species is in solution, it will not spontaneously come out of solution unless there is a driving reaction with another chemical or impurity to do so. Hence, through solid process design, this mechanism can be prevented from occurring. Finally, it was concluded that for bare silicon wafers, mechanism 3 is the most probable way deposition of metallic contaminants from process solutions occurs. Since the contaminants are present on the wafer surface as elements, with an oxidation state of 0, and were in solution as ions, there must have been an oxidation-reduction (redox) reaction that took place in order for the deposition to occur. It has been further shown by Kern that deposition can be predicted using half-cell potentials and the Nernst equation. Deposition experiments by

the University of Arizona SEMATECH Task 3 group have confirmed that deposition can be predicted by oxidation potential [80, 81].

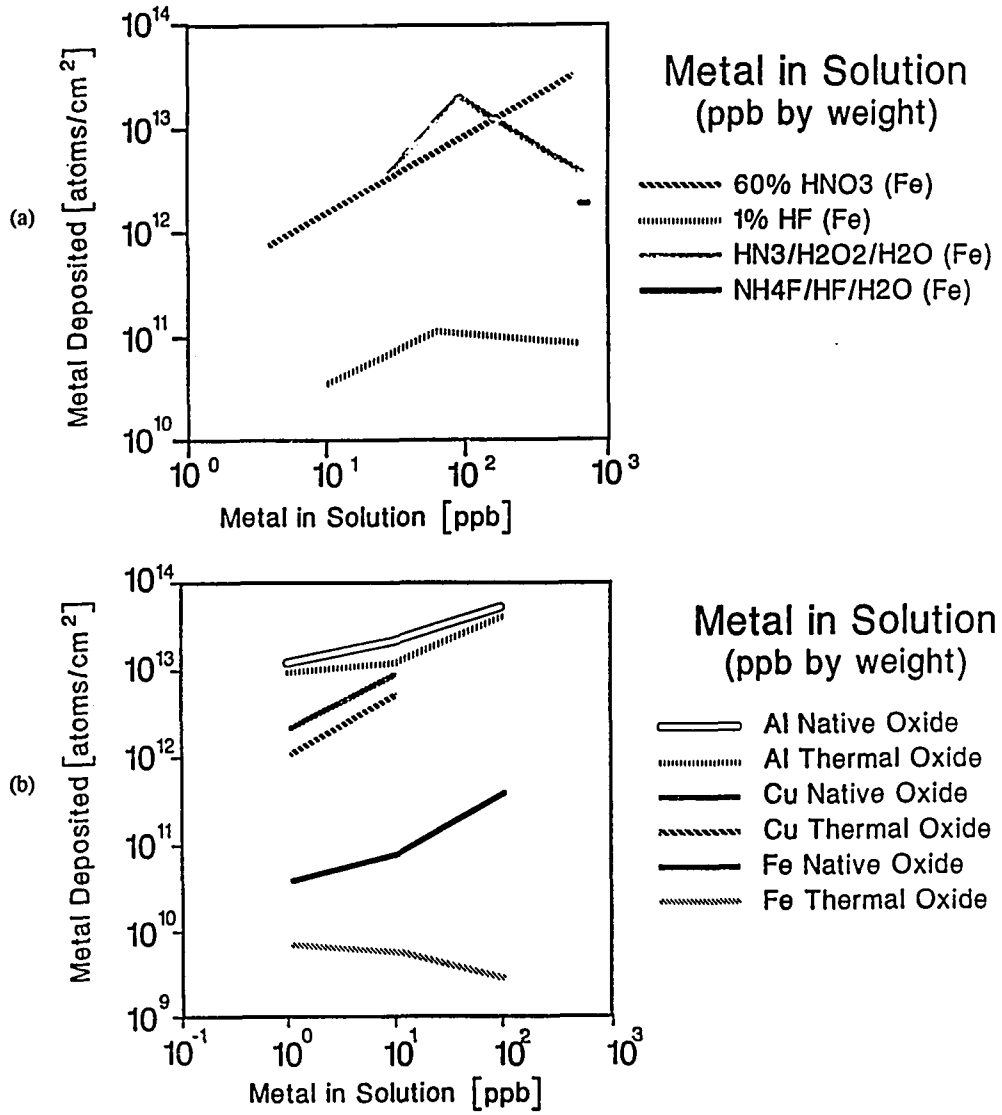


Figure 26. A summary of recently published results of deposition experiments of contaminants from various cleaning solutions onto a silicon wafer. (a) Looks at iron deposition from a variety of solutions [76, 82], (b) Shows that the deposition of aluminum, copper or iron also depends on whether the surface is native or thermal oxide [83].

There are a variety of studies that examine the transport of contaminants from the solution wet process baths to adsorption onto the wafer. Figure 26 shows some of the transport relationships found in

the literature, depicting the relationship between metallic impurities concentration in processing solutions to the subsequent contaminant concentration on the wafer surface. It is evident from the figure that the deposition process is highly sensitive to the species of metallic impurity, the stoichiometry of the cleaning solution, and the condition of the wafer surface. Also, the method of rinsing and drying following a clean can have a significant effect on the deposition results. Using data from various deposition experiments along with the leakage current model allows leakage current of junctions to be predicted, based on the concentration of metallic contamination in cleaning solutions.

Many currently used fabrication processes use BOE for wafer cleaning prior to the critical gate oxidation step. Of the components of BOE, NH_4F is by far the largest contributor of cation and metallic impurities. Transition metals in the BOE may become incorporated in the bulk silicon, in the gate oxide, or at the oxide-silicon interface. Experiments by this SEMATECH group [84] indicate that copper deposits on silicon surfaces from BOE more readily than iron or nickel. Figure 27 shows that the empirically determined relationship for Cu deposition on Si surfaces from BOE is:

$$Cu_W = 2.0x10^{11}(Cu_L)^{1.0}, \tag{3-23}$$

where Cu_W and Cu_L are the wafer surface copper concentration in atoms/cm^2 and the solution copper concentration in ppb respectively. This copper deposition data is in good agreement with other sources [82, 85].

Since the model predicts leakage based on a contaminant's concentration in the bulk silicon, in units of cm^{-3}, it is necessary to estimate the bulk concentration resulting from an areal or a deposition concentration of contaminant onto the wafer surface. As was discussed above, the deposition of metallic impurities onto the surface of the wafer often occurs during a particular processing step, for instance during a BOE etch. After the metallic impurity is deposited on the wafer's surface, subsequent processing involving heat can drive the impurity into the bulk silicon. The incorporation of an impurity into the bulk silicon is dependent on a number of factors including the post-deposition thermal processing and the species of contaminant, whether the impurity is an interstitial or substitutional diffuser. For fast diffusers,

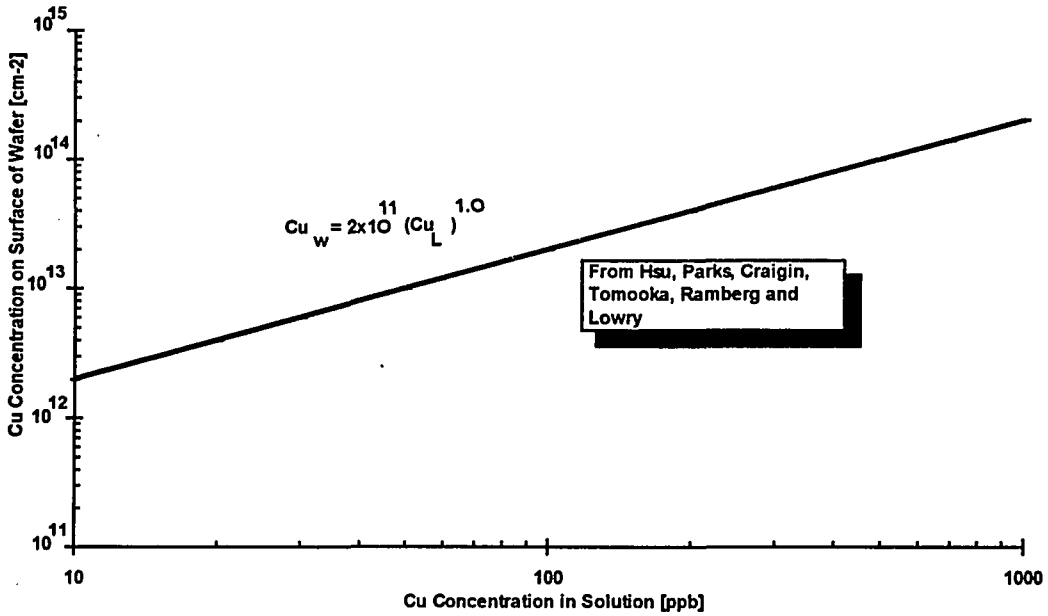


Figure 27. Copper deposition on Si surfaces from BOE, experimentally determined by the University of Arizona SEMATECH Task 3 group [85].

such as copper, which will diffuse through a 670μm wafer in less than one minute at 800°C [86], a simple relationship between surface and bulk concentrations can be established [87]. The bulk concentration, C (in cm^{-3}), is simply the surface or areal concentration, ϕ, resulting from the deposition divided by the wafer thickness, z; or, $C = \phi / z$. Using the data of Figure 27, and assuming Cu diffuses uniformly throughout the wafer thickness of 670 μm, the simple bulk leakage model produces a worst-case estimate of leakage current density as a function of Cu contamination in BOE; this is shown in Figure 28. Overlaid on the calculated leakage current density in Figure 28 are the DRAM leakage tolerances derived in Chapter 2. Using this figure it can be observed that processes for 64 Mbit DRAMs would probably require ultra-pure chemical reagents while processes for 1 Gbit DRAMs could possibly use semi-grade reagents.

While the relative effects of individual contaminants are clear from Figure 25, wafers in real processes will contain combinations of transition metal contaminants. Graphical presentation of the functional dependence of junction leakage current on a mixture of multiple contaminants would require a

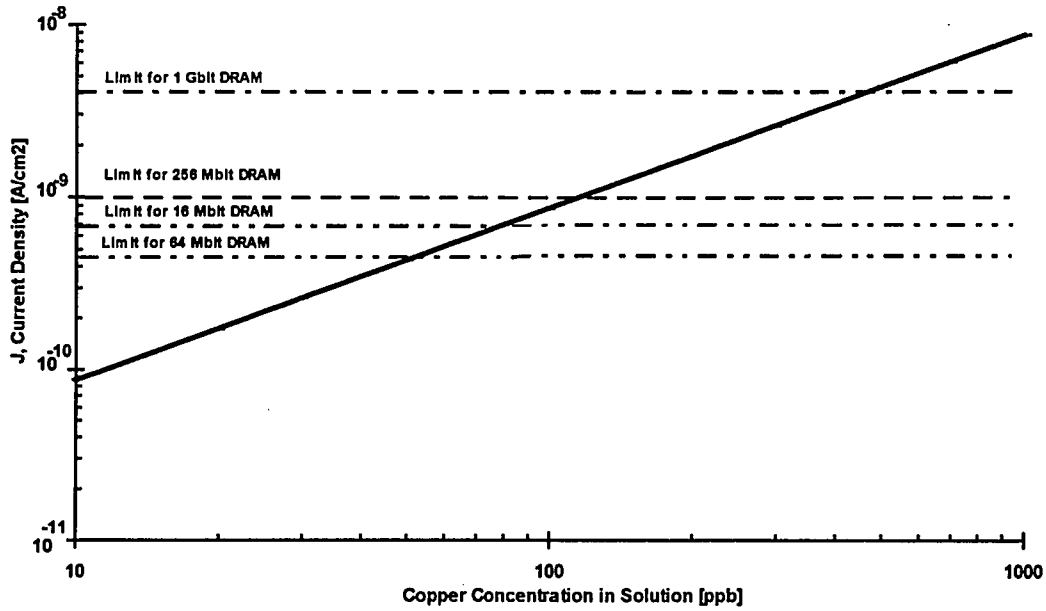


Figure 28. Leakage current density at 3.3V reverse bias vs. Cu concentration in BOE, with $N_A=1x10^{16}$ cm^{-3}.

multi-dimensional plot. Figures 27 and 28 show a bivariate calculation that can be illustrated stereometrically. The log of leakage current density as a function of copper and iron contaminants of various proportions is shown in Figure 29. Figure 30 shows the log of leakage current density as a function of titanium and tungsten concentrations. Titanium and tungsten may be important contaminants when used to fabricate silicide gates.

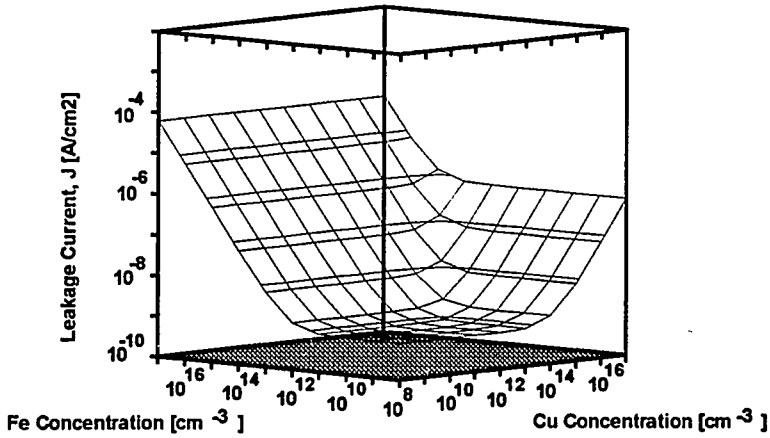


Figure 29 Relative effect of Fe and Cu in extrinsic bulk silicon on leakage current density. The reverse bias is 3.3V and N_A=1x10^{16}cm^{-3}.

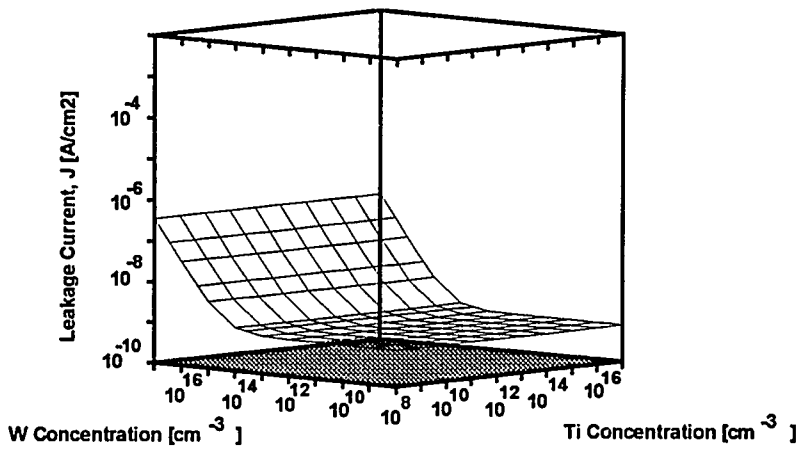


Figure 30. Relative effect of Ti and W in extrinsic bulk silicon on leakage current density. The reverse bias is 3.3 V and N_A=1x10^{16} cm^{-3}.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Now that a model which predicts leakage current of a reverse biased junction as a function of contaminant concentration has been developed, it is interesting to see if leakage values predicted by the model compare well to experimental leakage data taken from junctions having known metallic contamination levels. Controlled contamination experiments have been performed, by the SEMATECH Task 3 group at the University of Arizona, to examine the effects of homogeneous iron and copper contamination on the electrical properties of pn junctions and gate oxides. Most contamination-control efforts in integrated-circuit manufacturing have been directed toward identifying and eliminating particulates. However, as dimensions are reduced, homogeneous contamination in process chemicals is also an important source of yield loss and parametric degradation. In addition to enabling a determination of the accuracy of the leakage model, these controlled contamination experiments address a crucial two-part question: "can the use of ultra-clean chemicals enhance yield, and at which processing steps is their use mandated?" Contamination experiments that examine the effects of homogeneous contaminants on the electrical parameters of basic devices allow an insight into the answer of this question.

4.1 CMOS Process, Test Structures and Experiment

The test bed for the contamination experiment was a multi-purpose contamination monitor reticle with test structures designed jointly by Sandia National Laboratories (SNL) and the University of Arizona (UA). A description of SNL's CMOS process is given in Appendix F. The processing of the 1.25 μm CMOS wafers was performed at SNL in their Microelectronics Development Laboratory (MDL) facility.

All wafers in the experiment were processed together up to the pre-gate oxidation clean. At that point, the lot was split into child sublots to allow introduction of either iron or copper contaminant at a specific contamination level. Controlled quantities of iron or copper were introduced into the buffered oxide etch (BOE) used for the gate-oxidation pre-clean, which involved:

- Process in 5:1 Piranha at 95°C for 5 minutes.
- Rinse to 15 MΩ-cm resistivity in quick dump rinser.

- Dip in contaminated 30:1 BOE at 24°C for 1 minute, contamination level dictated by experimental design.
- Rinse to 15 MΩ-cm resistivity, cascade overflow, spin dry.
- Grow 18 nm gate oxide.
- Deposit 2KÅ of undoped polysilicon.
- Merge lot.

Ultra-clean chemicals served as a baseline for the experiment. The solutions' trace metallic make-up was determined by Inductively Coupled Plasma Mass Spectroscopy (ICP/MS). The etch bath was contaminated in a controlled manner by adding calculated aliquots of "doping" solutions. The doping solutions were 30:1 BOE solutions which contained known amounts of the desired contaminant. The first contaminated lot consisted of five splits, of three wafers each, contaminated with iron. The desired iron contamination levels were obtained from a stock solution prepared by dissolving 99.999% pure Fe_2O_3 in the 30:1 BOE. Iron levels in the BOE splits measured by ICP/MS were 17 ppb, 36 ppb, 106 ppb, 225 ppb and 480 ppb. The second contaminated lot used copper as the contaminant in three splits of three wafers each. Copper doping solutions were prepared by dissolving calculated quantities of 99.999% pure CuO in the 30:1 BOE solutions. Copper levels in the BOE splits measured by ICP/MS were <1 ppb, 8 ppb and 101 ppb. A third contamination lot also used copper, prepared in the same manner, with five splits of three wafers each at specified levels of 1, 10, 50, 100 and 500 ppb in solution. ICP/MS was also used to monitor 48 additional trace elements, giving an indication of the variability of the concentrations between the solutions.

All experimental splits were performed in a dedicated, filtered recirculating PVDF chemical bath. Prior to performing the experiments, the bath was drained and rinsed with DI water. A new 1 μm Millipore Teflon filter was installed, the bath was rinsed with DI water gain and drained, and a fresh 30:1 BOE bath was poured using General Chemical Class 10 ammonium fluoride and Ashland Chemical Megabit grade hydrofluoric acid. The bath was allowed to circulate through the filter overnight to assure low liquid-borne particle concentration. Whenever experimental design required that a fresh bath be poured, the bath was drained, rinsed with DI water, a new filter was installed, and the bath was rinsed with DI water a second time and drained.

All wafers were processed as a group up to the pre-gate oxidation clean with BOE. At this point, the wafers were divided into splits of three patterned wafers and one unpatterned monitor wafer (used for post-experiment analytical characterization) per split. The wafers were etched with their respective BOE solutions, then rinsed with DI water to 15 MΩ. A 18 nm gate oxide was grown, and each sublot was processed separately in the gate oxidation tube to reduce the risk of cross-contamination during oxide growth, after which the split lot was recombined for processing through first-level metal. Wafer position in the processing boats was randomized for pre-contamination processing, for separate processing of splits during the contamination steps, and for post-contamination processing; this minimized processing variation across the lot and reduced the affect of systematic errors. Analytical characterization of the unpatterned wafers was performed by Charles-Evans and Associates using TXRF and SIMS.

4.2 Wafer Characterization

Electrical testing of junction and capacitor test structures on the wafers was done with a HP4062C Parametric Test System and an Electroglas 1034X Automatic Wafer Prober (which included a temperature-controlled chuck). The system is well characterized and produces repeatable results at current levels in the pA range. Critical to measurements involving low currents was the ability to control the chuck temperature to an accuracy of better than 1°C. Additionally, all measurements were made in the dark and employed a guarding scheme to minimize parasitic current. The system is used to measure diode leakage current, activation energy, diode ideality factor, oxide leakage and tunneling current, oxide breakdown, capacitance-voltage (C-V), and capacitance-time transients (C-t). Data were transferred to a VAX for statistical analysis using the SAS software package.

The first contaminated lot consisted of five splits of three wafers each in BOE contaminated with iron at the pre-gate oxide clean. TXRF measurement of the surface concentration of iron on unpatterned test wafers processed with the lot splits showed no iron contamination above the noise level of the instrument. An alternate experiment using radio-tracer ^{59}Fe was conducted to quantify the expected

wafer contamination over this range of bath contamination levels. Resulting levels of iron surface concentrations ranged form $5x10^8$ atoms/cm^2 at 50 ppb Fe in solution to $7x10^9$ atoms/cm^2 at 470 ppb Fe in solution, confirming the lack of deposition of iron from BOE.

For the iron contaminated wafers, diode leakage measurements did not correlate with BOE iron concentration, as predicted by the deposition theory, further confirming the results of the deposition experiments. The mean reverse leakage current for 96 diodes at each contamination level was measured with a typical standard deviation of 0.015 nA.

Medium field breakdown of gate oxide MOS capacitor structures was measured by forcing 1.0 nA of current and measuring the voltage. Measured IV characteristics for baseline capacitors show an inherent breakdown of 13 MV/cm, indicating high-quality oxide, and a voltage drop of approximately 15V in the Fowler Nordheim conduction region at a forced current of 1.0 nA. Based on this criterion, capacitors for 17 ppb and 480 ppb BOE iron concentration splits showed 100 percent yield.

4.3 Experimental Results versus the Model

For the second lot, contaminated with copper, minority carrier generation lifetime and effective surface recombination velocity were measured using the Zerbst technique [88]. Figure 31 plots the mean lifetime of the splits versus the areal concentration of copper found on the wafer with TXRF. Mean lifetime of the pure BOE split was 17.2μs, compared to 15.0μs for the 10 ppb Cu split and 12.1μs for the 100 ppb Cu split. More than 100 devices were characterized from each lot, yielding 95 percent confidence intervals of about 1.3μs. No attributable effect on capacitor medium field breakdown by the copper split was observed, indicating that copper contamination did not degrade the gate oxides.

Figure 32 shows the median diode leakage current for Lot 2 wafers versus copper concentration in solution. Each data point represents the median of 147 diodes. The solid line in the figure was calculated by using the deep-trap model presented in Chapter 2. A nominal process lifetime of 115 μs was

assumed to enable fitting the nominal lifetime to the lifetime obtained from the lot split processed with the "pure" BOE.

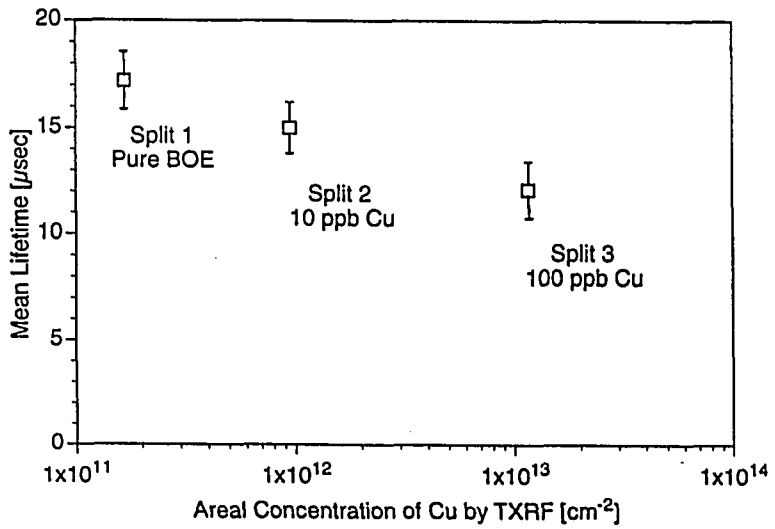


Figure 31. Areal concentration of copper (by TXRF) versus the mean generation lifetime.

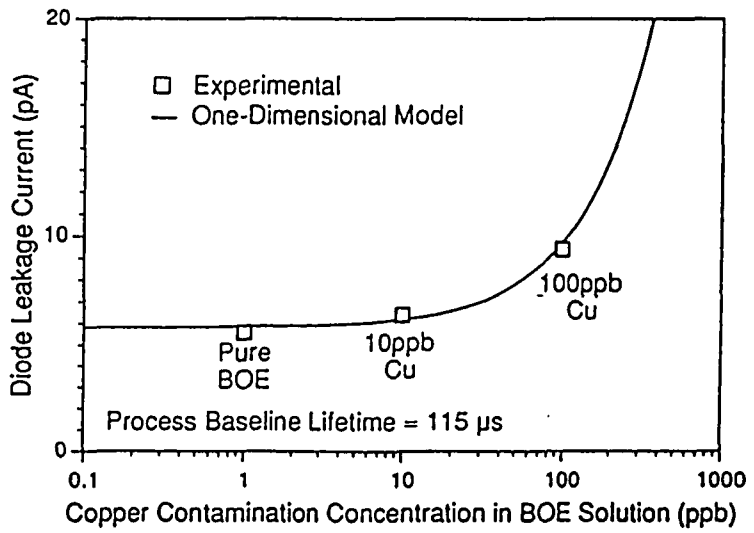


Figure 32. Reverse bias diode leakage current versus copper contamination level in the BOE; the solid line was calculated with the deep trap model.

It can be seen in the preceding figure that the leakage current for n^+p and pn^- diodes at low copper contamination levels corresponds to the nominal process lifetime. At a certain level, between 10

and 100 ppb copper, the contamination level begins to affect lifetime, and leakage increases with increasing copper contamination. The model is seen to fit nicely with the experimental data points. Members of the University of Arizona SEMATECH Task 3 group [67], in further work on the model, have defined a "critical concentration" which is a contaminant concentration level at the point where the lifetime changes from a baseline or nominal lifetime to the lifetime dominated by SRH center concentration (see Figure 33).

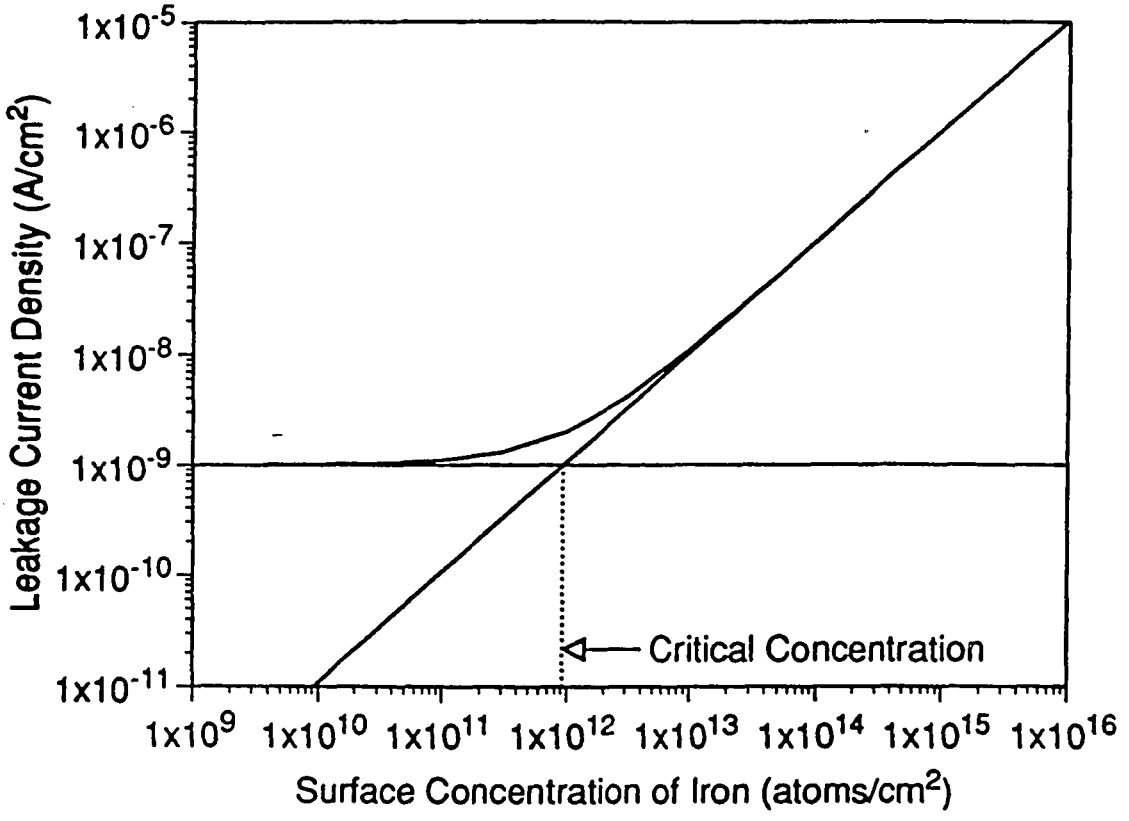


Figure 33. Surface concentration of a contaminant versus the resulting reverse bias leakage current density. The critical concentration is defined as a contaminant concentration level at the point where the lifetime changes from a baseline or nominal lifetime to the lifetime dominated by SRH center concentration.

5. SUMMARY AND CONCLUSIONS

DRAM CMOS processes, which produce a key component defining critical aspects of computer performance, are the subject of intensive design optimization efforts by major corporations. A variety of leading-edge research undertakings seek to improve many aspects of DRAM performance, one of which is the critical parameter, memory cell leakage current. Leakage from a memory cell basically determines how long a cell can hold a charge that represents the binary information. The storage capacitor's dielectric has reached thickness limits determined by physical phenomena. Additionally, the charge capacity of the storage capacitor has been close to minimum values for several generations of DRAMs, since noise margins have not allowed significant further reduction. This fact, coupled with the minimum oxide thickness, means the area of the storage capacitor is also near minimum value. Translated, this all means that the basic characteristic of the storage capacitor cannot scale significantly with further increases in the levels of integration. The refresh interval (the period between successive refresh operations) has doubled with each generation of DRAM, forced primarily by the system constraint of memory support time for the CPU. However, other critical aspects of refresh include the active power dissipation of the DRAM chip resulting from the need to drive more circuits (sense amplifiers) during the refresh operation. These requirements have forced a continued reduction in the leakage current a memory cell can tolerate. Fortunately, recent trends in cell architecture (the movement to the stacked capacitor cell) have relocated the capacitor into a dielectrically isolated region and significantly reduced the space-charge region within the memory cell. This results in a large reduction of the generation current, the major component of leakage from the memory cell. Based on these architectural trends and current 256 Mbit cell data, the leakage requirement for a 1 Gbit DRAM memory cell was estimated to be 1.95×10^{-17} A/bit or about 4×10^{-9} A/cm^2

The model for leakage current in a reverse biased junction as a function of contaminant concentration in the bulk silicon was developed from Shockley-Read-Hall deep-trap theory with application of empirical data on capture cross-section. Used in conjunction with various deposition studies in the literature which relate contaminant concentration in processing steps to the resulting

contaminant concentration on the wafer surface, it is possible to estimate, for instance, leakage as a function of a contaminant concentration in a process wet etch. This type of prediction would be extremely valuable for determining process cleanliness requirements.

Finally, controlled contamination experiments involving iron and copper were performed. The pre-gate oxidation clean of a 1.25μm CMOS process was intentionally contaminated with iron or copper. The wafers were fabricated at Sandia National Laboratory's Microelectronics Development Laboratory facility using a reticle of test structures developed jointly by the University of Arizona and Sandia National Laboratory. It was found that iron does not deposit significantly onto the wafer surface from a BOE solution. Copper contamination was found to deposit from BOE by the relationship $Cu_W=2x10^{11}Cu_{BOE}$. Subsequent leakage measurements of reverse biased diodes contaminated with different levels of copper found good agreement between measured values and those predicted by the model.

A crucial question that needs to be addressed is whether the use of ultra-clean chemicals can enhance yield and, if so, at which processing steps their use is mandated. The model presented is a useful tool to gage the results of homogenous contamination experiments directed at answering this question. Based on the results presented in Figure 25, an estimate of an upper bound for a metal contaminant concentration can be made for advanced processes. Contaminant concentration levels obtained from this plot demonstrate that, in addition to present yield-reducing particulates, homogeneous transition metal contamination must also be more tightly controlled in next-generation devices.

APPENDIX A. LIST OF SYMBOLS

ε_o	Permittivity in a vacuum.
ε_s	Permittivity constant for silicon.
ε_{ox}	Permittivity constant for silicon dioxide.
ϕ	Areal density.
ϕ_s	Surface potential for transfer transistor.
ϕ_B	Potential barrier between Si and SiO_2.
η	Array efficiency
μ_n	Mobility of an electron.
σ_n	Capture cross-section of G-R center for electrons.
σ_p	Capture cross-section of G-R center for holes.
τ_{no}	Minority carrier (electron) lifetime.
τ_{po}	Minority carrier (hole) lifetime.
τ_R	Recombination lifetime.
τ_G	Generation lifetime.
A_c	Area of the memory cell.
A_d	Area of the die.
A_p	Peripheral area of memory cell capacitor
A_s	Area of the memory cell capacitor.
B	Drain-induced barrier lowering coefficient.
C	Volumetric density.
C_{BL}	Bit-line capacitance
C_D	Depletion layer capacitance per unit area.
C_{it}	Fast state interface trap capacitance per unit area.
C_o	Transfer transistor gate dielectric capacitance per unit area.
C_s	Total capacitance of memory cell storage capacitor.
E_f	Fermi energy level.
E_i	Intrinsic energy level.
E_{ox}	Electric field across storage capacitor
E_t	Energy level of G-R center.
F	Minimum feature size.
G	Generation rate.
h	Planck's constant.
I_{diff}	Leakage current due to bulk diffusion
I_{gen}	Leakage current due to carrier generation.
I_{leak}	Total leakage current.
I_{ox}	Tunnel current through storage capacitor dielectric.
I_{sub}	Transfer transistor's subthreshold conduction current.
k	Boltzmann's constant.
L_j	Metallurgical channel length.
L_{eff}	Effective channel length.
L_n	Electron diffusion length.
m^*_n	Effective mass of an electron in the conduction band.
m^*_p	Effective mass of a hole in the valence band.
n	Electron concentration.
N_{it}	Density of fast state interface traps.
N_{ch}	Channel dopant concentration.

APPENDIX A. LIST OF SYMBOLS CONTINUED

n_i	Intrinsic carrier density.
N_{ss}	Surface trap density.
n_t	Excess electrons due to G-R centers.
N_T	Generation-recombination (G-R) center density.
p	Hole concentration.
p_t	Excess holes due to G-R centers.
q	Electronic charge.
Q_c	Storage charge.
S	Subthreshold constant.
S_o	Surface recombination velocity at memory cell.
S_p	Surface recombination velocity at the field oxide region around memory cell.
T	Temperature °K.
t_{RC}	Random read/write cycle time.
T_{ox}	Storage capacitor dielectric thickness.
T_r	Refresh interval.
U	Recombination rate.
Vdd	Supply voltage.
v_n	Average thermal velocity of an electron in the conduction band.
v_{nox}	Average thermal velocity of an electrons in SiO_2.
v_p	Average thermal velocity of a hole in the valence band.
V_s	Voltage at storage node.
V_T	Thermal voltage (kT/q).
W_d	Depletion layer thickness of drain junction.
W_s	Depletion layer thickness of source junction.
Z	Channel width or wafer thickness.

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL

Enter Donor and Acceptor Concentrations:

$N_A := 1.1 \cdot 10^{17} \cdot cm^{-3}$ $N_D := 10^{19} \cdot cm^{-3}$

Enter reverse bias on juncion:

$V_a := -3.3 \cdot volt$

Enter number for contaminant:
(Ag=0, Au=1, Cr=2, Cu=3, Fe=4, Mo=5, Ni=6, Pb=7, Sn=8, Ti=9, W=10, Zn=11) $i := 3$

Minority carrier concentrations:

$n_o := \frac{n_i^2}{N_A}$ $p_o := \frac{n_i^2}{N_D}$

$n_o = 9.091 \cdot 10^2 \cdot cm^{-3}$ $p_o = 1 \cdot 10^1 \cdot cm^{-3}$

Contaminant Concentrations:

$$\underline{N_t} := \begin{bmatrix} 10^8 \\ 10^9 \\ 10^{10} \\ 10^{11} \\ 10^{12} \\ 10^{13} \\ 10^{14} \\ 10^{15} \\ 10^{16} \\ 10^{17} \end{bmatrix} \cdot cm^{-3}$$

Base units:

cm ≡ 1L
kg ≡ 1M
sec ≡ 1T
coul ≡ 1Q

$farad \equiv 10^{-4} \cdot \frac{coul^2 \cdot sec^2}{kg \cdot cm^2}$

$volt \equiv 10^4 \cdot kg \cdot \frac{cm^2}{coul \cdot sec^2}$

$joule \equiv 10^4 \cdot kg \cdot \frac{cm^2}{sec^2}$

$amp \equiv \frac{coul}{sec}$

Constants:

$n_i := 10^{10} \cdot cm^{-3}$

$E_i := .548 eV$

$kT := 0.026 eV$

$\varepsilon_o := 8.854 \cdot 10^{-14} \cdot \frac{farad}{cm}$

$\varepsilon_s := 11.8$

$q := 1.6 \cdot 10^{-19} \cdot coul$

$eV \equiv 1.6 \cdot 10^{-19} \cdot joule$

Capture Cross-sections:

$$\underline{\sigma_p} := \begin{bmatrix} 2.1 \cdot 10^{-16} \\ 10^{-15} \\ 2.8 \cdot 10^{-17} \\ 5 \cdot 10^{-14} \\ 7 \cdot 10^{-16} \\ 1.5 \cdot 10^{-16} \\ 4 \cdot 10^{-14} \\ 1.1 \cdot 10^{-15} \\ 2.5 \cdot 10^{-16} \\ 2.2 \cdot 10^{-17} \\ 8.2 \cdot 10^{-16} \\ 1 \cdot 10^{-15} \end{bmatrix} \cdot cm^2 \quad \underline{\sigma_n} := \begin{bmatrix} 7.3 \cdot 10^{-16} \\ 3.5 \cdot 10^{-15} \\ 9.8 \cdot 10^{-17} \\ 10^{-17} \\ 1.5 \cdot 10^{-15} \\ 3 \cdot 10^{-16} \\ 1.6 \cdot 10^{-16} \\ 3.9 \cdot 10^{-15} \\ 8.9 \cdot 10^{-16} \\ 7.6 \cdot 10^{-17} \\ 6 \cdot 10^{-17} \\ 10^{-20} \end{bmatrix} \cdot cm^2$

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL CONTINUED

Energy Levels introduced in the Silicon Band Gap:

$$Et := \begin{bmatrix} 0.33 & 0.76 & 0.0 & 0.0 & 0.0 \\ 0.29 & 0.49 & 0.58 & 0.0 & 0.0 \\ 0.71 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.24 & 0.49 & 0.0 & 0.0 & 0.0 \\ 0.4 & 0.61 & 0.98 & 0.0 & 0.0 \\ 0.3 & 0.34 & 0.79 & 0.0 & 0.0 \\ 0.23 & 0.77 & 0.0 & 0.0 & 0.0 \\ 0.37 & 0.95 & 0.0 & 0.0 & 0.0 \\ 0.22 & 0.87 & 0.0 & 0.0 & 0.0 \\ 0.91 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.31 & 0.34 & 0.75 & 0.82 & 0.9 \\ 0.31 & 0.55 & 0.0 & 0.0 & 0.0 \end{bmatrix} \cdot eV$$

Junction built-in voltage:

$$V_{bi} := \left(\frac{kT}{q}\right) \cdot \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right) \qquad V_{bi} = 9.604 \cdot 10^{-1} \cdot volt$$

Depletion region width:

$$W := \sqrt{\frac{\left(2 \cdot \varepsilon_o \cdot \varepsilon_s\right) \cdot \left(V_{bi} - V_a\right) \cdot \left(N_A + N_D\right)}{\left(q \cdot N_A \cdot N_D\right)}} \qquad W = 2.261 \cdot 10^{-5} \cdot cm$$

Average thermal velocities of carriers:

$$V_{Tn} := \sqrt{\frac{3 \cdot kT}{m_n}} \qquad V_{Tn} = 1.116 \cdot 10^{7} \cdot \frac{cm}{sec}$$

$$V_{Tp} := \sqrt{\frac{3 \cdot kT}{m_p}} \qquad V_{Tp} = 1.524 \cdot 10^{7} \cdot \frac{cm}{sec}$$

Effective mass of carriers:

$$m_n := 1.002 \cdot 10^{-30} \cdot kg$$

$$m_p := 5.374 \cdot 10^{-31} \cdot kg$$

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL CONTINUED

Mean time to capture: $N := 9$ $j := 0..N$

$$\tau_{po_j} := \text{until}\left(N - j, \frac{1}{V_{Tp} \cdot \sigma_{p_i} \cdot N_{t_j}}\right) \qquad \tau_{no_j} := \text{until}\left[N - j, \frac{1}{\left(V_{Tn} \cdot \sigma_{n_i} \cdot N_{t_j}\right)}\right]$$

$$\tau_{po} = \begin{bmatrix} 1.313 \cdot 10^{-2} \\ 1.313 \cdot 10^{-3} \\ 1.313 \cdot 10^{-4} \\ 1.313 \cdot 10^{-5} \\ 1.313 \cdot 10^{-6} \\ 1.313 \cdot 10^{-7} \\ 1.313 \cdot 10^{-8} \\ 1.313 \cdot 10^{-9} \\ 1.313 \cdot 10^{-10} \\ 1.313 \cdot 10^{-11} \end{bmatrix} \cdot \text{sec} \qquad \tau_{no} = \begin{bmatrix} 8.961 \cdot 10^{1} \\ 8.961 \\ 8.961 \cdot 10^{-1} \\ 8.961 \cdot 10^{-2} \\ 8.961 \cdot 10^{-3} \\ 8.961 \cdot 10^{-4} \\ 8.961 \cdot 10^{-5} \\ 8.961 \cdot 10^{-6} \\ 8.961 \cdot 10^{-7} \\ 8.961 \cdot 10^{-8} \end{bmatrix} \cdot \text{sec}$$

Retrevie number of trapping levels for a particular contaminant:

$$a := \text{if}\left(Et_{i,1} > 0 \cdot eV, 1, 0\right) \quad b := \text{if}\left(Et_{i,2} > 0 \cdot eV, 2, a\right) \quad c := \text{if}\left(Et_{i,3} > 0 \cdot eV, 3, a\right) \quad e := \text{if}\left(Et_{i,4} > 0 \cdot eV, 4, a\right)$$

$$d := \begin{bmatrix} a \\ b \\ c \\ e \end{bmatrix}$$

$$f := \text{if}(\max(d) \geq 1, \max(d), 0)$$

Excess carriers resulting from the trapping levels:

$k := 0..f$

$$p_{t_k} := \text{until}\left(f - k, n_i \cdot \exp\left(\frac{E_i - Et_{i,k}}{kT}\right)\right)$$

$$n_{t_k} := \text{until}\left(f - k, n_i \cdot \exp\left(\frac{Et_{i,k} - E_i}{kT}\right)\right)$$

$$p_t = \begin{pmatrix} 1.395 \cdot 10^{15} \\ 9.307 \cdot 10^{10} \end{pmatrix} \cdot \text{cm}^{-3} \qquad n_t = \begin{pmatrix} 7.166 \cdot 10^{4} \\ 1.074 \cdot 10^{9} \end{pmatrix} \cdot \text{cm}^{-3}$$

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL CONTINUED

A lifetime is calculated for each energy level introducted:

$$\tau G_{j,k} := \text{until}\left(f - k, \frac{\tau \underline{po}_j \cdot \underline{n}_{\underline{t}_k} + \tau \underline{no}_j \cdot \underline{p}_{\underline{t}_k}}{n_i}\right)$$

$$\tau G = \begin{bmatrix} 1.250 \cdot 10^{7} & 8.340 \cdot 10^{2} \\ 1.250 \cdot 10^{6} & 8.340 \cdot 10^{1} \\ 1.250 \cdot 10^{5} & 8.340 \\ 1.250 \cdot 10^{4} & 8.340 \cdot 10^{-1} \\ 1.250 \cdot 10^{3} & 8.340 \cdot 10^{-2} \\ 1.250 \cdot 10^{2} & 8.340 \cdot 10^{-3} \\ 1.250 \cdot 10^{1} & 8.340 \cdot 10^{-4} \\ 1.250 & 8.340 \cdot 10^{-5} \\ 1.250 \cdot 10^{-1} & 8.340 \cdot 10^{-6} \\ 1.250 \cdot 10^{-2} & 8.340 \cdot 10^{-7} \end{bmatrix} \cdot \text{sec}$$

Reciprocal sum of lifetimes from trapping levels, the shortest lifetime dominants.

$$\tau \underline{Gsum}_j := \frac{1}{\sum_{k=0}^{f} \frac{1}{\tau G_{j,k}}} \cdot \text{sec}$$

$$\tau \underline{Gsum} = \begin{bmatrix} 8.339 \cdot 10^{2} \\ 8.339 \cdot 10^{1} \\ 8.339 \\ 8.339 \cdot 10^{-1} \\ 8.339 \cdot 10^{-2} \\ 8.339 \cdot 10^{-3} \\ 8.339 \cdot 10^{-4} \\ 8.339 \cdot 10^{-5} \\ 8.339 \cdot 10^{-6} \\ 8.339 \cdot 10^{-7} \end{bmatrix} \cdot \text{sec}$$

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL CONTINUED

Lifetime, T_{nom}, becomes constant with doping below 10^{17} cm^{-3} and low contaminant concentration:

$$v := \begin{pmatrix} N_A \\ N_D \end{pmatrix} \qquad \min(v) = 1.1 \cdot 10^{17} \cdot cm^{-3}$$

$$\tau_{nom} := \text{if}\left(\min(v) < 3 \cdot 10^{16} \cdot cm^{-3},\ 1.5 \cdot 10^{-4} \cdot sec,\ 10^{18.257 - 1.34 \cdot \log\left(\frac{\min(v)}{cm^{-3}}\right)} \cdot sec\right) \qquad \tau_{nom} = 2.64 \cdot 10^{-5} \cdot sec$$

Lifetime:

$$\tau_{\underline{Gtotal}_j} := \frac{\tau_{\underline{Gsum}_j} \cdot \tau_{nom}}{\tau_{nom} + \tau_{\underline{Gsum}_j}}$$

$$\underline{G}_{\underline{tot.}_j} := \frac{n_i}{\tau_{\underline{Gtotal}_j}} \qquad \underline{G}_{\underline{tot.}} = \begin{bmatrix} 3.788 \cdot 10^{14} \\ 3.788 \cdot 10^{14} \\ 3.789 \cdot 10^{14} \\ 3.789 \cdot 10^{14} \\ 3.79 \cdot 10^{14} \\ 3.8 \cdot 10^{14} \\ 3.908 \cdot 10^{14} \\ 4.988 \cdot 10^{14} \\ 1.578 \cdot 10^{15} \\ 1.237 \cdot 10^{16} \end{bmatrix} \cdot time^{-1} \cdot cm^{-3}$$

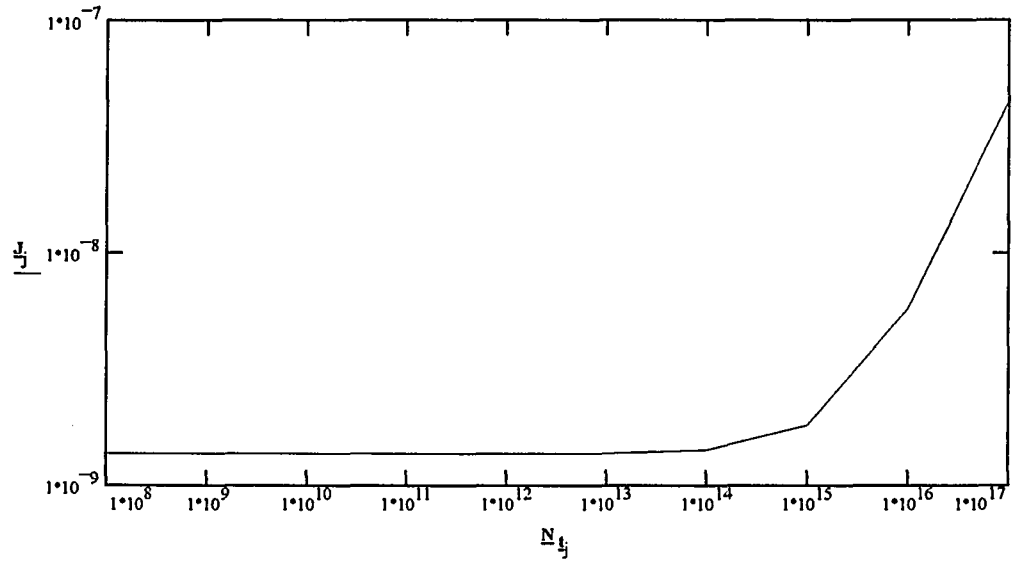
Current density:

$$\underline{J}_j := q \cdot \underline{G}_{\underline{tot.}_j} \cdot W$$

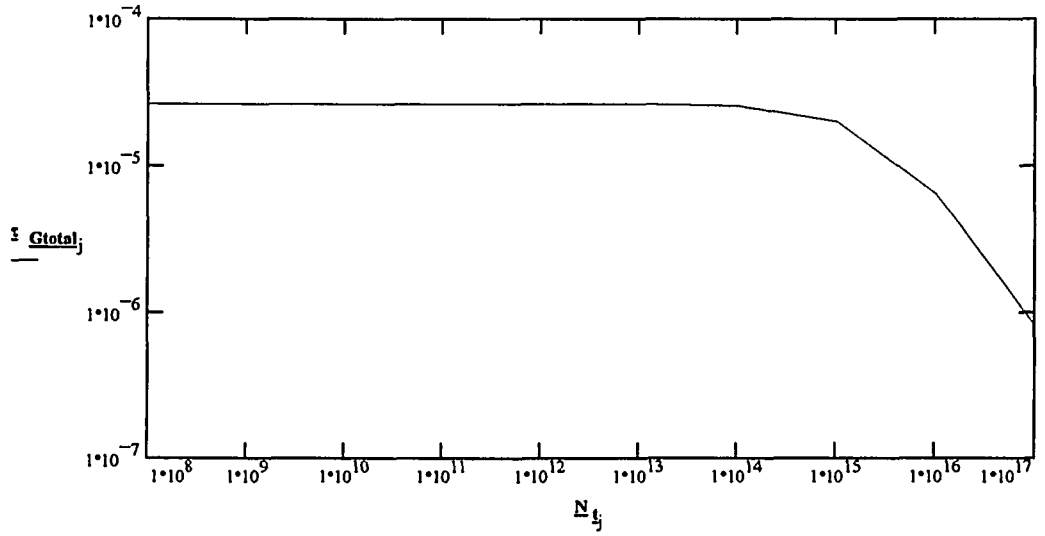
$$\tau_{\underline{Gtotal}} = \begin{bmatrix} 2.640 \cdot 10^{-5} \\ 2.640 \cdot 10^{-5} \\ 2.640 \cdot 10^{-5} \\ 2.639 \cdot 10^{-5} \\ 2.639 \cdot 10^{-5} \\ 2.631 \cdot 10^{-5} \\ 2.559 \cdot 10^{-5} \\ 2.005 \cdot 10^{-5} \\ 6.337 \cdot 10^{-6} \\ 8.084 \cdot 10^{-7} \end{bmatrix} \cdot sec \qquad \underline{J} = \begin{bmatrix} 1.371 \cdot 10^{-9} \\ 1.371 \cdot 10^{-9} \\ 1.371 \cdot 10^{-9} \\ 1.371 \cdot 10^{-9} \\ 1.371 \cdot 10^{-9} \\ 1.375 \cdot 10^{-9} \\ 1.414 \cdot 10^{-9} \\ 1.805 \cdot 10^{-9} \\ 5.709 \cdot 10^{-9} \\ 4.476 \cdot 10^{-8} \end{bmatrix} \cdot \frac{amp}{cm^2}$$

APPENDIX B. MATHCAD 4.0 IMPLEMENTATION OF MODEL CONTINUED

Graph of junction leakage vs. contaminant concentration:



Graph of Lifetime versus Contaminant Concentration:



APPENDIX C. DRAM MEMORY CELL TYPES

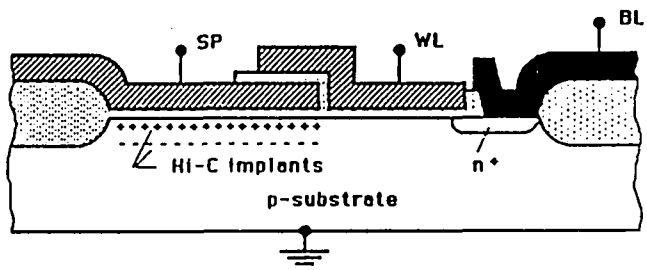
This appendix compiles various types of DRAM memory cell architectures. For each cell type, the level of integration of common use, the advantages and disadvantages, structure and in some cases some process details are given.

PLANAR CAPACITOR CELLS

High Capacitance Cell, Hi-C

1 Mbit

ADV	DISADV
Simple Process	Cell Area larger than Capacitor Area
Depletion Capacitance Increased	Large Space Charge Volume
High Charge Capacity	High Leakage
	High SER



From reference [57]

TRENCH CAPACITOR CELLS

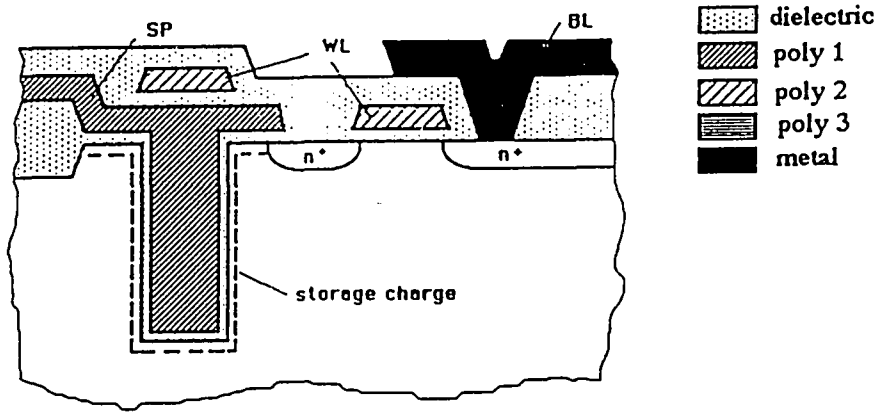
Corrugated Capacitor Cell, CCC

1 Mbit

ADV	DISADV
Reduced Cell Area	Punch-Through to Adjacent Trenches
	Large Space Charge Volume
	High Leakage
	High SER

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

Corrugated Capacitor Cell, CCC, Continued

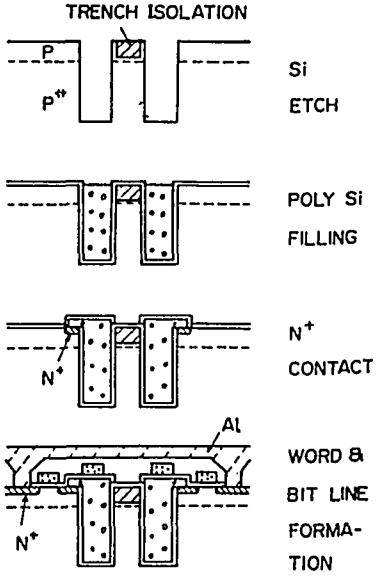


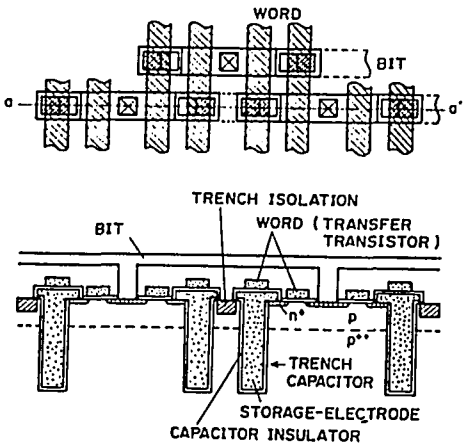
From reference [57]

Substrate Plate Trench Cell, SPT
1 Mbit, 4 Mbit

ADV
Reduced Cell Area

DISADV
Still Large Space Charge Volume
(although reduced compared to CCC)
Relatively High Leakage
Relatively High SER





From reference [14]

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

Buried Storage Electrode Cell, BSE

Another name for SPT cell (see above).

Buried Stacked Capacitor Cell, BSC

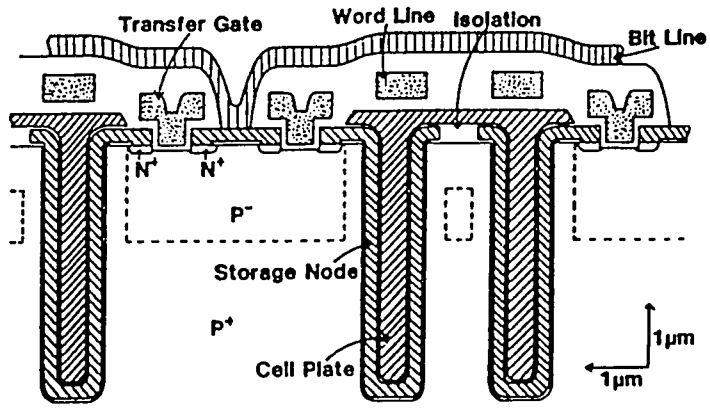
4 Mbit, 16 Mbit, 64 Mbit

ADV

Reduced Cell Area
Large Relative Reduction in
Space Charge Volume Compared
to Preceding Cells

DISADV

Complex Process



From reference [31]

Trench Transistor Cell, TTC

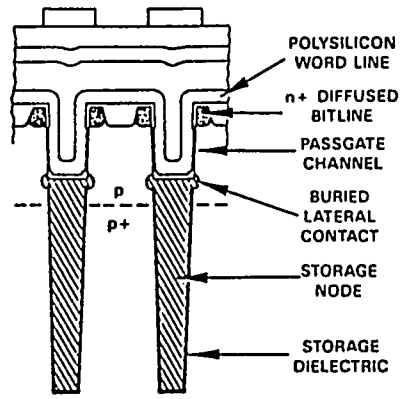
4 Mbit

ADV

Large Cell Area Reduction

DISADV

Very Complex Process
Still Large Space Charge Volume
(although reduced compared to CCC)
Relatively High Leakage
Relatively High SER



From reference [23]

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

Folded Capacitor Cell, FCC

4 Mbit, 16 Mbit

ADV

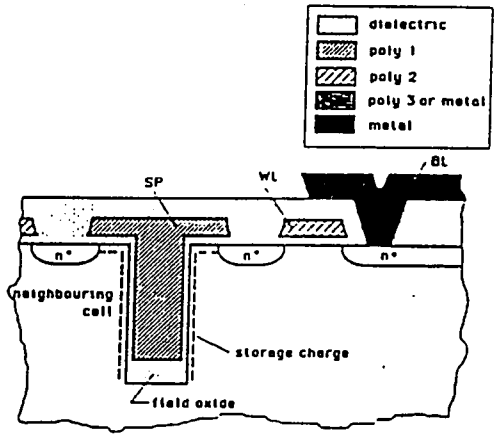
Reduced Cell Area

DISADV

Large Space Charge Volume

High Leakage

High SER



From reference [57]

Isolation Merged Vertical Cell, IVEC

16 Mbit, 64 Mbit, 256 Mbit

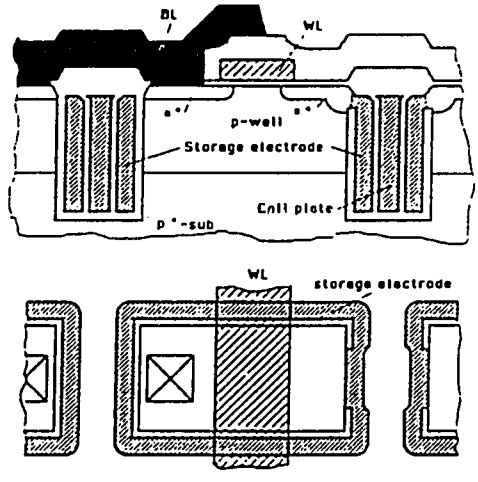
ADV

Reduced Cell Area

Large Relative Reduction in Space Charge Volume Compared to Preceding Cells

DISADV

Complex Process



From reference [57]

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

Buried Trench Capacitor Cell, BTC

4 Mbit, 16 Mbit, 64 Mbit

ADV

Large Cell Area Reduction

DISADV

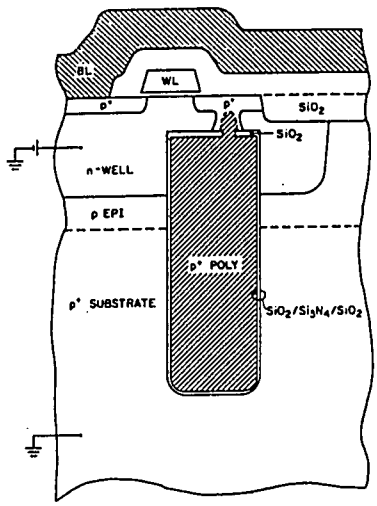
Very Complex Process

Still Large Space Charge Volume

(although reduced compared to CCC)

Relatively High Leakage

Relatively High SER



From reference [56]

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

STACKED CAPACITOR CELLS

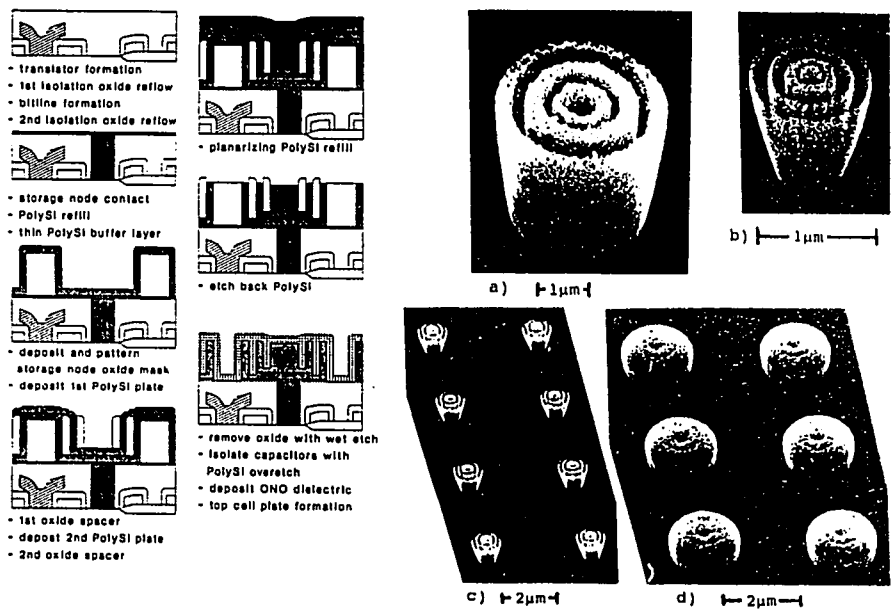
Stacked Capacitor Cell, STC

Most of the stacked cell variants were shown previously in Figure 12. The majority of reported 64 and 256 Mbit cells are variants of this cell type. The following cells are examples of architectures being used in advanced cells.

Multilayer Vertical Stacked Capacitors, MVSTC

64 Mbit, 256 Mbit

ADV	DISADV
Large Cell Area Reduction	Poor Topography
Large Reduction in Space Charge Volume	
Low Leakage	
Low SER	
Easily Adaptable for Alternate Dielectric Processes	



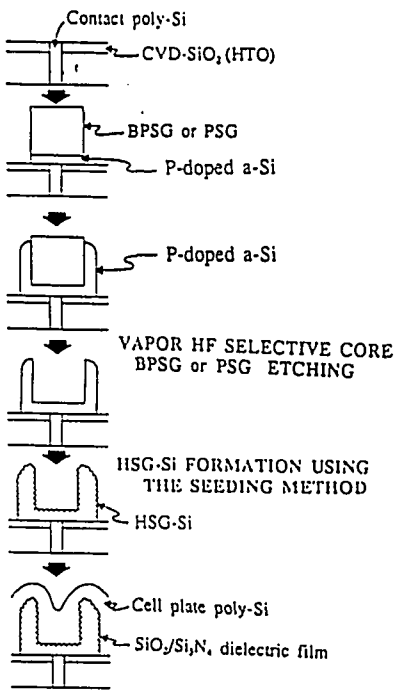
From reference [89]

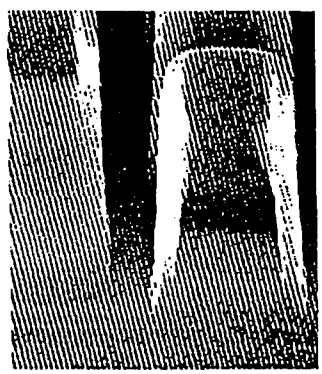
APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

Cylindrical, Hemispherical Grained Silicon, HSG-Si

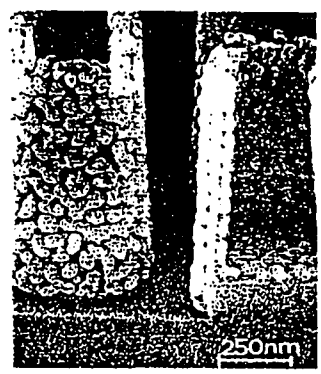
256 Mbit

ADV	DISADV
Large Cell Area Reduction	Poor Topography
Large Reduction in Space Charge Volume	
Low Leakage	
Low SER	
Easily Adaptable for Alternate Dielectric Processes	





After selective core BPSG etching by vapor HF.



After HSG-Si formation.

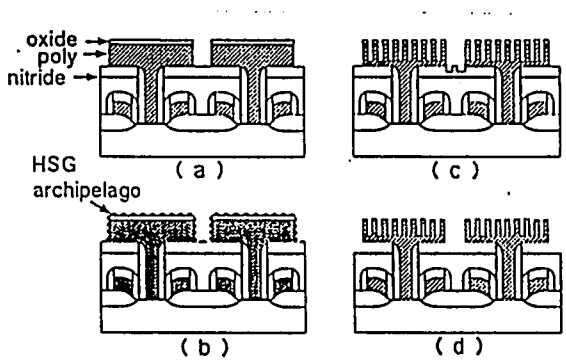
From reference [50]

APPENDIX C. DRAM MEMORY CELL TYPES CONTINUED

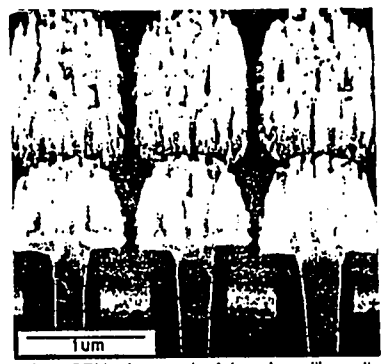
Micro Villus Patterning, MVP

256 Mbit

ADV	DISADV
Large Cell Area Reduction	Poor Topography
Large Reduction in Space Charge Volume	
Low Leakage	
Low SER	
Easily Adaptable for Alternate Dielectric Processes	



Process sequence of the cell structure using MVP technology



SEM micrograph of the micro villus cell structure with Fin undercut

From reference [52]

APPENDIX D. CAPTURE CROSS-SECTIONS FROM LITERATURE

ELEMENT	σ_p [cm^2]	σ_n [cm^2]	REFERENCE
Au	$1.0x10^{-15}$	$3.5x10^{-15}$	90
	$1.0x10^{-13}$	$1.0x10^{-16}$	91
	$2.4x10^{-15}$	$6.3x10^{-15}$	92
Ag	$2.1x10^{-16}$	$7.3x10^{-16}$	68*
Bi	$1.0x10^{-15}$	$1.0x10^{-15}$	94
Co	$1.0x10^{-16}$	$1.0x10^{-15}$	68*
Cu	$7.0x10^{-19}$	$1.7x10^{-16}$	94
	$3.5x10^{-20}$		93
Cr	$2.8x10^{-17}$	$9.8x10^{-17}$	68*
Fe	$7.0x10^{-16}$	$1.5x10^{-15}$	94
In	$1.0x10^{-16}$	$1.5x10^{-13}$	94
Mn	$5.0x10^{-15}$	$5.0x10^{-15}$	68*
Mo	$1.5x10^{-16}$	$5.3x10^{-16}$	68*
Ni	$4.0x10^{-14}$	$1.0x10^{-18}$	68*
Pb	$1.1x10^{-15}$	$3.9x10^{-15}$	68*
Sn	$2.5x10^{-16}$	$8.9x10^{-16}$	68*
Ti	$2.2x10^{-17}$	$7.6x10^{-17}$	68*
W	$8.2x10^{-16}$	$6.0x10^{-17}$	94
Zn	$1.0x10^{-14}$	$1.0x10^{-18}$	94

* Capture cross-sections computed based on elements where experimental data were not available, capture cross-sections were estimated using gold as a reference element in Yassievich's model which shows capture cross-section proportional to the trapping element's atomic number cubed, Z^3 [68].

APPENDIX F. SANDIA NATL. LAB.'S 1.25μm CMOS PROCESS

```
1  PREDIFFUSION CLEAN
2  PAD OXIDATION - 365 Ang
3  NITRIDE DEPOSIT - 1175 Ang
4  LASER SCRIBE
5  CLEAN
6  PTUB RESIST
7  CLEAR N-WELL
8  IMPLANT 3E12 @ 100KEV  PHOS
9  IMPLANT 2.5E11 @180KEV ARS
10 OXIDIZE 6300 Ang
11 STRIP NITRIDE OFF PWELL
12 IMPLANT PWELL 5E12 @50KEV BORON
13 DRIVE IN - 750 Ang
14 IMPLANT 6E12 @100KEV BORON
15 STRIP ALL OXIDE
16 OXIDIZE FIELD 600 Ang
17 CVD FIELD OX - 4400 Ang
18 STRIP BACKSIDE
19 THINOX PR/ETCH
20 GATE 0 GROWTH
21 IMPLANT 1.5E12 @35KEV BORON
22 STRIP GATE 0
23 GROW GATE OXIDE 180 Ang
24 DEPOSIT GATE POLY 4.4K Ang
25 GATE PR/ETCH
26 STRIP BACKSIDE
27 SCREEN OXIDE
28 LDD IMPLANT 5E13 @80KEV PHOS
29 P+ PR
30 P+ IMPLANT 2E15 @30KEV BORON
31 LDD SPACER TEOS 3K Ang
32 ETCH SPACER
33 SCREEN OX
34 N+ PR
35 N+ IMPLANT 5E15 @80KEV ARS
36 N+ IMPLANT 3E14 @35KEV PHOS
37 ANNEAL IMPLANT 850C/30MIN N2
38 TEOS 2500 Ang
39 BPTEOS 5500 Ang
40 DENSIFY
41 STRIP BACKSIDE (POLY/OXIDE)
42 WINDOW1 PR/ETCH
43 METAL 1  5K Ang (.75% S1/.5% Cu)
44 METAL1/ETCH
45 ALLOY 375 C 30MIN N2
46 DEPOSIT 12000 Ang 3 % CAP GLASS
47 CAP GLASS PR/ETCH
48 TEST
```

SNLA DIV 2132

MDT 1/13/89

REFERENCES

[1] J. Schmid, R. Craigin, C. Damianou, J. Hohl, R. Schrimpf, H. Parks and J. Ramberg, "*A Simple Model for Estimating Allowable Transition Metal Contamination Levels in DRAMs*," TECHON '90 Extended Abstract Volume, pp. 263-266, 1990.

[2] K. Yagi and S Shimizu, "*Trends in Submicron Silicon Process Technology*," Hitachi Hyoron, vol. 71, pp. 1-8, May 1989.

[3] M. Inoue et al, "*A 16-Mbit DRAM with a Relaxed Sense-Amplifier-Pitch Open-Bit-Line Architecture*," IEEE J. of Solid State Circuits, Vol. 23, No. 5, pp. 1104-1111, October 1988.

[4] P. Lin, R. Marcus and T. Sheng, "*Leakage and Breakdown in Thin Oxide Capacitors - Correlation with Decorated Stacking Faults*," J. Electrochem. Soc., Vol. 130, pp. 1878-1883, September 1983.

[5] A. Tasch and L. Parker, "*Memory Cell and Technology Issues for 64- and 256-Mbit One-Transistor Cell MOS DRAMs*," Proceedings of the IEEE, Vol. 77, No. 3, pp. 374-388, March 1989.

[6] W. Regitz and J. Karp, "*The Three-Transistor Cell 1024-Bit 500 nsec MOS RAM*," IEEE Jour. of Solid-State Circuits, Vol. SC-5, P.181, October 1970.

[7] T. Sugibayashi et al, "*A 30 ns 256 Mbit DRAM with Multi-Divided Array Structure*," 1993 IEEE ISSCC, pp. 50-51, 1993.

[8] T. Ozaki et al, "*A Surrounding Isolation-Merged Plate Electrode (SIMPLE) Cell with Checkered Layout for 256 Mbit DRAMs and Beyond*," 1991 IEEE IEDM, pp. 469-472, 1991.

[9] T. Lowery, G. Cloud, D. Kelly, P. Zagar and M. Seyyedy, "*The 64 Megabit DRAM Challenge*," Semiconductor International, Vol. 15, No. 6, pp. 48-52, May 1993.

[10] M. Kumanoya et al, "*A 90ns 1 Mbit DRAM with Multi-Bit Test Mode*," 1985 IEEE ISSCC, pp. 240-241, 1985.

[11] C. Webb et al, "*A 65ns CMOS 1 Mbit DRAM*," 1986 IEEE ISSCC, pp. 262-263, 1986.

[12] D. Yaney et al, "*Technology for the Fabrication of a 1 Mbit CMOS DRAM*," 1985 IEEE IEDM, pp. 698-701, 1985.

[13] Y. Takemae et al, "*A 1 Mbit DRAM with 3-Dimensional Stacked Capacitor Cells*," 1985 IEEE ISSCC, pp. 250-251, 1985.

[14] M. Sakamoto et al, "*Buried Storage Electrode (BSE) Cell for Megabit DRAMs*," 1985 IEEE IEDM, pp. 710-712, 1985.

[15] ICE, "*Construction Analysis of the Texas Instruments TMS4C1024-12L 1 Megabit DRAM Date Code 8822*," ICE, Scottsdale, AZ, 1989.

[16] M. Taguchi et al, "*Dielectrically Encapsulated Trench Capacitor Cell*," 1986 IEEE IEDM., pp. 136-139, 1986.

[17] M. Takada et al, "*A 4 Mbit DRAM with Half Internal-Voltage Bitline Precharge*," 1986 IEEE ISSCC, pp. 270-271, 1986.

[18] M. Nagatomo et al, "*A High Density 4 Mbit DRAM Process Using folded Bitline Adaptive Side-Wll Isolated Capacitor (FASIC) Cell*," 1986 IEEE IEDM, pp. 144-147, 1985.

[19] K. Nakamura et al, "*Buried Isolation Capacitor (BIC) Cell for Megabit MOS Dynamic RAM*," 1984 IEEE IEDM, pp. 236-239, 1984.

[20] K. Rao et al, "*Trench Capacitor design Issues in VLSI DRAM Cells*," 1986 IEEE IEDM, pp. 140-143, 1986.

[21] W. Richardson et al, "*A Trench Transistor Cross-Point DRAM Cell*," 1985 IEEE IEDM, pp. 714-717, 1985.

[22] A. Shah et al, "*A 4 Mbit DRAM with Cross-Point Trench Transistor Cell*," 1986 IEEE ISSCC, pp. 268-269, 1986.

[23] S. Banerjee, D. Coleman, W. Richardson and A. Shah, "*Leakage Mechanisms in the Trench Transistor DRAM Cell*," IEEE Trans. Elec. Dev., Vol. 35, No. 1, pp. 108-116, January 1988.

[24] M. Wada, K. Hieda and S. Watanabe, "*A folded Capacitor Cell (F.C.C.) for Future Megabit DRAMs*," 1984 IEEE IEDM, pp. 244-247, 1984.

[25] P. Murkin et al, "*Isolation Related Leakage in a 4 Mbit DRAM Cell*," Solid State Dev., 17th ESSDERC, pp. 173-176, 1987.

[26] T. Furuyama et al, "*An Experimental 4 Mbit CMOS DRAM*," 1986 IEEE ISSCC, pp. 272-273, 1986.

[27] K. Kusters et al, "*Stacked Capacitor in trench cell for 16 Mbit DRAM*," 19th ESSDERC, pp. 907-910, 1989.

[28] T. Takeshima et al, "*A 55ns 16 Mbit DRAM*," 1989 IEEE ISSCC, pp. 246-247, 1989.

[29] M. Ohkura et al, "*A Three-dimensional DRAM Cell of Stacked Switching-Transistor in SOI (SSS)*," 1985 IEEE IEDM., pp. 718-721, 1985.

[30] S. Nakajima, K. Miura, K. Minegishi and T. Morie, "*An Isolation-Merged Vertical Capacitor cell for Large Capacity DRAM*," 1984 IEEE IEDM., pp. 240-243, 1984.

[31] K. Tsukamoto et al, "*Double Stacked Capacitor with Self-Aligned Poly Source/Drain Transistor (DSP) Cell for Megabit DRAM*," 1987 IEEE IEDM., pp. 328-331, 1987.

[32] T. Kaga et al, "*A 4.2μm^2 Half-Vcc Sheath-Plate Capacitor DRAM Cell with Self-Aligned Buried Plate-Wiring*," 1987 IEEE IEDM., pp. 332-335, 1987.

[33] Y. Kawamoto et al, "*A Half Micron Technology for an Experimental 16 Mbit DRAM using i-Line Stepper*," 1987 Symp. on VLSI Tech., pp. 17-18, 1987.

[34] M. Aoki et al, "*An Experimental 16 Mbit DRAM with Transposed Data-Line Structure*," 1988 IEEE ISSCC, pp. 250-251, 1988.

[35] S. Watanabe et al, "*An Experimental 16 Mbit CMOS DRAM Chip with a 100MHz Serial Read/Write Mode*," 1988 IEEE ISSCC, pp. 248-249, 1988.

[36] S. Kimura et al, "*A New Stacked Capacitor DRAM Cell Characterized by a Storage Capacitor on a Bit-Line Structure*," 1988 IEEE IEDM, pp. 596-599, 1988.

[37] P. Bakeman et al, "*A High Performance 16 Mbit DRAM Technology*," 1990 Symp. on VLSI Tech., pp. 11-12, 1990.

[38] P. Singer and P. Burggraaf, "*IBM First with 16 Mbit DRAM*," Semiconductor International, Vol. 15, No. 5, pp. 15-16, April 1992.

[39] K. Arimoto et al, "*A 60 ns 3.3 V 16 Mbit DRAM*," 1989 IEEE ISSCC, pp. 244-245, 1989.

[40] S. Fujii et al, "*A 45 ns 16 Mbit DRAM with Triple-Well Structure*," 1989 IEEE ISSCC, pp. 248-249, 1989.

[41] T. Yamada et al, "*A 64 Mbit DRAM with Meshed Power Line and Distributed Sense-Amplifier Driver*," 1991 IEEE ISSCC, pp. 108-109, 1991.

[42] S. Mori et al, "*A 45 ns 64 Mbit DRAM with a Merged Match-Line Test Architecture*," 1991 IEEE ISSCC, pp. 110-111, 1991.

[43] M. Taguchi et al, "*A 40 ns 64 Mbit DRAM with Current-Sensing Data-Bus Amplifier*," 1991 IEEE ISSCC, pp. 112-113, 1991.

[44] Y. Oowaki et al, "*A 33 ns 64 Mbit DRAM*," 1991 IEEE ISSCC, pp. 114-115, 1991.

[45] Y. Kawamoto et al, "*A 1.28μm^2 Bit-Line Shielded Memory Cell Technology for 64 Mbit DRAMs*," 1990 Symp. on VLSI Tech., pp. 13-14, 1990.

[46] Y. Kohyama et al, "*Buried Bit-Line Cell for 64 Mbit DRAMs*," 1990 Symp. on VLSI Tech., pp. 17-18, 1990.

[47] D. Kenney et al, "*A Buried-Plate Trench Cell for a 64 Mbit DRAM*,", 1992 Symp. on VLSI Tech., pp. 14-15, 1992.

[48] H. Itoh et al, "*Two Step Deposited Rugged Surface (TDRS) Storage node and Self Aligned Bitline-Contact Penetrating Cellplate (SABPEC) for 64 Mbit DRAM STC Cell*," 1991 Symp. on VLSI Tech., pp. 9-10, 1991.

[49] H. Koike, "*A 30ns 64 Mbit DRAM with Built-In Self-Test and Repair Function*," 1992 IEEE ISSCC, pp. 150-151, 1992.

[50] H. Watanabe et al, "*A New Cylindrical Capacitor using Hemispherical Grained Si (HSG-Si) for 256 Mbit DRAMs*," 1992 IEEE IEDM, pp. 259-262, 1992.

[51] N. Shimizu et al, "*A Poly-Buffer Recessed LOCOS Process for 256 Mbit DRAM Cells*," 1992 IEEE IEDM, pp. 279-282, 1992.

[52] J. Ahn et al, "*Micro Villus Patterning (MVP) Technology for 256 Mbit DRAM Stack Cell*," 1992 Symp. on VLSI Tech., pp. 12-13, 1992.

[53] K. Sagara et al, "*A 0.72μm^2 Recessed STC (RSTC) Technology for 256 Mbit DRAMs using Quarter Micron Phase-Shift Lithography*," 1992 Symp. on VLSI Tech., pp. 10-11, 1992.

[54] S. Kamiyama et al, "*Highly Reliable 2.5nm Ta_2O_5 Capacitor Process Technology for 256 Mbit DRAMs*," 1991 IEEE IEDM, pp. 827-830, 1991.

[55] K. Koyama et al, "*A Stacked Capacitor with $(Ba_xSr_{1-x})TiO_3$ for 256 Mbit DRAM*," 1991 IEEE IEDM, pp. 823-826, 1991.

[56] N. Lu, "*Advanced Cell Structures for Dynamic RAMs*," IEEE Circuits & Dev. Mag., Vol. 15, No. 1, pp. 27-36, January 1989.

[57] H. Maes, G. Groeseneken, H. Lebon and J. Witters, "*Trends in Semiconductor Memories*," Microelectronics Jour., Vol. 20, No. 1-2, pp. 9-58, 1989.

[58] A. Mohsen et al, "*A High Density, High Performance 1T DRAM Cell*," 1982 IEEE IEDM., pp. 616-619, 1982.

[59] W. Lee, T. Osakama, K. Asada and T. Sugano, "*Design Methodology and Size Limitations of Submicrometer MOSFET's for DRAM Application*," IEEE Trans. on Elect. Dev., Vol. 35, No. 11, pp. 1876-1884, November 1988.

[60] B. Majkusiak and A. Jakubowski, "*Very Thin Oxides in VLSI Technology: Properties and Device Implications*," Microelectronics Jour., Vol. 21, No. 2, pp.21-40, 1990.

[61] K. Schuegraf, C. King and C. Hu, "*Ultra-Thin Silicon Dioxide Leakage Current and Scaling Limit*," 1992 Symp. on VLSI Tech., pp. 18-19, 1992.

[62] A. Tasch, P. Chatterjee, H. Fu and T. Holloway, "*The Hi-C RAM Cell Concept*," IEEE Trans. on Elect. Dev., Vol. 25, No. 1, p.33, 1978.

[63] P. Fazan, H. Chan and V. Mathews, "Stacked Capacitor Structures for 64 Mbit DRAMs and Beyond," Semiconductor International, Vol. 15, No. 6, pp. 108-112, May 1992.

[64] W. Trickle, "*Analysis of an Unusual Leakage Phenomenon in a 1 Mbit DRAM Cell*," 1988 ICMPC, pp. 461-465, 1988.

[65] K. Sunouchi et al, "*A Self-Amplifying (SEA) Cell for Future High Density DRAMs*," 1991 IEEE IEDM, pp. 465-468, 1991.

[66] D.K. Schroder, "*The Concept of Generation and Recombination Lifetimes in Semiconductors*," IEEE Trans. Elect. Dev., Vol. 29, No. 8, pp. 1336-1338, August 1982.

[67] B. Craigin, H.G. Parks and R. Schrimpf, "*Metallic Impurity Requirements based on Shockley-Read-Hall Theory*,"Submitted for publication to IEEE Trans. on Elec. Dev., July 1992.

[68] I. Yassievich, "*Carrier Capture by Attraction Centers in Semiconductors*," Semiconductor Physics, Consultants Bureau, NY, pp. 519-530, 1986.

[69] M. Lax, "*Cascade Capture of Electrons in Solids*," Phys. Rev, Vol. 119, No. 5, pp. 1502-1523, September 1960.

[70] R. Pierret, "*Advanced Semiconductor Fundamentals*," Addison-Wesley Pub. Co. Reading, MA., pg. 155, 1988.

[71] J. Feldman, "*The Physics and Circuit Properties of Transistors*," Wiley & Sons, NY, pg. 239, 1972.

[72] T. Warner and B. Grung, "*Transistors, Fundamentals for the Integrated-Circuit Engineer*," Wiley & Sons, NY, pg. 195, 1983.

[73] A. Ohsawa, K. Honda, R. Takizawa, T. Nakanishi, M. Aoki, and N. Toyokura, "*Effects of Impurities on Microelectronics Devices*," Proc. Sixth Int. Symp. Silicon Material Science and Tech.: Semiconductor Silicon 1990, pp. 601-613, Electrochem. Soc., 1990.

[74] T. Hattori, "*Contamination Control: Problems and Prospects*," Solid State Technology, Vol. 33, No. 7, pp. s1-s8, July 1990.

[75] S. Fujimura and H. Yano, "*New Device Degradation Mechanism, Heavy Metal Contamination from Resists During Plasma Stripping*," Proceedings - The Electrochem. Soc., Vol. 87, No. 6, pp. 352-365, 1987.

[76] R. Takizawa, T. Nakanishi and A. Ohsawa, "*Degradation of Metal-Oxide-Semiconductor Devices Caused by Iron Impurities on the Silicon Wafer Surface*," J. Appl. Phys., Vol. 62, No. 12, pp. 4933-4935, December 1987.

[77] T. Shiaiwa, N. Fujino, S. Sumita and Y. Tanizoe, "*Chemical Analysis of Metallic Impurity on the Surface of Silicon Wafers*," Semiconductor Fabrication: Technology and Metrology, ASTM STP 990, pp. 314-323, 1989.

[78] J. Pribley, M. Ghezzo, H.G. Parks and D. Brown, "*Advanced CMOS Process Technology*," VLSI Electronics Microstructure Science, Vol. 19, Academ. Press, Inc., NY, 1989.

[79] F. Kern, M. Itano, I. Kawanabe, M. Miyashita, R. Rosenburg and T. Ohmi, Paper presented at 37th Annual IES Meeting, San Diego, CA, May 1991.

[80] H.G. Parks, R. Schrimpf, R. Craigin, R. Jones and P Resnick, "*Device Degradation due to Process Chemical Contamination*," Presented at 183rd meeting of Electrochem. Soc, Honolulu HI, 16-21 May 1993.

[81] H.G. Parks, R. Schrimpf, R. Craigin, R. Jones and P. Resnick, "*Test Structure Metrology of Homogeneous Contamination*,"Proc. IEEE Internat'l. Meeting on Microelec. Test Struct., pp. 207-212, 1993.

[82] V. Penka and W. Hub, "*Detection of Metallic Trace Impurities on Si(100) Surfaces with Total Reflection X-ray Fluorescence (TXRF)*," Fresenius Z. Anal. Chem. 333, pp. 568-589, 1989.

[83] T. Shiraiwa, N. Fujino, S. Sumita and Y. Tanizoe, "*Chemical Analysis of Metallic Impurity on the Surface of Silicon Wafers*," Semiconductor Fabrication: Technology and Metrology, ASTM STP 990, pp. 314-323, 1989.

[84] E. Hsu, H.G. Parks, R. Craigin, S. Tomooka, J. Ramberg and R. Lowry, "*Deposition Characteristics of Metal contaminants from HF-Based Process Solutions onto silicon Wafer Surfaces*," J. Electrochem. Soc., Vol. 139, No. 12, pp3659-3664, December 1992.

[85] P. Eichinger, H.J. Rath, and H. Schwenke, "*Application of Total Reflection X-ray Fluorescence Analysis for Metallic Trace Impurities on Silicon Wafer Surfaces*," Semiconductor Fabrication: Technology and Metrology, ASTM STP 990, pp. 305-313, 1989.

[86] S. Sze, "*VLSI technology*, " McGraw-Hill, NY, pg. 309, 1988.

[87] S. Coffa, L. Calcagno and S. Campisano, "*Control of Minority Carrier Lifetime by Gold Implantation in Semiconductor Devices*," J. Electrochem. Soc., Vol. 136, No. 7, pp. 2073-2075, July 1989.

[88] D. Schroder, "*Semiconductor Material and Device Characterization*," Wiley & Sons, NY, pp. 360-367, 1990.

[89] D. Temmler, "*Multilayer Vertical Stacked Capacitors (MVSTC) for 64 Mbit and 256 Mbit DRAMs*," 1991 Symp. on VLSI Tech., pp. 13-14, 1991.

[90] G. Bemski, Phys. Rev., 111, pg 1515, 1958.

[91] F. Richou, G. Pelous, and D. Lecrosnier, "*Thermal Generation of Carriers in Gold-Doped Silicon*," Jour. of Appl. Phys., Vol. 51, pg. 6252, December 1980.

[92] J. Fairfield and B. Gokhale, "*Gold as a Recombination Center in Silicon*," Solid-State Electronics, Vol. 8, pg. 685, 1965.

[93] Y. Zibuts, L. Partiskii, S. Ryvkin and Z. Dokholyan, Soviet Phys. - Solid State, Vol. 8, pg. 2041, 1967.

[94] E. Schibli and A. Milnes, "*Lifetime and Capture cross-section Studies of Deep Impurities in Silicon*," Mater. Sci. Eng., Vol. 2, pp. 229-241, 1967/68.