Advanced Circuit Design of Gigabit-Density Ferroelectric Random-Access Memories

Von der Fakultät für
Elektrotechnik und Informationstechnik der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur

Jürgen Thomas Rickes

aus Neuwied

Berichter: Universitätsprofessor Dr.-Ing. Rainer M. Waser

Universitätsprofessor Dr.-Ing. Bruce F. Cockburn

Tag der mündlichen Prüfung: 6. Dezember 2002

PREFACE

This dissertation arose during my work at the Institut für Werkstoffe der Elektrotechnik (RWTH Aachen), the research center IMEC (Leuven, Belgium), and Agilent Technologies (Palo Alto, US).

First, I would like to thank my promoter Prof. Rainer Waser for his encouragement and guidance through the three years of my doctorate. He manages to strike the perfect balance between providing direction and encouraging independence. I am especially grateful to him for giving me the opportunity to participate in the FeRAM development program with Agilent Technologies and Texas Instruments and spending a significant part of my thesis in a company R&D environment of high scientific level.

I would like to express my appreciation to Ralph Lanham; the many discussions, guidance and friendly encouragements were essential for carrying out this work.

I want to thank my collaborators and colleagues, especially Andrei Bartic (IMEC), James Grace, Hugh McAdams (TI), and Scott Summerfelt (TI) for many inspiring and fruitful discussions.

I am also indebted to Prof. Bruce Cockburn who kindly agreed to be co-examiner in the jury.

Finally, I want to thank my wife Christine for her support through these years of study.

Jürgen Rickes
July 2002

Meinen Eltern,
Monika und Werner Rickes,
meinen Schwiegereltern,
Monika und Lothar Siebel
und meiner Frau Christine
gewidmet

TABLE OF CONTENTS

I. INTRODUCTION

FUTURE applications demand very large memories with extremely low power dissipation but high performance and that are, ideally, non-volatile. Of course, all of these requirements need to be achieved at a very low cost. The electronics industry is seeking to reduce the size of systems by putting more and more integrated circuits on a single semiconductor chip. In addition, many applications demand a single-chip solution. For example, ultra-portable devices like cellular phones, pagers, smart cards, personal digital assistants (PDAs), or wearable computers, greatly benefit from single-chip solutions because of the reduced weight and smaller space required. The main advantages of system-on-a-chip (SoC) solutions are reduced system cost, increased performance, and reduced power consumption. The market for these devices is expected to grow rapidly in the near future. Embedded memory, especially, is becoming increasingly important. The most important requirements that an embedded memory has to satisfy are logic process compatibility, minimum added process cost, and low voltage operation. Small cell size, infinite number of read and write cycles, and for all battery-driven applications, non-volatility and ultra-low power consumption in order to extend the time to the next required re-charge of the battery, are also very desirable. Typical specifications of different memory types are compared in Table 1.

Table 1: Comparison of existing memory technologies

	SRAM	DRAM	Flash
Read cycles	$> 10^{15}$	$> 10^{15}$	$> 10^{15}$
Write cycles	$> 10^{15}$	$> 10^{15}$	$\sim 10^{6}$
Write voltage	V_{DD}	V_{DD}	10 - 18 V
Write time	6-10 ns	30-70 ns	1 µs - 1 ms
Access time	6-10 ns	30-70 ns	30-70 ns
Standby current	10-30 µA	< 400 µA	< 10 µA
Cell size	~ 100 F^2	8 F^2	6-10 F^2
Data retention	none	none	> 10 years
Memory size	16 Mbit	1 Gbit	1 Gbit

Static random access memory (SRAM) is particularly suitable for embedded memory since its memory cell is logic based. Consequently, it is fully compatible with the logic process with no added process cost. It has very fast read and write operation times, a non-destructive read, and operates at low voltages. Power consumption during standby is low as well. However, SRAM has a very large memory cell with 4 or 6 transistors (4-T or 6-T) and low yield is a problem for embedded macro sizes larger than 16-Mbit.

It is therefore very costly to realize large embedded memories with SRAM. Despite the fact that it has low standby power consumption, it is still volatile. To maintain stored information, it must continue to be powered on. Moreover, the leakage current is increasing dramatically with every process generation and is becoming a problem.

Dynamic random access memory (DRAM) is volatile like SRAM. Moreover, it requires periodic refresh operations to maintain stored information. The power-consuming refresh operation has to be executed sufficiently often during standby, which is reflected in higher overall standby power consumption. On the other hand, its 1 transistor (1-T) memory cell is much smaller than the SRAM cell and is only slightly larger than the NAND Flash cell. The read and write operations are about 4-6x slower than for SRAM but are still very acceptable for most applications. Additional masks are required for embedded DRAM – resulting in increased process cost.

Flash memory is non-volatile. When powered off, it maintains stored information with zero power consumption. NAND-Flash also has a much smaller cell size than SRAM or even DRAM. This allows the storage of more information per chip area. Nevertheless, Flash memory typically requires large erase and programming voltages and has slow erase and program operations. Read operations are fast, but write operations are extremely slow, which makes Flash unacceptable for applications that require a fast write. Also the number of write cycles is limited.

Due to different weaknesses in each case, none of the existing memory technologies is capable of satisfying all of the requirements simultaneously. It is therefore common practice to combine two memory types in order to balance out weaknesses of any single type (e.g. Flash/SRAM combinations). This allows the specific demands of a particular application to be met. However, each type of memory requires a different manufacturing process and specialized processes are required to achieve optimal per-

formance. Clearly there is a need for a single memory technology that can satisfy all (or a larger part) of the above requirements at the same time.

It is widely believed that ferroelectric memory has the potential to be this kind of memory as it offers small cell size, fast operation, logic process compatibility, low added process cost, low voltage operation, non-volatility, and low power consumption. However, as of today, it is uncertain if ferroelectric memories will become competitive to existing memory technologies in the near future since commercially available ferroelectric random access memories (FeRAM / FRAM™) are still well below their expectations. For example, commercial FeRAM is still limited to small memory capacities of less than 256-kbit for IC-card applications. On the other hand, FeRAM is non-volatile and – since no refresh is necessary – standby power consumption for FeRAM is much smaller than for DRAM or even SRAM.

Table 2: Stand-alone FeRAM memory status

	1Mbit		4 Mbit	8 Mbit	32 Mbit
Company	NEC	Fujitsu	Samsung	Toshiba	Samsung
Source	JSSCC '96	ISIF '99	ISSCC '00	ISSCC '01	ISSCC '02
Ferroelectric	SBT	PZT	PZT	PZT	PZT
V_{DD}	3.3 V	5V / 3.3V	3.3 V	3.0 V	3.0 V
Access time	60 ns	60 ns	40 ns	40 ns	45 ns
Cycle time	100 ns	-	80 ns	70 ns	-
Technology	1.0 µm	0.5 µm	0.4 µm	0.25 µm	0.25 µm
Cap. Size	-	-	0.64 µm²	0.86 µm²	0.44 µm²
Cell Size	34.72 µm²	15.80 µm²	4 µm²	-	0.94 µm²
Chip size	91 mm²	-	52 mm²	76 mm²	90 mm²
Production	-	Yes	-	-	-

Performance and memory capacity have improved remarkably in recent years thanks to achieved process milestones for both ferroelectric and CMOS processes (see Table

™ FRAM is a trademark of Ramtron International Corporation. (Colorado Springs, CO, USA)

2). At present, the largest reported experimental chip size is 32-Mbit [1]. This was achieved in 0.25-µm technology using a 1T1C architecture. Access time is 45 ns at 2.7V. Also, an experimental 8-Mbit Chain FeRAM has been presented in 0.25-µm technology with an access time of 40 ns and a cycle time of 70 ns [2]. Table 3 compares the reported properties of other embedded memories assuming 0.18-µm design rules, and estimates what properties a hypothetical embedded FeRAM in these design rules must have in order to be competitive. When comparing the required properties for a 0.18-µm FeRAM (Table 3) to the reported properties of existing FeRAMs (Table 2), there seems to be a large discrepancy. Part of this could be attributed to the fact that ferroelectric memory technology is relatively new and many design and process problems still have to be solved. Other issues are inherent to the technology itself. In order to become competitive with existing memory technologies, improvements in several areas are mandatory for FeRAMs. For instance, the cell size needs to be reduced to less than 10 F^2, where F is the minimum feature size. At the same time, access and cycle times have to be substantially improved.

Table 3: Requirements for 0.18-µm generation embedded FeRAM

	eFeRAM	eFLASH	eDRAM	SRAM
Source		Philips	United	IBM
Macro size	16 Mbit	16 Mbit	16 Mbit	18 Mbit
Chip size	< 25 mm²		15.8 mm²	114 mm²
Cell size	< 0.6 µm²	0.78 µm²	0.37 µm²	4.23 µm²
Cell size in F^2	< 10 F^2	24 F^2	11 F^2	130 F^2
V_{DD}	< 1.8 V	1.8 V	1.5 V	1.8 V
Access time	< 12 ns	41 ns	10 ns	2.7 ns
Active power	< 1W		0.7 W	1.5 W

This work investigates new circuit design techniques for the development of future ferroelectric memories. Some of the presented techniques aim to solve well-known design issues; others aim to overcome material-related weaknesses. In Chapter III, state-of-the-art circuit design techniques relevant for the implementation of ferroelectric memories are reviewed. Chapter IV describes the ferroelectric capacitor model that has been developed as part of this thesis and that was used for all circuit simula-

tions. Chapter V describes the design of a first 0.35-µm FeRAM test chip that has been developed in cooperation with the research center IMEC. Chapters VI to VIII form the main part of this study. These chapters document an aggressive but systematic approach that could lead to the implementation of Gigabit-density embedded ferroelectric memory. The planning and design phase as well as part of the testing of a 0.13-µm 4-Mbit embedded FeRAM module is presented. This ambitious project was the result of a joint development program between Agilent Technologies and Texas Instruments. Chapter IX discusses the potential and design issues concerning chain-type FeRAM. Chapter X presents a scaling theory for ferroelectric memories and compares it to existing memory technologies.

II. BACKGROUND TO SEMICONDUCTOR MEMORIES

A. BASIC MEMORY ARCHITECTURE

A semiconductor memory is composed of three blocks: an array, a peripheral circuit, and an I/O circuit (Figure 1).

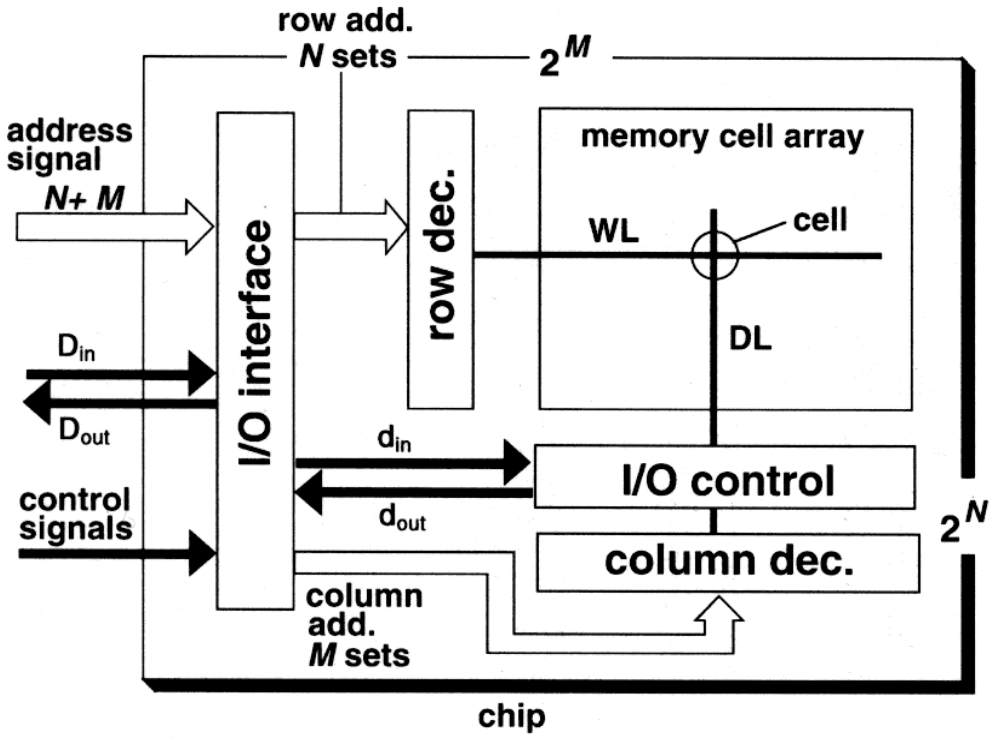


Figure 1: Memory chip architecture [3]

An array comprising a matrix of 2^N rows and 2^M columns can store binary information of $2^{N+M+K-1}$ bits if each element can store K bits (usually K=1). For example, 4-Mbit of information can be stored for N+M=22 and K=1. Any cell can be accessed at random by selecting both the corresponding row and column. The memory cell array is also called the matrix or core. The row is called an X line or word-line, while the column is called a Y line, bit-line, or data line. The matrix arrangement minimizes the number of driv-

ing circuits because a single word-line driver is shared by all cells of the same row, while a sense amplifier is shared by all cells of the same column.

The peripheral circuitry bridges between the memory array and the I/O interface circuit so that they can communicate with each other. It sends write data to a memory cell in the array under the control of the I/O circuit. A typical circuit is a decoder. It selects a logic circuit corresponding to one row or one column, based on the address signals from the I/O circuit.

The I/O circuit converts external signals, such as addresses, clocks, control signals, and data inputs, to the corresponding internal signals that activate the peripheral circuit. In addition, it outputs read data from the array as the data output of the chip. Data input and output buffers and clock control circuits are also typical components of the I/O interface circuit. [3]

B. CATEGORIES OF MEMORY CHIPS

The semiconductor memory [4] now widely used is categorized as random-access memory (RAM) and read-only memory (ROM), as shown in Figure 2.

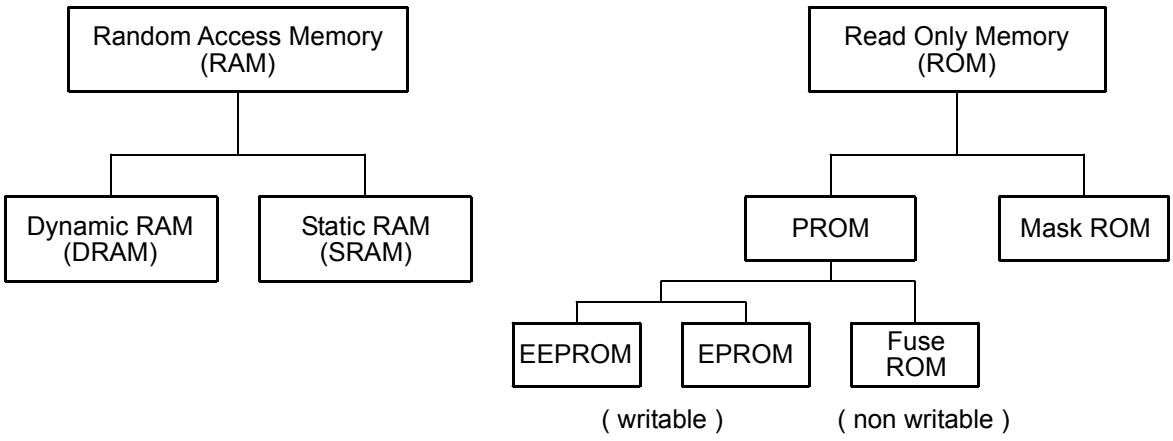


Figure 2: Categories of semiconductor memories

RAM permits random write and read operations for any memory cell in a chip. The stored data are usually volatile when the power-supply voltage is turned off. RAM is further classified into DRAM and SRAM. Due to the advantage of low cost, despite medium speed, DRAM is widely and extensively used for the main memory in personal computers. SRAM, which features high speed and ease of use, despite high cost, is also

used for the main memory of supercomputers, the cache memory in mainframe computers, microprocessors, and memory in handheld equipment. ROM, dedicated to read operation with non-volatile stored data, is classified into two categories, depending on the process of the write procedure: Programmable ROM (PROM), in which data are written after chip fabrication; and Mask ROM, in which data are written during chip fabrication by the use of a photo mask that contains the write data. PROM is further categorized into erasable PROM (EPROM), electrically erasable PROM (EEPROM), and Fuse ROM. EPROM erases the data by exposing the memory cells to ultraviolet rays, while it writes the data by electrical means. In EEPROM, the erase and write operations are both performed by electrical means. There are some drawbacks: the write speed is two to three orders slower than that of RAM, and there is an upper limit on the number of write operations of 10^4 – 10^5. Note that Flash memory – a kind of EEPROM – is being intensively developed and has emerged in the market offering the potential of high density and low cost, in some usages, it may replace magnetic disk memory. In Fuse ROM, users cannot rewrite data once they are written, because the fuses will be blown. ROM, which is generally cheaper than RAM, is used as memory for fax machines, telephones, and so on. [3]

C. MEMORY CELLS

The following section describes memory cells, which store binary information, “1” or “0”, in the above-described memory chips.

DRAM

A DRAM cell [5] comprises an n-channel transistor that works as a switch, and a capacitor for storing charge (Figure 3). For example, non-existence of charge at the capacitor corresponds to “1”, while existence of charge corresponds to “0”. In other words, for the voltage expression, a high stored voltage corresponds to “1” while a low stored voltage corresponds to “0”. The write operation is performed by turning on the switch and applying a voltage corresponding to the write data from the data line, DL, to the capacitor. Here, the switch is turned on by applying a sufficiently high voltage to the word-line, WL. The read operation is performed by turning on the switch. A resultant

signal voltage developed on the data line, depending on the stored data at the capacitor, is discriminated by a detector on the data line. In principle, the cell holds the data without power consumption. Actually, however, a leakage current at the p-n junction in the storage node degrades an initial high stored voltage, finally causing the loss of information. This loss can be avoided by a "refresh" operation: the cell is read before the stored voltage has become excessively decayed, and then it is rewritten by utilizing the resultant read information, so that the voltage is restored to its initial value. A succession of read-rewrite operations at a given time interval retains the data. The time interval, which is determined by the leakage current, is about 2–64 ms. The name DRAM is derived from the fact that data is dynamically retained by refresh operations, which differs from SRAM.

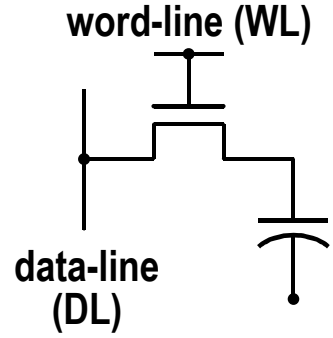


Figure 3: DRAM memory cell

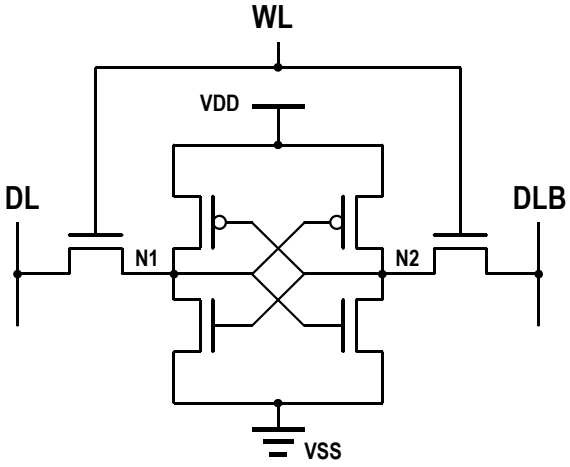


Figure 4: SRAM memory cell

SRAM

An SRAM cell consists of a flip-flop circuit that is constructed between a power supply voltage (V_{DD}) and the ground (V_{SS}), and two switching transistors (Figure 4). Data is communicated between a pair of data lines, DL and DLB, and a flip-flop by turning on two transistors. The write operation is performed by applying a differential voltage between a high voltage (H) and a low voltage (L) to a pair of data lines and thus to the

storage nodes, N_1 and N_2. For example, "1" is written for a polarity of H at DL (N_1) and L at DL bar (N_2) while "0" is written for an opposite polarity of L at DL and H at DL bar. The read operation is performed by detecting the polarity of a differential signal voltage developed on the data lines. No refresh operation is needed because the leakage currents at N_1 and N_2, if any, are compensated by a static current from the power supply, as long as V_{DD} is supplied. Thus, SRAM is easy to use, although the use of more transistors in a cell increases the memory cell area to more than four times that of DRAM.

PROM

Unlike RAMs, a PROM cell (Figure 5) needs an additional erase operation because it must be initialized to a state of zero net charge by extracting electrons from the floating gate (i.e. the storage node). The succeeding write operation is achieved by either injecting electrons to the floating gate or not doing so. For example, "0" is written for the injection of electrons, while "1" is written for the non-injection state, that is, the erased state. The read operation is performed by capacitive coupling between the word-line and the floating gate. For "0", the transistor is kept off even with the help of a positive pulse coupled from WL to the floating gate because electrons at the floating gate prevent the transistor from turning on. For "1", however, the transistor turns on. Thus, a detector on the data line can differentiate the currents to discriminate the information. Note that, in principle, the injected electrons never discharge because they are stored at the floating gate and surrounded by pure insulators. Data retention is insured even when the power supply is off, thus realizing a non-volatile cell.

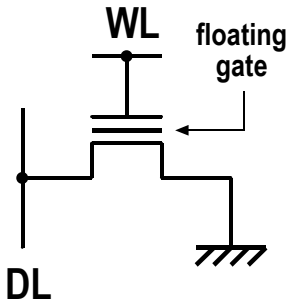


Figure 5: PROM memory cell

EPROM, EEPROM and Flash

In an EPROM chip, the stored data of all of the cells are simultaneously erased by exposing the memory cells on the chip to ultraviolet rays through the crystal glass lid of the chip package. In an EEPROM chip, the data are electrically erased by a tunneling current. The erase operation is normally done for every unit of 8 bits. In particular, an

EEPROM in which all the data in a chip are simultaneously erased is called Flash memory. In EEPROM, there are two kinds of write-operation mechanisms: the injection of hot electrons, generated by avalanche breakdown phenomena at the drain of the cell transistor; and the injection of electrons generated by a tunneling effect.

Fuse ROM and Mask ROM

In a Fuse ROM cell, the write operation is accomplished by the blowing of program devices, such as fuses or p-n diodes connected to the cell transistor. Thus the on or off state of the cell transistor can be programmed according to the write data. The resultant destroyed program devices, however, permit only one write operation, as described before. In a Mask ROM (Figure 6), a mask pattern that programs the on or off state of each cell transistor is used. It offers the smallest cell area and ease of fabrication, allowing the largest memory capacity and the lowest cost, despite its limited function. [3]

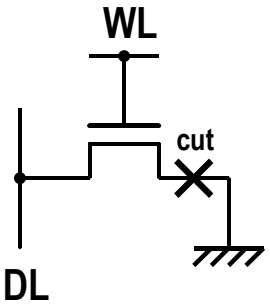


Figure 6: Mask ROM memory cell

D. DRAM

The operation principle of FeRAM is very similar to that of DRAM because both are charge-based memories, i.e., information is stored in terms of charge. Their memory cells also share the same configuration: a series connection of one transistor and one capacitor. The only difference is that the storage element used in FeRAMs is a ferroelectric capacitor. It is physically distinguished from a dielectric capacitor by substituting the dielectric with a ferroelectric material. Moreover, the memory architecture of present FeRAMs has been in large part derived from DRAM. Many popular circuits that are used for DRAM can be found in FeRAMs, too. The following section describes the operating principles and basic architecture of DRAM as well as the most basic circuits.

D.1 The Basic Operation of the DRAM Cell

Figure 7 shows a conceptual 1-T cell array of n rows by m columns, and an actual data line configuration. Multiple memory cells, a pre-charge circuit and equalizer (PE), and a latch-type CMOS sense amplifier (SA) are connected to each pair of data lines (DL) which communicate with a pair of common data input/output lines through a column switch. The 1-T cell operation comprises read, write, and refresh operations. All operations entail common operations: pre-charging all pairs of data lines to a floating reference voltage of a half-V_{DD} by turning off the pre-charge circuit and equalizer, and then activating a selected word-line. [3]

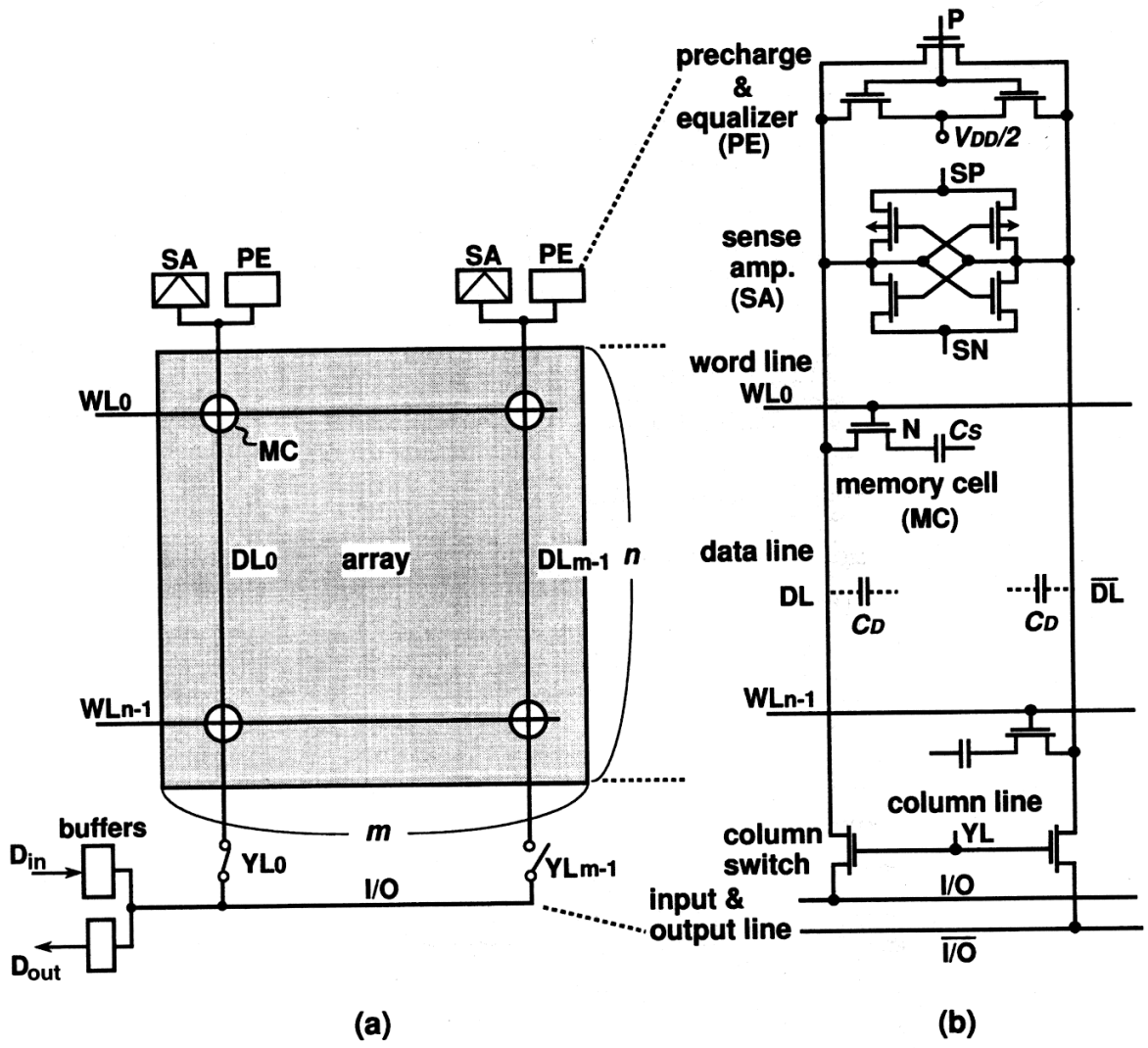


Figure 7: Conceptual DRAM array and data line configuration [3]

In the read operation (Figure 8), a stored data voltage, V_{DD} (="1") or V_{SS} (="0"), at the cell node (N) of each cell along the word-line is read out on the corresponding data line.

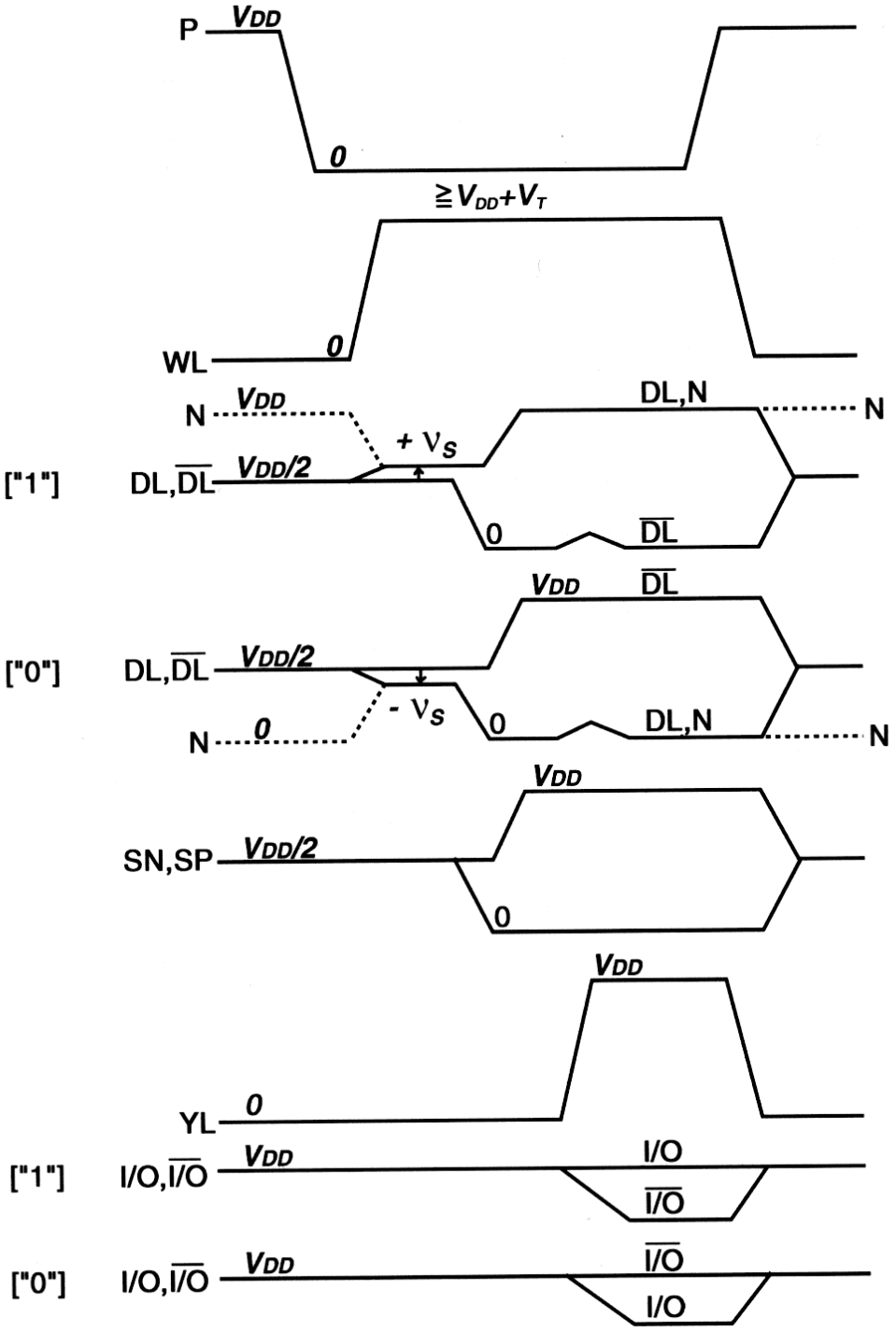


Figure 8: The DRAM read operation [3]

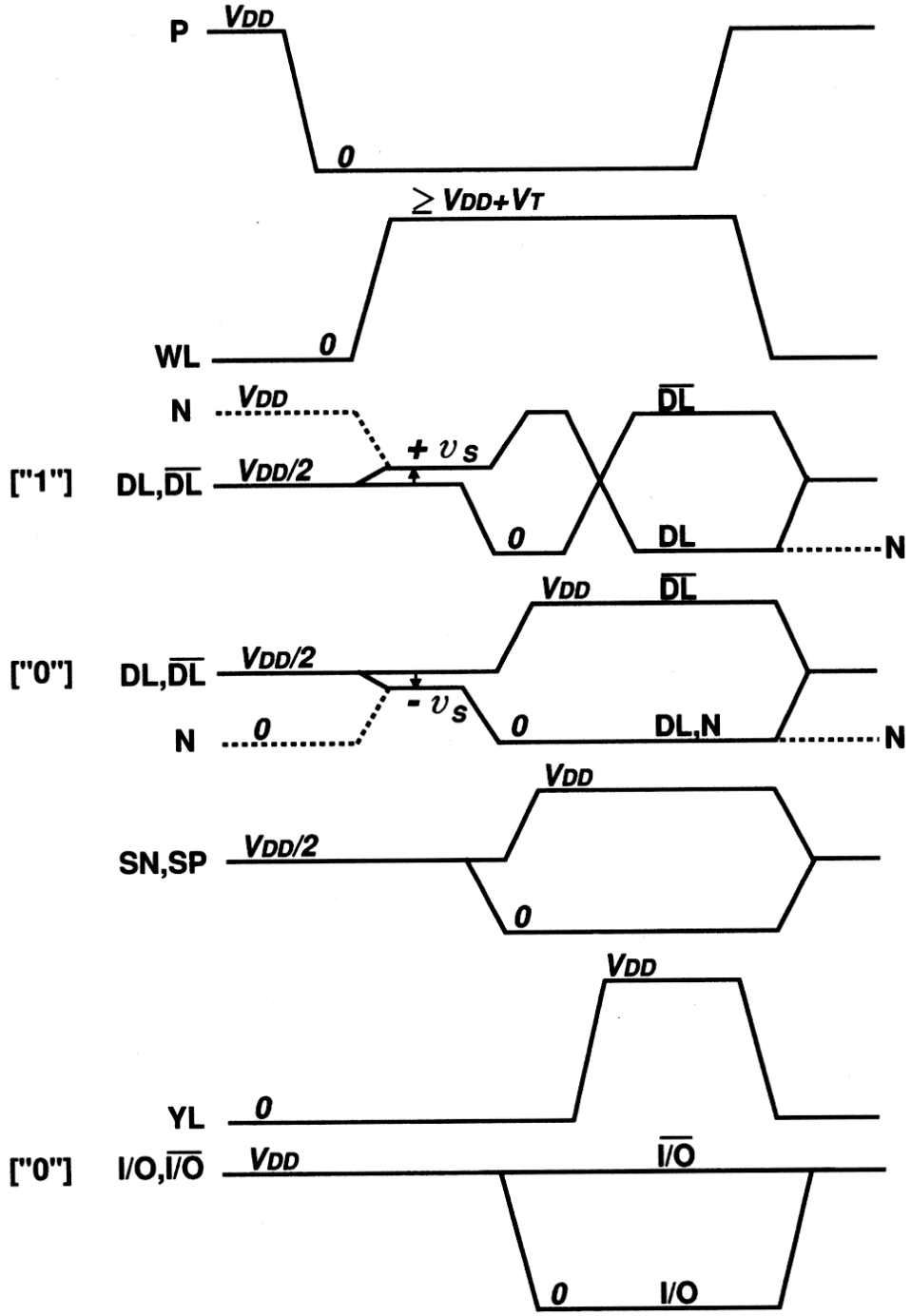


Figure 9: The DRAM write operation for "0" [3]

As a result of charge sharing, the signal voltage ($\pm v_S$) developed on the floating data line (for example, DL) is expressed by

$$\upsilon_S = \frac{V_{DD}}{2} \cdot \frac{C_S}{C_D + C_S}$$

Unfortunately, v_S is inherently small (100 - 200 mV) because the data line parasitic capacitance C_D is much larger than the cell storage capacitance C_S. A small C_S and large C_D result from the need for a small cell area and for connecting a large number of cells to a data line, respectively. Hence, the original large signal component ($V_{DD}/2$, usually 1.0 – 2.5 V) at the storage node collapses to v_S. The destructive readout (DRO) characteristics necessitate successive amplification and restoration for each one of the cells along the word-line. This is performed in parallel by a latch-type differential CMOS sense amplifier on each data line, with the other data line (DLB) used as reference. Then, one of the amplified signals is outputted as a differential voltage to the I/O lines by activating a selected column line, YL.

The write operation (Figure 9) is always accompanied by a preceding read operation. After almost completing the above amplification, a set of differential data-in voltages of V_{DD} and 0 V is inputted from the I/O lines to the selected pair of data lines. Hence, the old cell data are replaced by the new data. Note that the above read operation (i.e. amplification and restoration) is done simultaneously for each of the remaining cells on the selected word-line to avoid loss of information.

The stored voltage of each cell degraded by the leakage current is restored by a refresh operation that is almost the same as for the read operation, except that all YLs are kept inactive. This is done by reading the data of cells on the word-line and restoring them for each word-line so that all of the cells retain the data for at least t_{REFmax}. Here t_{REFmax} is the maximum refresh time for the cell, which is guaranteed in catalog specifications, and is exemplified by t_{REFmax} = 64 ms for a 64-Mbit chip. Thus, each cell is periodically refreshed at intervals of t_{REFmax}, although each cell usually has a data-retention time much longer than t_{REFmax}.

The fundamental circuit design issues of the 1-T cell can be summarized as the signal-to-noise ratio (S/N), power dissipation, and speed because of the following inherent cell characteristics [3]:

1. A small read signal, and relatively large levels of noise. Thus, the read operation is unstable unless a high S/N design is achieved. A small read signal is caused by the cell having no gain. There are many noise sources during the read operation:

- The difficulty in accommodating the sense amplifier and pre-charge circuit within the small layout pitch of the data lines tends to generate electrical imbalances onto a pair of data lines.
- A large number of sense amplifiers results in a large deviation in the threshold-voltage mismatch (the offset voltage) between pairs of transistors in the sense amplifier.
- Simultaneous charging and discharging of many heavily capacitive data lines with a high voltage invariably introduces many kinds of noise.
- In addition, cell leakage currents and alpha particle hits, which degrade the stored charges, effectively work as noises.

2. The above charging and discharging of data lines also causes high power dissipation.
3. Slow sense amplifier operation. The relatively poor driving capability of the sense amplifier, which stems from the need for a small area, and operation based on a low voltage of a half-V_{DD}, makes the sense amplifier operation slow. Thus, the sensing time is the largest component of the access time for the chip.

D.2 The Multi Division of a Memory Array

To realize stable operation with a higher S/N ratio, lower power dissipation, and a higher speed, multi-divisions of both the data line and the word line are essential. The resulting multi-divided memory array provides high performance despite an increased memory capacity. When the number of divisions increases, however, the total memory array area increases because an additional area is necessary at each division. Figure 10 shows an example, in which the memory array is divided in 16 sub-arrays, with four divisions in both the data line and the word-line. Obviously, the bundles of row and column decoders increase at every division, causing an increased chip area. In addition, the resulting increased load capacitance of the address line reduces the speed of the address buffer. Up to the 4-Mbit generation, such a simple division had been popular as a result of giving the first priority to a simpler fabrication process.

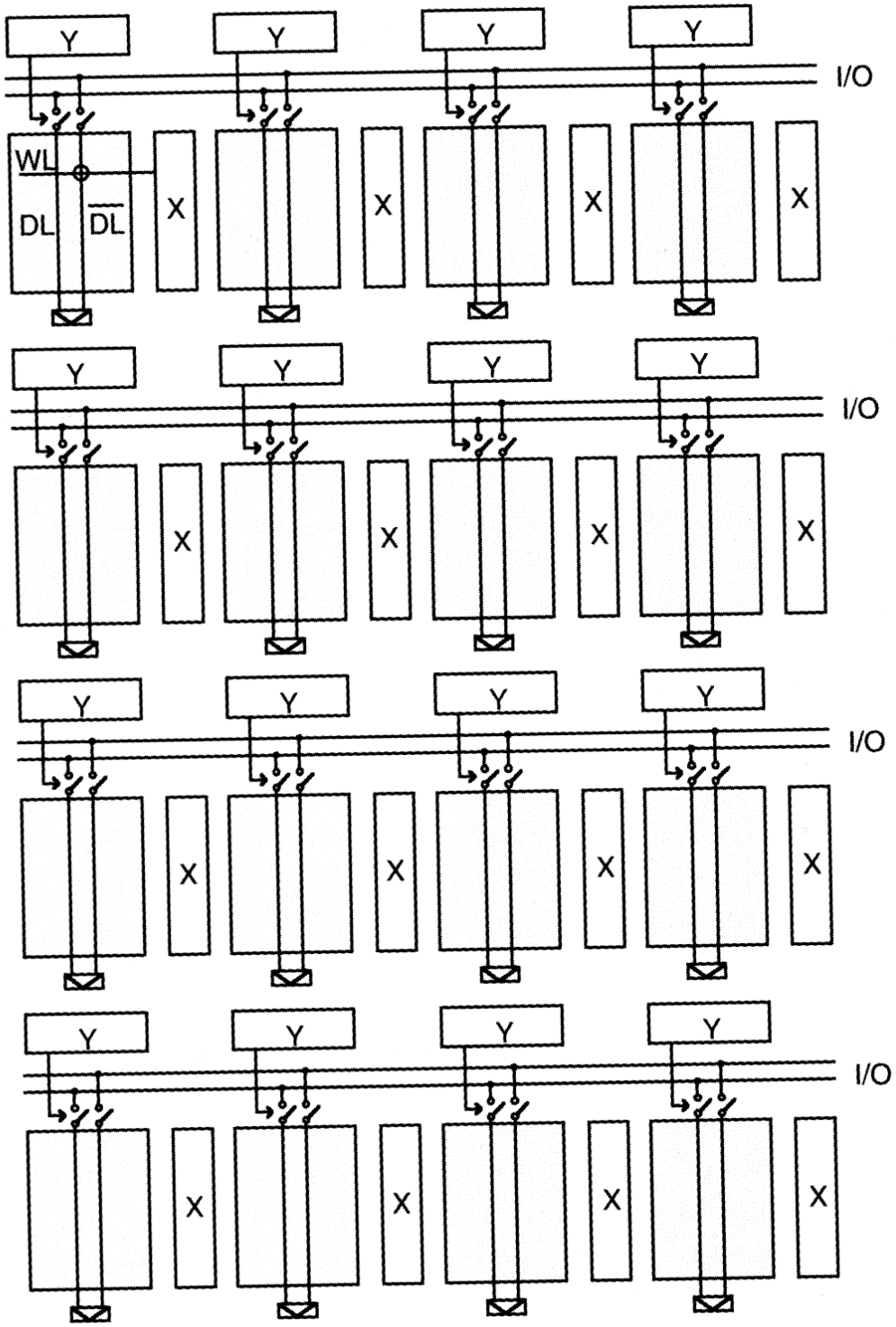


Figure 10: Simply divided memory cell arrays [4]

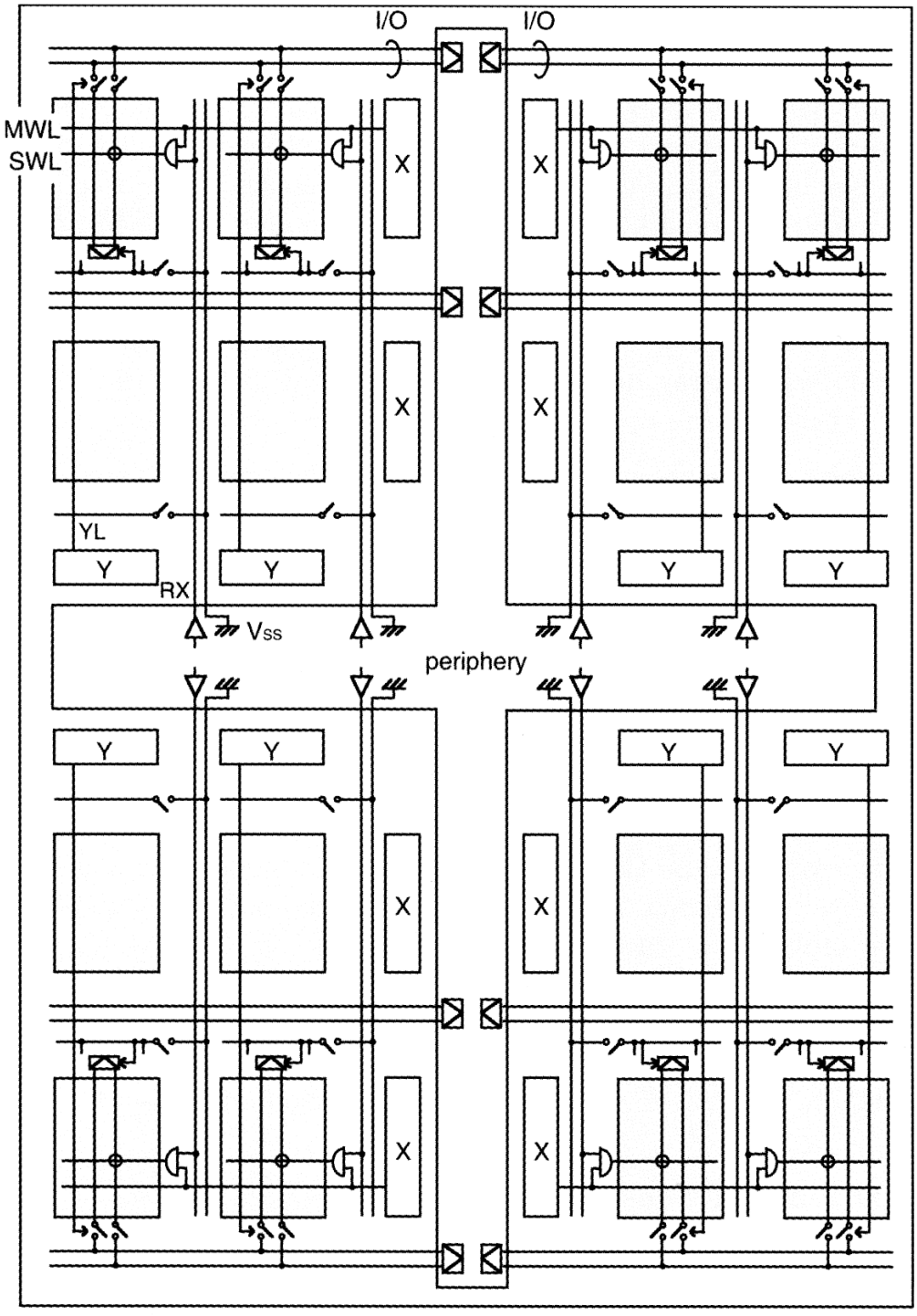
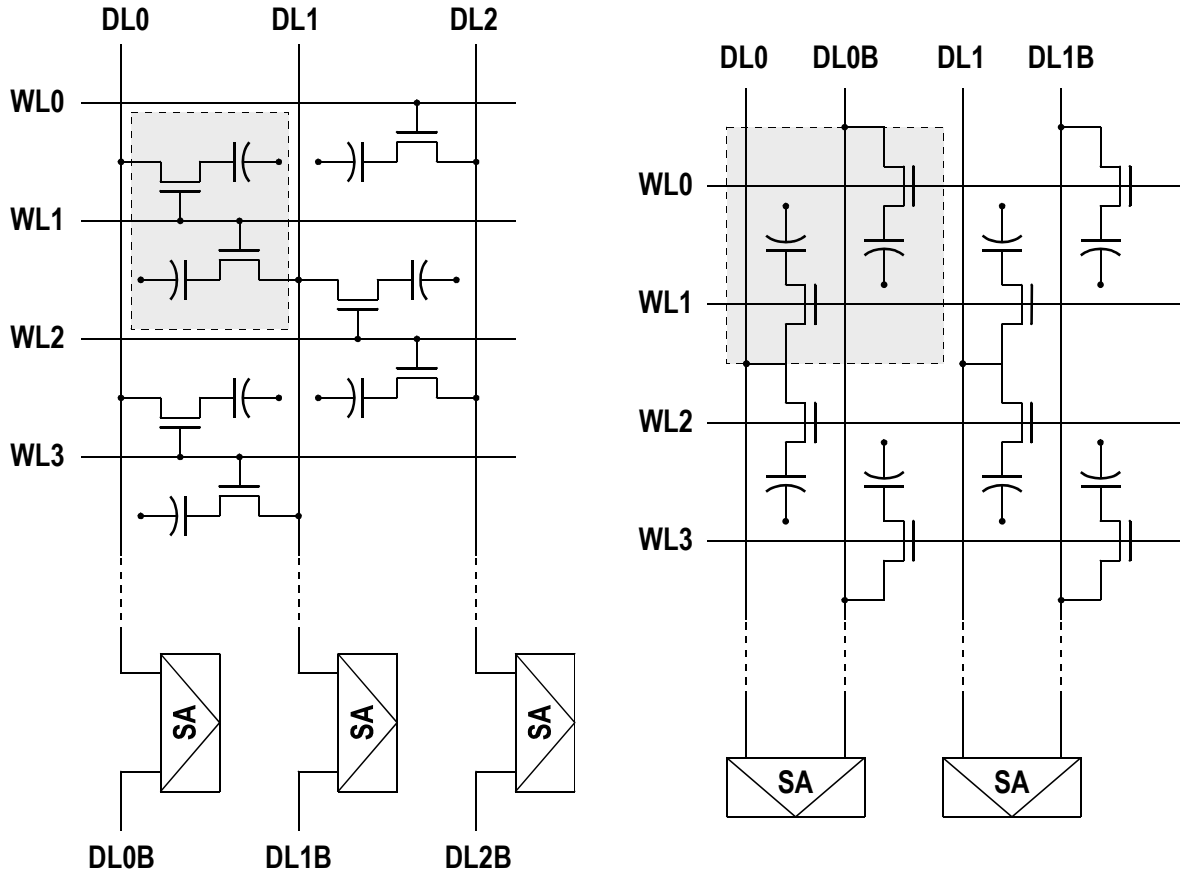


Figure 11: Multi-divided memory-cell arrays featuring shared decoders, distributed SA driving and cross-shaped layout areas for peripheral circuit blocks. [4]

Recently, however, multilevel metal wiring, using Al or W, has been widely used to suppress the increase in the area due to the division and to obtain a higher S/N ratio and a higher speed. Multilevel metal wiring offers new multi-divided array architec-

tures to minimize the number of decoders, and realizes low-resistance layouts for power and signal lines. Figure 11 shows one of the most advanced multi-divided arrays [4], although many kinds of array divisions have been proposed. [3]

D.3 Data Line Arrangements



(a) open *(b) folded*

Figure 12: Data line arrangements

Figure 12 shows two kinds of data line arrangements, the open data line arrangement and the folded data line arrangement. The open data line arrangement allows a very small $6F^2$ memory-cell. However, because data line and data line bar lie in different sub-arrays, there is likely to be a capacitance mismatch between these two. In addition, coupled differential noise can become a serious problem. For the folded data line, the differential noise is lowered because of cancellation at two crossing points, where a word line and a pair of data lines intersect. To reduce differential noise and capacitance mismatch the folded data line arrangement was widely adopted. Unfortunately,

this also resulted in the larger $8F^2$ cell that can be found in today's DRAMs. At present, there is a strong desire to go back to the smaller $6F^2$ cell [6].

D.4 Sensing and Amplification

A cross-coupled differential amplifier connected to a pair of data lines can amplify a small signal voltage on one data line up to V_{DD}, with respect to a reference voltage on the other data line. Thus, the read information is discriminated by the polarity of the read signal for the reference voltage (V_{ref}), as shown in Figure 13, while any common-mode voltage coupled to both data lines is canceled. The discrimination stability is closely related to the data line arrangement and the reference voltage generation. [3]

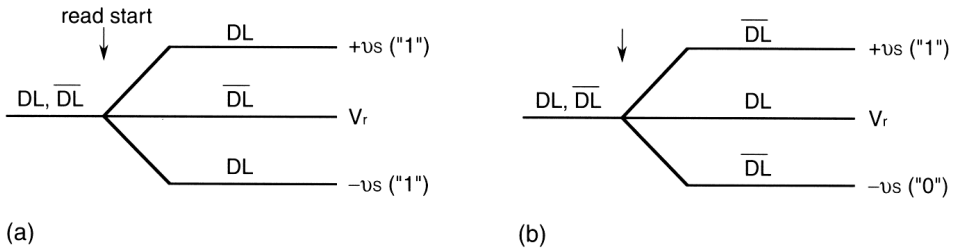


Figure 13: A signal voltage on a pair of data lines [4]

Figure 14 shows half-V_{DD} pre-charge using a CMOS cross-coupled amplifier that works as a sense amplifier and an active restoring circuit. Fast SNL activation is shown in the figure. The CMOS amplifier can be regarded as a parallel connection of an NMOS cross-coupled amplifier and a PMOS cross-coupled amplifier. Thus, the NMOS amplifier tends to strongly discharge the data line to a lower voltage, while the PMOS amplifier tends to strongly charge up the data line to a higher voltage. Consequently, a small differential signal voltage is finally amplified to a large differential voltage of V_{DD}. Note that when the PMOS amplifier is activated, DL bar is momentarily charged up by Q_P bar. However, the charging up is suppressed by the ever-rising DL voltage, and thus the increasing Q_N bar transconductance. In practice, both amplifiers are activated, with a short time interval at high speed, despite the increased dc current that traverses both amplifiers. In principle, activation of the PMOS amplifier could precede that of the NMOS amplifier unless the noise generated in an NMOS

memory-cell array and the offset voltage of the PMOS amplifier are risks. A dummy cell is usually unnecessary. [3]

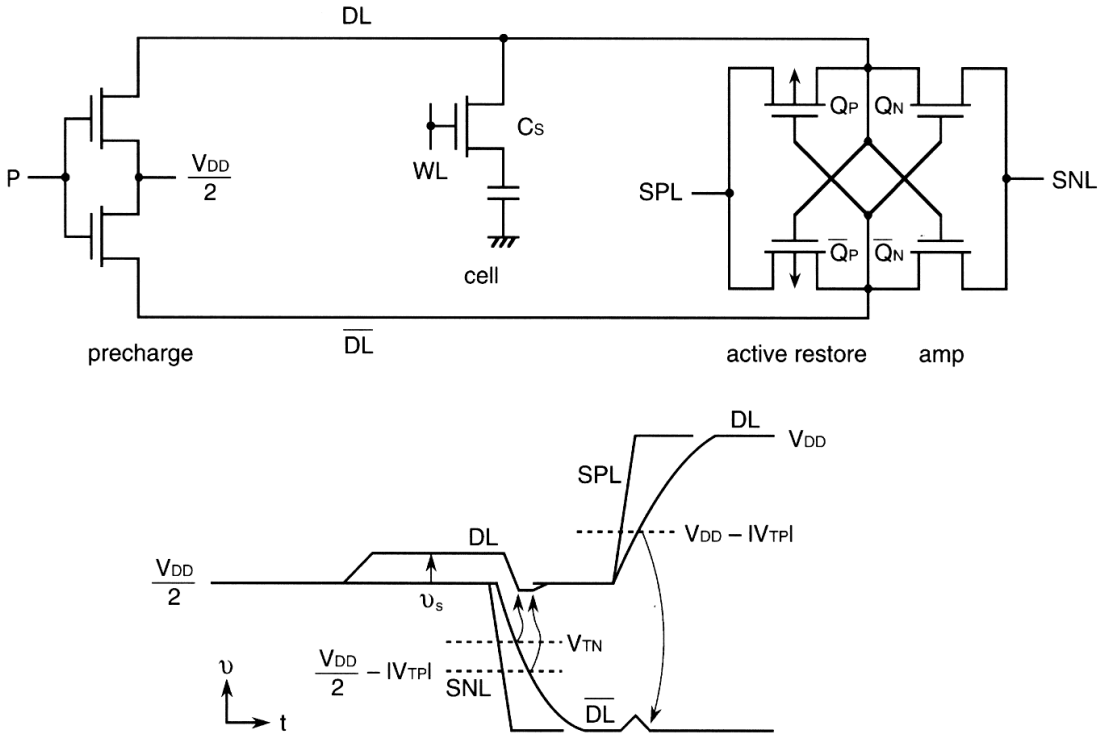


Figure 14: DRAM sense amplifier and amplification mechanism for half-V_{DD} pre-charging [4]

E. State-of-the-Art and Short-Term Trends

Figure 15 shows trends in memory cell area for various memories, which have been presented at major conferences. Both DRAM and SRAM cells have been miniaturized at a pace of about one-fiftieth per 10 years. Recently, however, there has been a saturation due to the ever-more-difficult process of device miniaturization. Flash memory is about to catch up with DRAM, by using a one-transistor, one capacitor cell (the 1-T cell). Figure 16 shows trends in the memory chip capacity of R&D prototype memories. DRAM has quadrupled its memory capacity every 2.5 years, although in production parts it has quadrupled only every 3 years. As a result, commodity and stand-alone DRAMs have reached 1-4 Gbit in R&D, and 64-256 Mbit in volume-production. In addition, the throughputs have been boosted, as exemplified by the 1.6 Gbit/s 1-Gbit chip [7], by new memory-subsystem architectures.

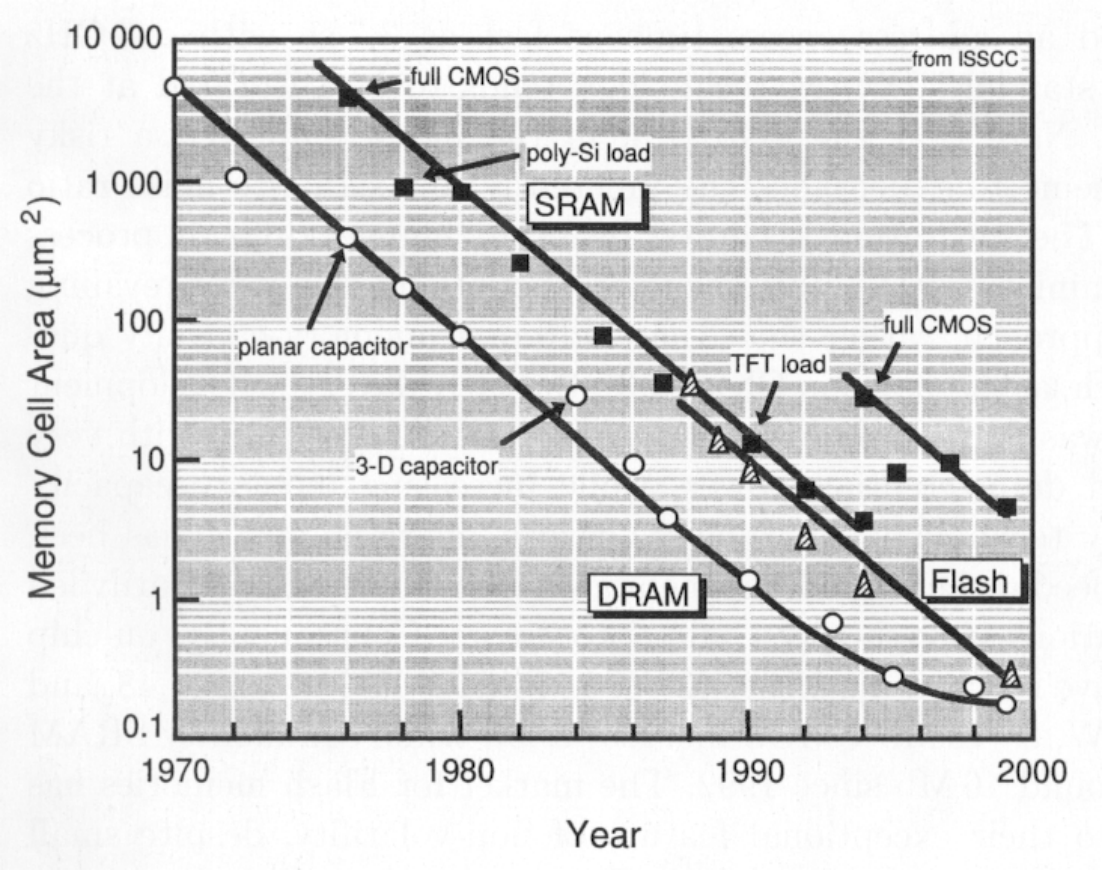


Figure 15: Trends in the memory-cell area of VLSI memories [4]

A 8-Mbit embedded DRAM [8] using the standard 1-Gbit DRAM technology has also revealed an address access time as fast as 3.7 ns with a 1 GHz clock. However, standard DRAM technology tends to be saturated at the 1-Gbit generation. Note that the 4-Gbit [9] in the figure employed a risky 2-bit-per-cell scheme that degrades the inherently low signal-to-noise ratio of the 1-T cell. The saturation is caused by both the ever-more-difficult process of memory cell miniaturization (as discussed above) and the ever-prevailing chip-shrinking approach, rather than the traditional memory-capacity quadrupling approach. In the early days, the development of SRAM chips was focused on low-power applications, especially with very low standby and data-retention power, while increasing memory capacity with high-density technology. Nowadays, however, more emphasis has been placed on high speed rather than large memory capacity, a trend primarily led by cache applications in high-speed microprocessors. Consequently, on-chip SRAM caches have reached a 0.55 ns, 43 W, 1-Mbit BiCMOS macro [10] and 1.8-3.4 ns, 1-7 W, 8-18 Mbit CMOS macros [11], while com-

modity SRAM has stayed around 16-Mbit since 1992. The market for Flash memories has expanded, due to their exceptional feature of non-volatility, despite small memory-chip capacities of 16-Mbit or less. Front-end commercial chips of 64 and 256-Mbit [12,13] will soon help to expand the market as Flash memories almost catch up with DRAMs in memory capacity, with their inherently small cell structures. The pace of increase in memory capacity of Flash memory will eventually follow the trend in DRAM, due to the limitations of their common lithography technology, although it has skyrocketed impressively since the advent of Flash memory.

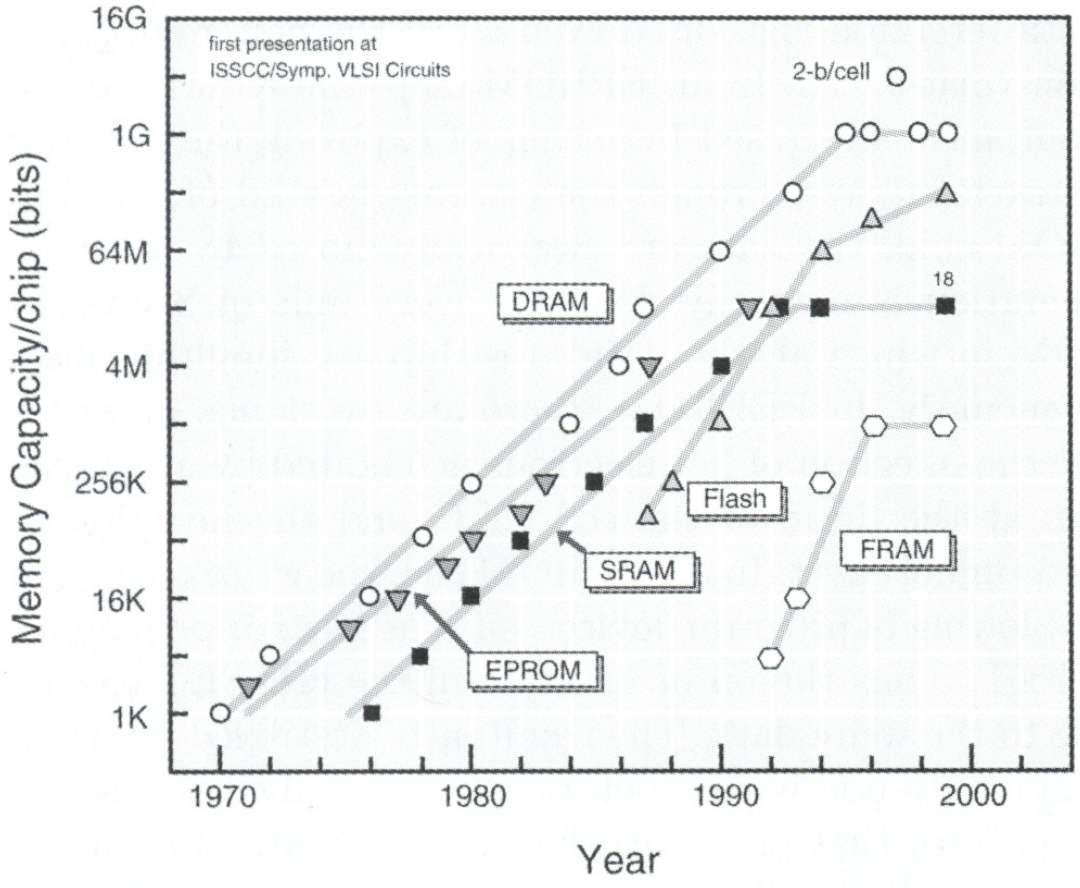


Figure 16: Trends in the memory capacity of VLSI memories [4]

III. REVIEW OF PREVIOUS FERROELECTRIC MEMORY TECHNOLOGY

A. FERROELECTRIC MATERIALS

"Ferroelectrics are substances in which a spontaneous polarization occurs in a certain range of temperature and the direction of this polarization can be changed by an electric field." [14]

A.1 Electrical Properties

The hysteresis loop shown in Figure 17 is the most characteristic electrical property of a ferroelectric. It describes the relationship between the resulting polarization P and the applied electric field E and is similar to the hysteresis loop of ferromagnetic materials from where the name stems.

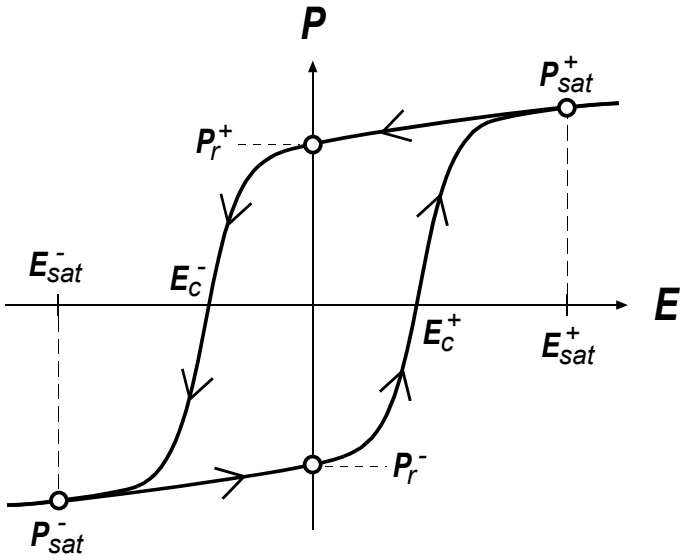


Figure 17: Ferroelectric hysteresis loop

As indicated by the arrows on the hysteresis loop, the ferroelectric can be brought into one of the two stable states marked as P_r^- and P_r^+ at E=0, i.e., when the electric field is zero. P_r^- and P_r^+ are the negative and positive remanent polarizations. For example, if

an external field of $E=E_{sat}^{+}$ is applied to the ferroelectric, it is polarized to $P=P_{sat}^{+}$. P_{sat}^{+} is the positive saturation polarization. After the external field is removed (i.e. E=0), the polarization settles back to $P=P_r^{+}$. In the opposite case where an external field of $E=E_{sat}^{-}$ is applied, the ferroelectric is polarized to $P=P_{sat}^{-}$. After the field is removed, the polarization settles back to $P=P_r^{-}$. Thus the polarization of a ferroelectric depends not only on the present value of E but also on its previous values. The points where the hysteresis loop crosses the x-axis are marked as E_c^{-} and E_c^{+}. E_c^{-} and E_c^{+} are the negative and positive coercive fields, respectively.

A.2 Materials for Memory Applications

Two ferroelectric materials that are most promising for memory applications are lead zirconate-titanate, $Pb(Zr_xTi_{1-x})O_3$ (PZT) and strontium bismuth tantalite, $SrBi_2Ta_2O_9$ (SBT) layered perovskite. Both belong to the family of perovskite crystals. The spontaneous polarization in perovskites is caused by a displacement of the cation from its central position in the middle of the oxygen octahedra. As shown in Figure 18, the cation resides in one of the two stable positions, and the position is reversible by an electric field.

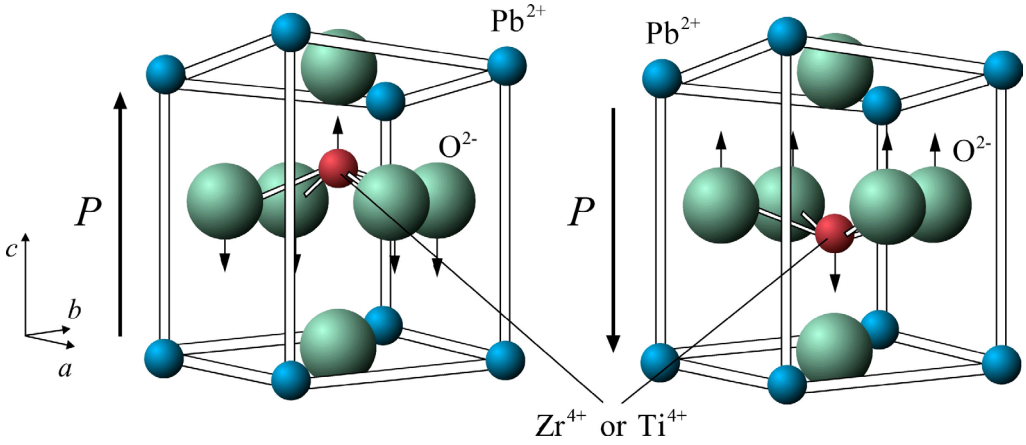


Figure 18: Perovskite structure of tetragonal Pb(ZrTi)O3

PZT is a solid solution of lead zirconate, $PbZrO_3$, which is anti-ferroelectric, and lead titanate, $PbTiO_3$, which is ferroelectric. The ratio of $PbZrO_3$ to $PbTiO_3$ determines the properties of PZT, as shown in the phase diagram in Figure 19. Since PZT undergoes a phase transition at T_C, the Curie temperature; it exhibits ferroelectricity only below

T_C. PZT compositions with a larger fraction of $PbTiO_3$ (e.g. $x < 0.4$) achieve higher polarization values due to the higher polarization of 81 $\mu C/cm^2$ for pure $PbTiO_3$ and are therefore preferable for memory applications. A remanent polarization between 18 and 30 $\mu C/cm^2$ is typical for 40/60 PZT.

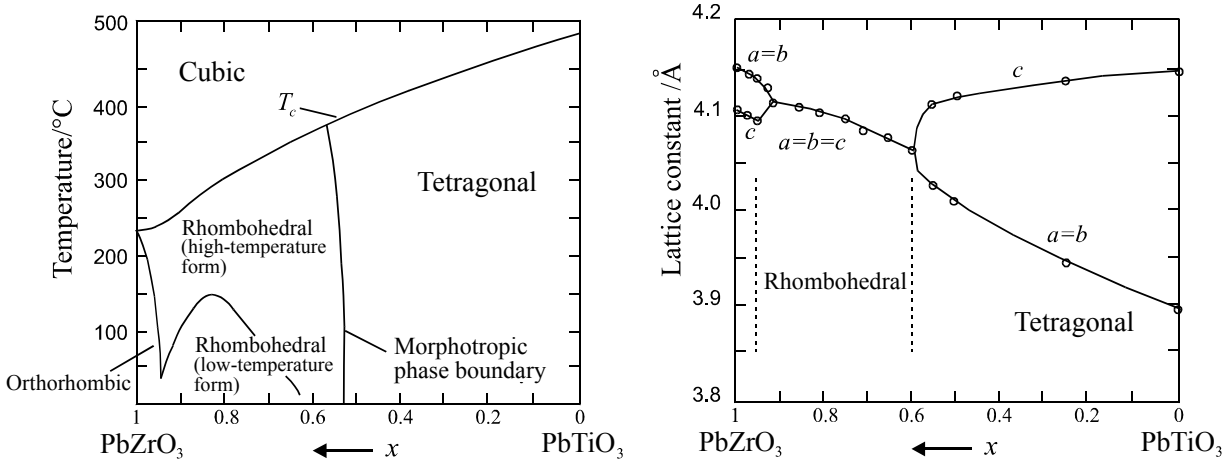


Figure 19: The phase diagram for $Pb(Zr_xTi_{1-x})O_3$ (PZT)

A.3 Fatigue, Imprint and Retention Loss

Three major material issues that are known to cause failure in ferroelectric memories are fatigue, imprint and retention loss.

Fatigue is one of the most serious issues and is observed as a decrease of remanent polarization with the number of read/write cycles (see Figure 20). The stress induced by the continuous switching between positive and negative polarization states causes the ferroelectric material to fatigue.

The loss in remanent polarization depends on switching frequency, amplitude, temperature, etc. and may be as large as 50% or more after only 10^8 cycles for PZT with platinum (Pt) electrodes. A loss of P_r directly translates into a loss of signal voltage and ultimately leads to the inability of the sense amplifier to discern between the two different polarization states. It has been found that the electrode material

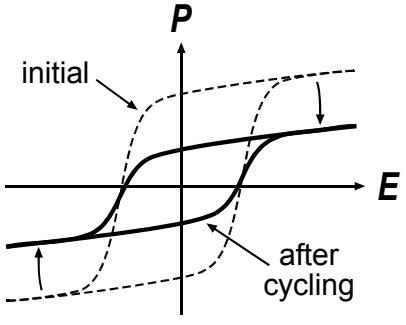


Figure 20: Fatigue

plays a crucial role in fatigue performance. For example, the recent change from Pt to IrO_2 and RuO_2 electrode material allowed extending the point where a loss of P_r starts to appear up to ~10^{12} cycles. Another major issue is known as *imprint* and is described as the preference of one polarization state over the other [15]. In general, two different effects are observed. First, imprint leads to a shift of the hysteresis loop on the voltage axis, and second, to a loss of the polarization state complementary to the established one. Hence, both effects can cause a loss of signal voltage and ultimately failure of the ferroelectric memory cell. Memory failure caused by imprint is also further investigated in Chapter IV.

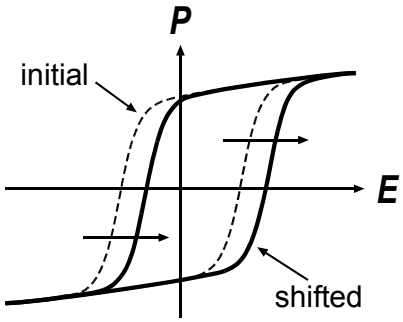


Figure 21: Imprint

Immediately after a write operation, the polarization inside the ferroelectric capacitor is either P_r^- or P_r^+. However, it has been observed that the ferroelectric is not able to preserve the initial remanent polarization over time without loss. This loss in remanent polarization is called *retention loss*. Although less probable than memory failure due to fatigue or imprint, retention loss may lead to memory failure as well.

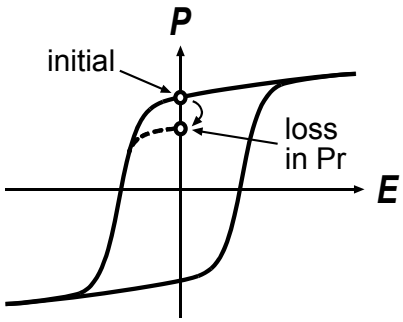


Figure 22: Retention

B. Ferroelectric Memory

A conventional FeRAM has memory cells containing ferroelectric capacitors. Each ferroelectric capacitor contains a ferroelectric thin film sandwiched between two conductive plates.

B.1 The Ferroelectric Capacitor

The typical Q-V characteristic of a ferroelectric capacitor is shown in Figure 23. To store data in a ferroelectric capacitor, a write operation applies a write voltage (typically $-V_{DD}$ or $+V_{DD}$) to the ferroelectric capacitor to polarize the ferroelectric material

in a direction associated with the data being written. After the write voltages are removed a remanent polarization of P_r^+ (P_r^-) remains in the ferroelectric material, which in turn maintains a complementary charge of Q_r^- (Q_r^+) on the conductive capacitor plates. This property is used for the non-volatile storage of information. For readout, one plate of the ferroelectric capacitor is connected to a floating data line while the other one is raised to a read voltage (typically V_{DD}). If the remanent polarization in the ferroelectric capacitor is in a direction corresponding to the read voltage, the read voltage causes a relatively small current through the ferroelectric capacitor, resulting in a small voltage change on the data line. If the remanent polarization initially opposes the read voltage, the read voltage flips the direction of the remanent polarization, discharging the plates and resulting in a relatively large current and voltage increase on the data line. Unfortunately, the direction of the polarization can only be detected by a destructive read operation (DRO). Therefore, a write-back of the original data after read-out is required.

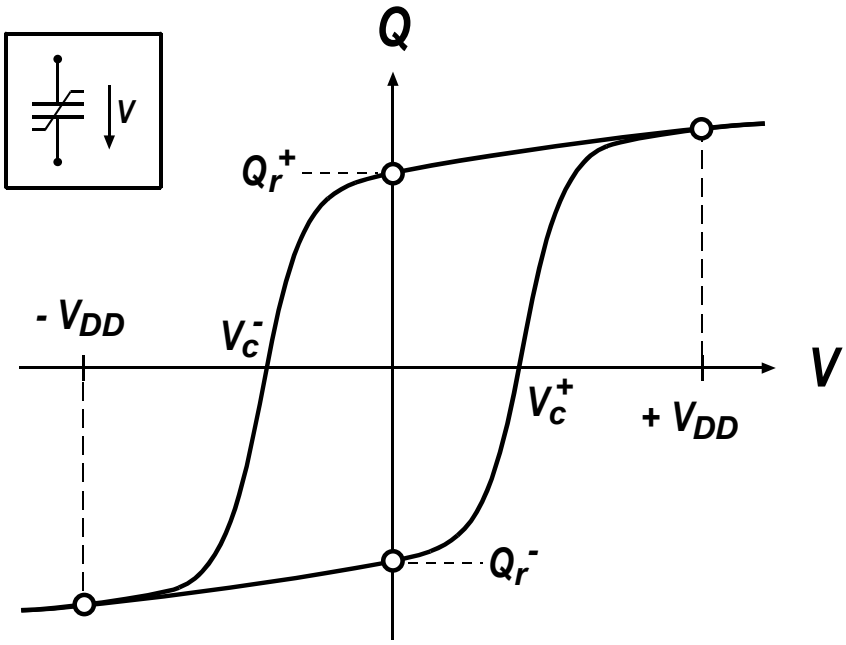


Figure 23: Symbol and Q-V characteristic of ferroelectric capacitor

B.2 Ferroelectric Memory Cells

The 1T1C cell

The memory cell of a ferroelectric memory consists of an access transistor connected in series with a ferroelectric capacitor (Figure 24). This configuration is similar to the 1-T DRAM cell and is known as the 1T1C cell. In contrast to the DRAM cell, the second plate of the capacitor is also connected to a signal line, which is usually called the drive- or plate-line (PL). During operation, this line is pulsed from V_{SS} to V_{DD} to improve the voltage range that can be applied to the ferroelectric capacitor from $\pm\frac{1}{2}V_{DD}$ to $\pm V_{DD}$. A constant plate voltage of $V_{DD}/2$ would only give a voltage range of $\pm\frac{1}{2}V_{DD}$. Sample write- and read-operations with this memory cell are described next.

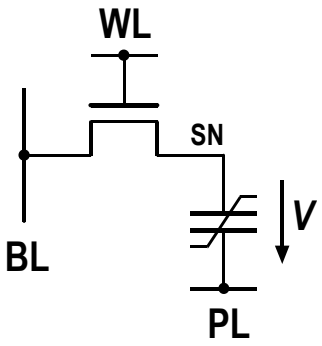


Figure 24: FeRAM memory cell (1T1C)

Write and read operation

The write access is initiated by driving the bit-line to either V_{SS} or V_{DD} depending on whether a "0" or "1" is to be written (Figure 25). Here we assume that the "0" corresponds to the polarization state that generates the non-switching charge and the "1" to the polarization state that generates the switching charge. The write is performed by taking the word-line to V_{PP}, followed by pulsing the plate-line from V_{SS} to V_{DD}. V_{PP}, the boosted word-line voltage, must be larger than $V_{DD} + V_{th}$ to prevent degradation of V_{DD} from BL to the storage node SN. Finally, the bit and word-line are brought back to V_{SS}.

The read access begins with disabling the bit-line pre-charge, after which the bit-line is floating at near V_{SS}. To read out the stored information, the word-line is activated and the plate-line is pulsed to V_{DD}. At this time, charge from the ferroelectric capacitor is dumped onto the bit-line. The amount of charge depends on the previous polarization direction of the ferroelectric. The bit-line voltage is established according to the capacitor divider that is formed by the ferroelectric capacitor and the parasitic bit-line capacitance, C_{BL}, between the plate-line and ground. Next, the voltage on the

bit-line is compared against a reference voltage by a sense amplifier and the bit-line is restored to either a full V_{SS} or a full V_{DD} level, depending on the outcome of the comparison. This is also the beginning of the write-back phase. Write-back is accomplished, once the plate-line is returned to V_{SS}. Finally, the bit-line is discharged to V_{SS} and the word-line is turned off.

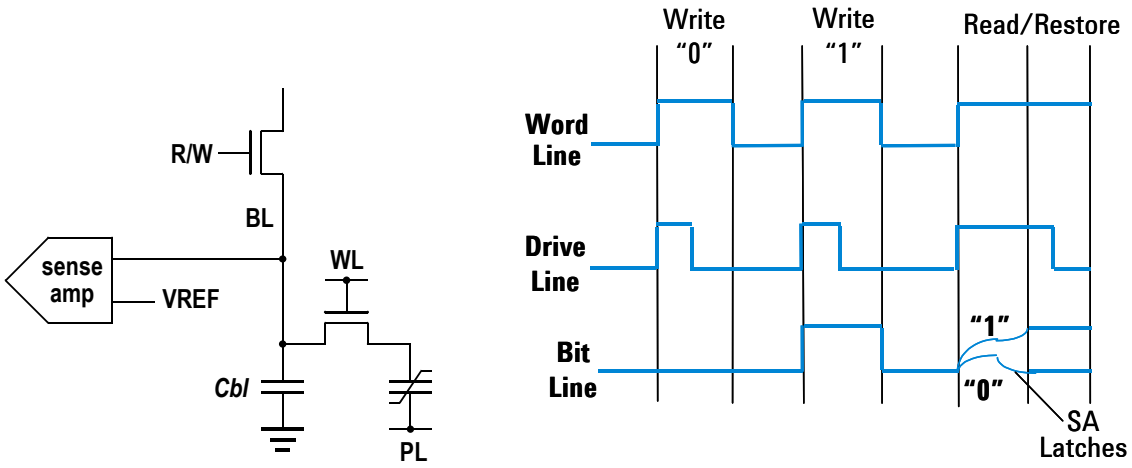


Figure 25: Simplified FeRAM write- and read-operation

Typically, the reference voltage is common for all the storage cells in a memory. If the polarization in the cell capacitor produces a bit-line voltage that is greater than the reference voltage, the cell is deemed to be in the "1" polarization state and if it is smaller than the reference voltage, the cell is deemed to be in the "0" polarization state. The amount of remanent polarization in a ferroelectric capacitor typically varies due to variations in the manufacturing process as well as material fatigue over time and other factors. A relatively high variation increases the difficulty in generating a reference voltage, which is suitable for reading all the storage cells in a memory. As of today, most commercially available FeRAM products employ a memory cell, which is known as the 2T2C cell (2 transistors, 2 capacitors).

The 2T2C cell

The 2T2C cell comprises two 1T1C cells that have a common word- and plate-line but are connected to two separate bit lines (Figure 26). This memory cell does not require a reference voltage because it stores both data and data complement. For a read, the charge of both ferroelectric capacitors is simultaneously dumped onto bit-line and bit-

line bar (BLB) and a sense amplifier performs a differential comparison between the voltages on the bit lines. This memory cell is less susceptible to problems caused by manufacturing variations and material fatigue because such variations tend to influence the stored polarization in a complementary fashion.

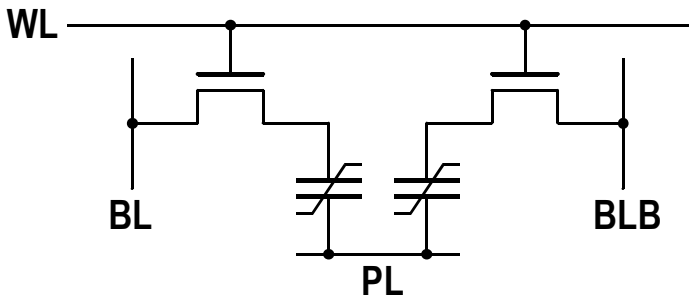
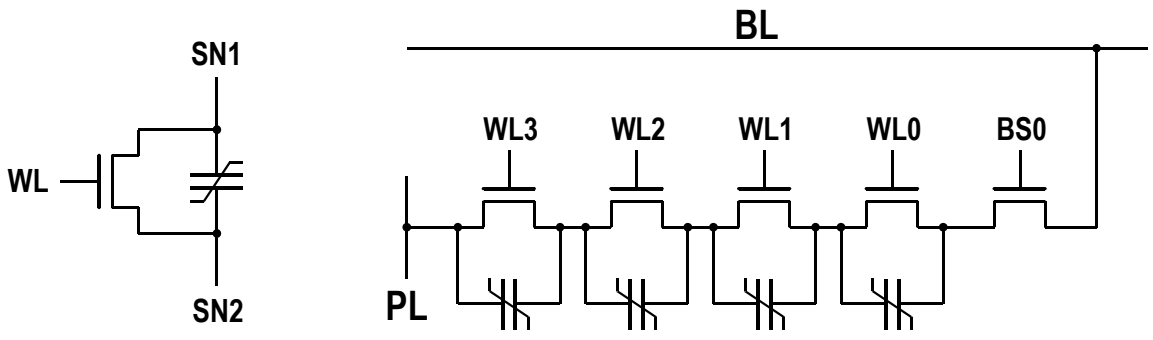


Figure 26: FeRAM memory cell (2T2C)

The CFRAM cell

To further improve density and speed, a chain-type ferroelectric memory (CFRAM) has been proposed [16]. In contrast to conventional FeRAM, the transistor and capacitor of the CFRAM cell are connected in parallel instead of in series (Figure 27a). Single cells are then lined up to a chain (Figure 27b) enabling a very compact layout (sharing of diffusion regions), thus reducing the average area per bit compared to conventional FeRAM. As cell size is a key parameter for semiconductor memory, CFRAM appears to be a leading candidate for high-density FeRAMs.



(a) *(b)*

Figure 27: CFRAM memory cell and cell chain

On the other hand, the serial connection of cells generates several new design issues, which have to be taken into account and are discussed in Chapter IX.

C. CIRCUIT DESIGN OF 1T1C FERAM

Due to the many analogies between DRAM and FeRAM, several design issues for FeRAM are already known from DRAM and have been solved by applying prior DRAM solutions. However, there are also a number of issues that are unique to FeRAM. Often these issues become critical design problems that require innovative solutions. This chapter presents some of the most relevant issues affecting present and future deep sub-micron FeRAMs.

C.1 Memory Cell Design

The memory cell itself has the largest impact on chip size and many other chip properties. Therefore finding the optimum cell configuration is crucial for overall chip performance. The most important criteria are cell size, capacitor size and cell efficiency. Second priority criteria include maximizing bit- and word-line pitch and minimizing parasitic capacitances.

Capacitor size considerations

Determining the adequate capacitor size for a given number of cells per bit-line or vice versa is not trivial for FeRAM. In contrast to DRAM, increasing the size of the cell capacitor does not always achieve a larger signal voltage. This is because during readout the plate-line voltage (e.g. V_{CC}) is shared according to the capacitor divider that is formed by the cell capacitor and the parasitic bit-line capacitance C_{BL} (see page 29ff.). In order to switch the polarization inside the ferroelectric, the voltage across the ferroelectric capacitor V_{FE} must exceed the coercive voltage V_C, hence, $V_{FE} >= V_C$. This circumstance imposes a first requirement for the bit-line capacitance that is fulfilled if the following relation is satisfied [17]:

$$V_{FE} \geq V_C \quad \Rightarrow \quad C_{BL} \geq \frac{V_C \cdot C_S}{V_{CC} - V_C} \tag{1}$$

(Note: C_S is a dielectric capacitor that represents only the dielectric charge of the bit cell. It does not include the remanent charge Qr [17].)

In practice, the above equation only represents an absolute minimum requirement and typically $V_{FE} = V_C$ is not sufficient. As a rule of thumb, the voltage across the capacitor – after the capacitor has dumped its signal charge (= $2Q_r$) onto the bit-line – should be twice as large as its coercive voltage in order to drive the ferroelectric into full saturation [18]. Otherwise, the signal voltage will be too small. A stricter requirement for C_{BL} is therefore:

$$C_{BL} \geq \frac{2Q_r + 2V_C \cdot C_S}{V_{CC} - 2V_C} \qquad (1^*)$$

On the other hand, the bit-line capacitance has to remain small enough in order to allow a large enough signal voltage to be established. The minimum signal voltage V_{SE} at which the sense amplifiers can correctly discern data is typically about 70 mV. This requires a second relation to be satisfied [17,18]:

$$C_{BL} \leq \frac{Q_r}{V_{SE}} - C_S \qquad (2)$$

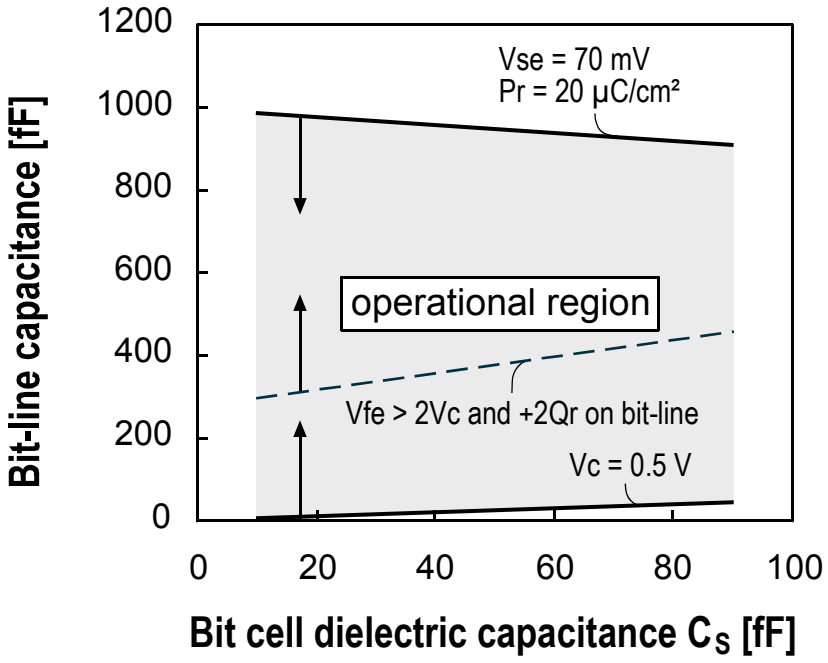


Figure 28: Relationship to be satisfied by C_{BL} and C_S (V_{CC} = 1.5 V)

The relationship between C_{BL} and C_S is illustrated in Figure 28. It may be confusing that the smaller C_S becomes, the larger C_{BL} is allowed to be. This is correct since in relation (2) it was assumed that Q_r remains constant while C_S changes. Again, C_S represents only the dielectric charge of the ferroelectric capacitor. However, this awkward situation might be misleading and suggest the conclusion that a smaller cell capacitor may allow a larger bit-line capacitance. Of course, this is not the case, which can be easily shown if C_S and Q_r are expressed with their dependence on capacitor area A:

$$Q_r = P_r \cdot A \quad \text{and} \quad C_S = \frac{P_{ns} \cdot A}{V_{CC}}$$

P_r is the remanent polarization and P_{ns} is the non-switching polarization of the ferroelectric capacitor. Now, relations (1) and (2) can be transformed into:

$$C_{BL} \geq \frac{V_C \cdot P_{ns}}{(V_{CC} - V_C) \cdot V_{CC}} \cdot A \tag{3}$$

$$C_{BL} \leq \left(\frac{P_r}{V_{SE}} - \frac{P_{ns}}{V_{CC}} \right) \cdot A \tag{4}$$

This illustrates that a smaller capacitor also requires a smaller bit-line capacitance. The relationship between bit-line capacitance and capacitor area is shown in Figure 29.

Signal voltage optimization

The later representation is very useful for initial memory cell design as it quickly allows one to visualize the boundaries for bit-line capacitance and cell capacitor area, A. However, it is still unclear how to determine the optimum values for C_{BL} and A within these boundaries in order to maximize the signal voltage. This step requires the use of a ferroelectric capacitor model, as presented later in Chapter IV, and a circuit simulator. The signal voltage for different values of bit-line capacitance and capacitor area can be obtained by simulation. Typically, the signal voltage exhibits a maximum if plotted versus bit-line capacitance or capacitor area (Figure 30).

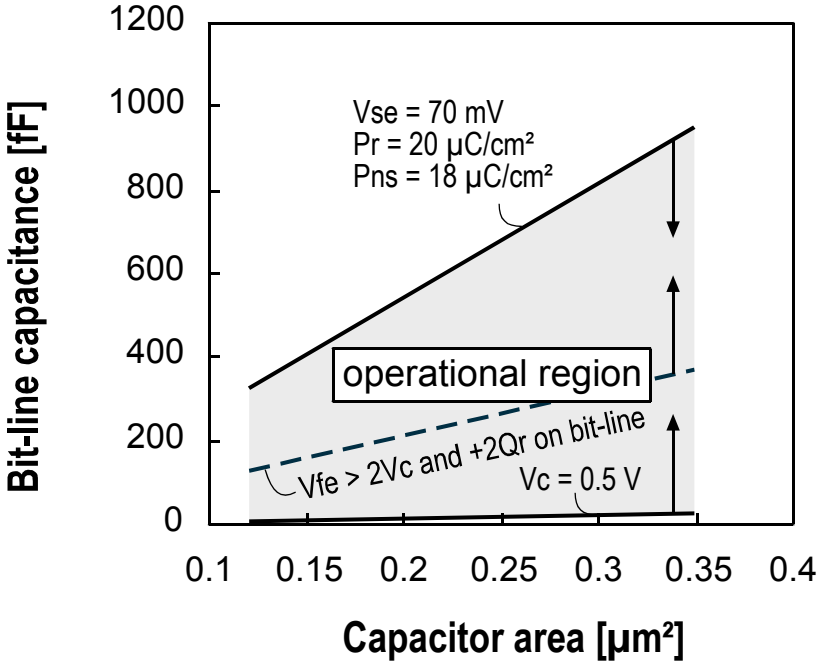


Figure 29: Relationship to be satisfied by C_{BL} and A (V_{CC} = 1.5 V)

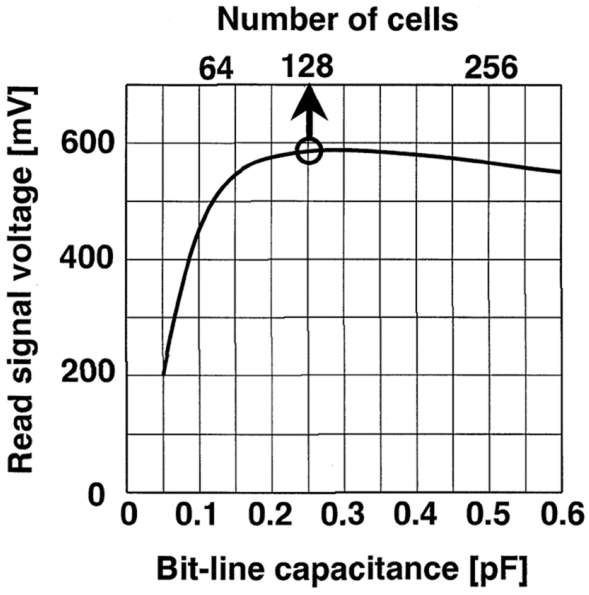


Figure 30: FeRAM read signal voltage versus bit-line capacitance [19]

C.2 Reference Voltage Generation

In contrast to the 2T2C cell, where both data and data complement are stored, a 1T1C cell requires a separately generated reference voltage. The ideal reference voltage for a

single memory cell lies halfway between its bit-line voltage for a "0" (V_{BL0}) and its bit-line voltage for a "1" (V_{BL1}). Typically, the memory cells of a chip exhibit slightly different electrical properties due to manufacturing process variations and, as a consequence, a distribution of cell charges is observed. Figure 31 illustrates a typical cell charge distribution for a FeRAM. It includes a distribution for the binary value "0" that corresponds to a ferroelectric capacitor having a polarization that is not flipped during read, and a distribution for the binary value "1" that corresponds to a ferroelectric capacitor having a polarization that is flipped during read. In Figure 31, the two distributions are cleanly separated, which allows selection of a reference voltage V_{REF} for read operation. Since the chip reference voltage is usually common to all memory cells, it should be halfway between the maximum bit-line voltage for a "0" and the minimum bit-line voltage for a "1".

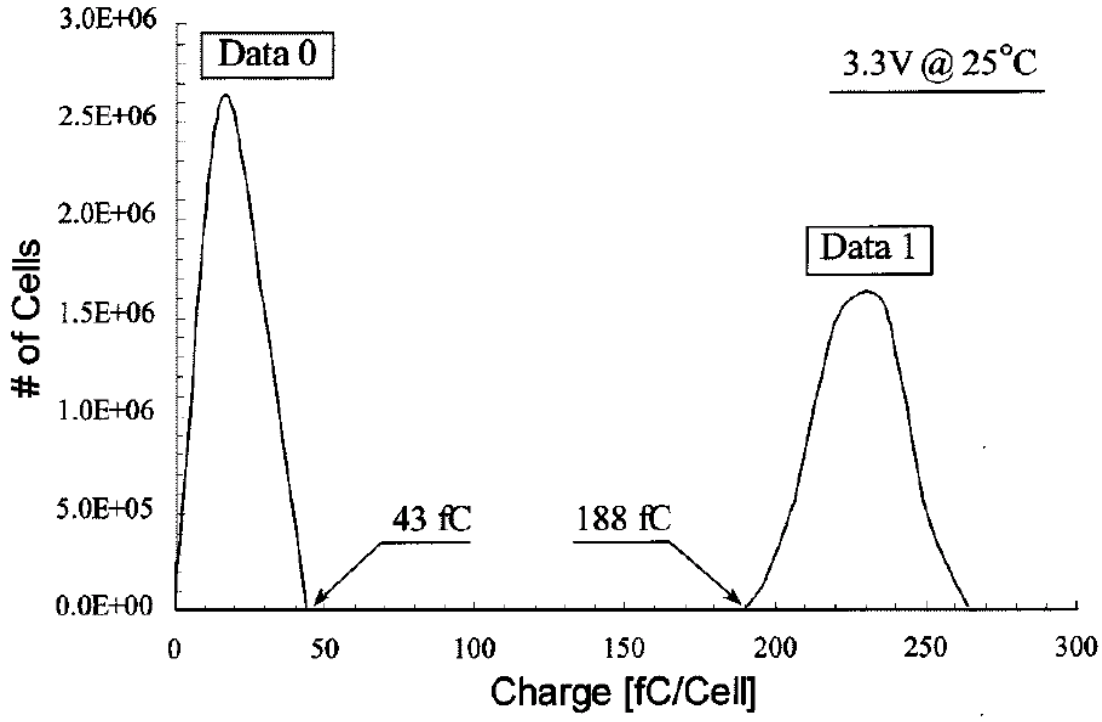


Figure 31: Example FeRAM memory cell charge distribution [20]

Moreover, V_{BL0} and V_{BL1} are also subject to material fatigue, temperature, voltage, and even time. Thus, the reference voltage should try to track these variations to ensure a sufficiently large signal differential at all times. Determining the best way to generate the reference voltage can therefore be a difficult task and numerous techniques have

been proposed to accomplish this [21]. In the following, only two of the most popular techniques will be reviewed.

Oversized ferroelectric capacitor as reference

This older technique makes use of a ferroelectric capacitor whose top electrode is connected via one transistor to the bit-line and via another one to a pre-charge signal. In a first step, the capacitor is always pre-charged to the same polarity and, in the second step, its non-switching charge is dumped onto the bit-line. Because this capacitor is larger than the memory cell capacitor, the voltage that is developed on the bit-line is always larger than V_{BL0} and can be adjusted to be halfway between V_{BL0} and V_{BL1}. This circuitry also tracks temperature variations. However, determining the proper dimensions for the reference capacitor can be difficult.

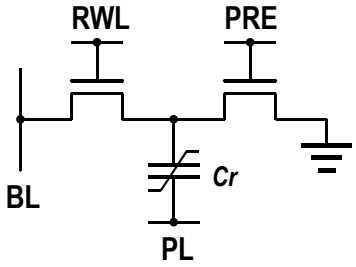


Figure 32: Oversized ferroelectric capacitor

Dielectric capacitor reference

Another reference voltage generation circuit is shown in Figure 33. Instead of a ferroelectric capacitor, it employs a dielectric capacitor to generate the reference voltage, but it is operated in a similar manner. The dielectric capacitor generates a tighter reference voltage distribution, because it is not subject to imprint and other ferroelectric material variations. The generated reference voltage can be adjusted via the reference capacitor pre-charge voltage V_{REFP}:

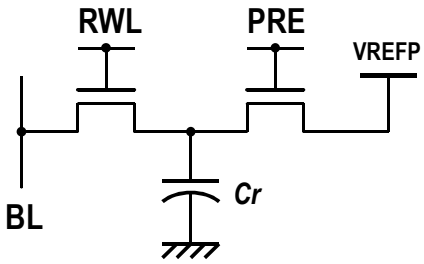


Figure 33: Dielectric reference voltage capacitor

$$V_{REF} = \frac{C_r \cdot V_{REFP}}{C_r + C_{BL}}$$

C.3 Bit-line Capacitance Imbalance During Sensing

In the 1T1C configuration, the charge of the addressed memory cell is read out to bit-line while the reference voltage is generated on bit-line bar. The difference in voltage across the bit lines is then amplified by the sense-amplifier to detect the stored information. However, the capacitive loading of the two bit lines is different (see Figure 34). This constitutes an imbalance to the sense amplifier that is reflected in a loss of signal margin as described next.

When the even word-line (WLE) is selected, the charge of C_S is read out to BL through transistor M0. Simultaneously, the charge of the reference capacitor (not shown) is dumped onto BLB to establish the reference level. At this time, prior to sensing, the capacitive load of BL is C_{BL} + C_S, but the load of BLB is only C_{BL}. This load imbalance, which makes it easier for the sense amplifier to move BLB than BL, degrades the sensing ability of the amplifier. A possible solution to overcome this problem is to disable the word-line before starting the sense operation, and then enabling it after sensing for write-back.

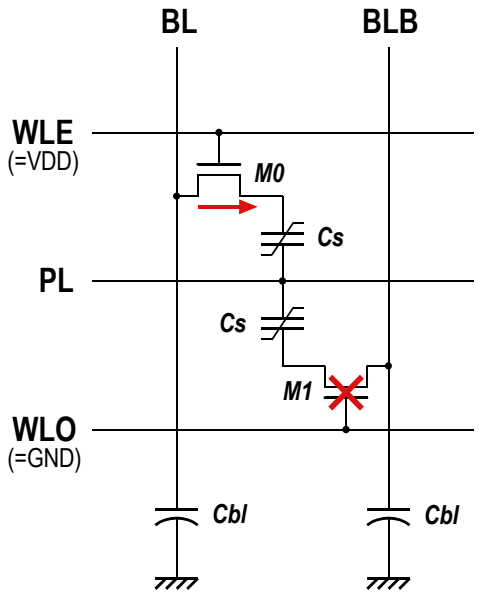


Figure 34: Schematic to illustrate imbalance problem

However, this also results in an access time penalty due to the additional word-line pulsing.

C.4 Plate-line Architecture

In present FeRAMs, the plate-line rise and fall time is known to be very large due to the large switching currents of ferroelectric capacitors. A measured sensing waveform, that has been recently published [22], is cited as an example (Figure 35). According to the graph, the local plate-line SPL has a rise time of approximately 25 ns. As an immediate consequence, the minimum cell access time is limited by the slow plate-line

signal. Choosing the optimum plate-line architecture is therefore crucial for the development of high-performance FeRAM.

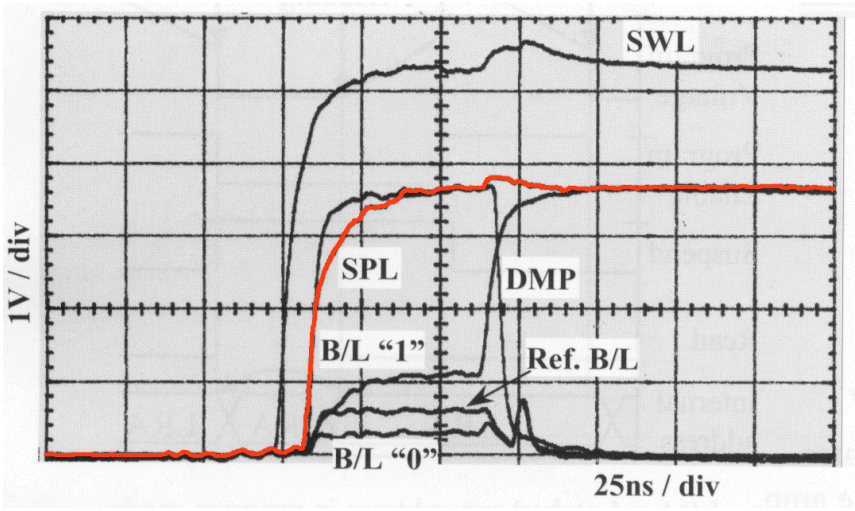


Figure 35: Measured FeRAM plate pulse up sensing waveform [22]

As one possibility, a global plate-line driver could be connected to all plate lines of the array (see Figure 36).

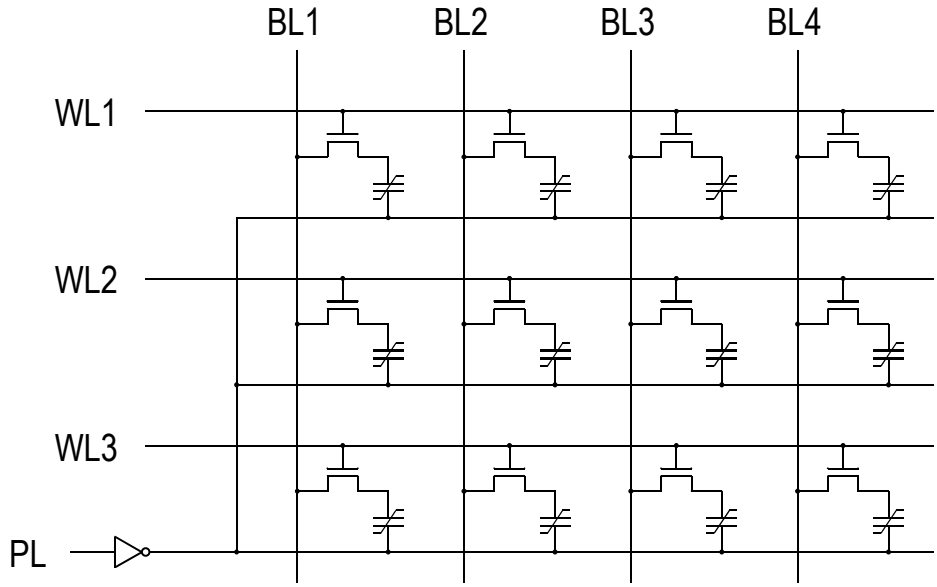


Figure 36: Memory array with global plate-line driver

The advantage of this architecture is that the single driver – despite being usually very large – does not require a large amount of chip real estate relative to the cell array, so the area efficiency can be relatively high. A disadvantage is that at any time

there is one active row and a large number of inactive rows that nevertheless contribute to the total capacitive load. The resulting high capacitive load slows the rise and fall times of the plate-line signal down. Thus, the speed of the FeRAM must be slowed accordingly.

Time constant of plate-line signal

The characteristic time constant of the plate-line signal is approximated by multiplying the on-resistance of the plate-line driver R_{PLD}, by the capacitance of the plate-line C_{PL}. The resistance of the plate-line itself is usually negligible for a metal plate-line (Al or Cu).

$$\tau_{PL} = R_{PLD} \cdot C_{PL} \qquad \text{with} \qquad R_{PLD} \approx \frac{V_{DD}}{I_{on} \cdot W}$$

W is the width of the corresponding transistor and I_{on} is the nominal on current per channel width. The plate-line capacitance is found to be:

$$C_{PL} = C_{wire} + columns \cdot \left(C_S + (rows - 1) \cdot \frac{C_S \cdot C_{jn}}{C_S + C_{jn}} \right)$$

C_S is the ferroelectric capacitor and C_{jn} the junction capacitance of the word-line transistor. C_S is usually ~100 times larger than the junction capacitance C_{jn}. For the worst case it is assumed that all ferroelectric capacitors in the active row switch during plate-line pulsing.

Segmented plate-line architecture

A commonly used plate-line architecture is shown in Figure 37. In the *segmented plate-line architecture*, there is a single global plate-line driver, but pass gates are used to isolate local plate lines LPLx from the global plate-line GPL except for the local plate-line of the row that is accessed.

The advantage of this architecture relative to the global plate-line architecture is that the use of the pass gates reduces the capacitive load on the global plate-line driver, so that the operational speed of the memory array may be increased. However, the pass-gates have to be relatively wide in order to achieve good performance and

larger pass gates also add more capacitive load to the global plate-line. Moreover, additional logic is required to drive the local plate lines of unselected rows to V_{SS} – otherwise they would be floating.

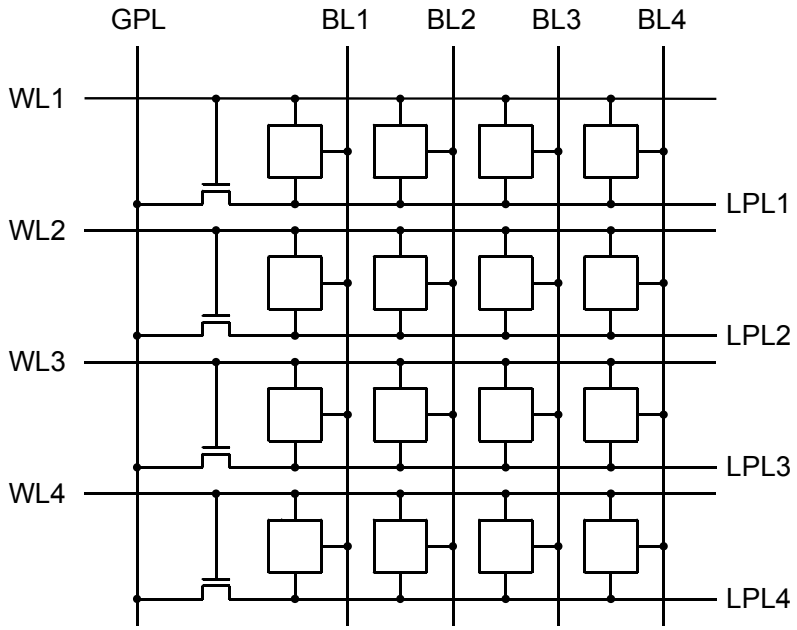


*Figure 37: Segmented plate-*line *architecture*

Non-driven plate-line architecture (NDP)

Another possibility to improve plate-line speed is to employ the *non-driven plate-line architecture* [17]. In this approach the plate-line is hard-wired to½V_{DD} and no longer pulsed. This is similar to DRAM, where the second capacitor plate is also connected to a constant supply voltage. The required bipolar voltage drop across the ferroelectric capacitor is achieved by forcing the bit-line to V_{SS} or V_{DD}. The presenters demonstrated a substantial improvement for access time using the NDP architecture (see Figure 38). Nevertheless, several disadvantages make this scheme less attractive for use in future FeRAMs. First, the voltage range that can be applied to polarize the ferroelectric capacitor is reduced from [-V_{DD}; +V_{DD}] to only [-½V_{DD}; +½V_{DD}].

Therefore the coercive voltage of the ferroelectric needs to be about two times(!) smaller than with a driven plate-line architecture. Another disadvantage is the required refresh for the storage node (SN) of the memory cell. The storage node voltage of all unselected capacitors has to remain at $\frac{1}{2}V_{DD}$ at all times; otherwise, the voltage drop across the capacitor would destroy the stored information. Since the bit lines are pre-charged to V_{SS}, leakage current from the storage node through the cell transistor to the bit-line causes the voltage on the storage node to decrease with time. A refresh operation is necessary to regularly restore the storage node voltage to a full $\frac{1}{2}V_{DD}$ level. This causes higher power dissipation.

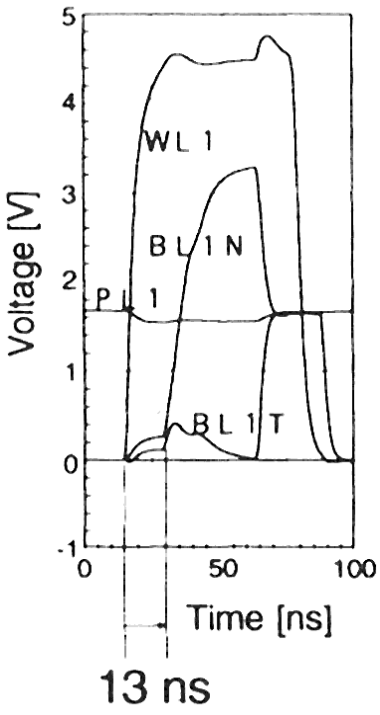


Figure 38: Simulation of non-driven plate line [17]

IV. FERROELECTRIC CAPACITOR MODEL

In cooperation with IMEC[1] a model for ferroelectric capacitors has been developed and implemented into a standard circuit simulator. The model is based on the Preisach theory of hysteresis for magnetic systems [23]. This chapter presents the integration issues and application of the model. For a complete description of the model theory and algorithm, please refer to the Ph. D. thesis of Andrei Bartic [24] and related publications [25-28].

A. MODEL IMPLEMENTATION

The ferroelectric capacitor model has been integrated into the Cadence Spectre circuit simulator using AHDL – an analog hardware description language. The model consists of three components in parallel: a dielectric capacitor, a leakage resistor, and a pure ferroelectric capacitor. The dielectric capacitor represents the dielectric properties of the real ferroelectric capacitor, while the pure ferroelectric capacitor represents only the ferroelectric properties. The saturation curve of the pure ferroelectric capacitor is separated into an ascending and a descending branch. Each branch is normalized to 1 and approximated by the function F:

$$F(V) = \frac{2}{\pi} \cdot \arctan\left[a \cdot (V - V_{CO})\right]$$

$$\text{with: } F(+\infty) = 1 \ , \quad F(-\infty) = -1$$

Vco is the coercive voltage and is Vcp for the ascending branch and is Vcn for the descending branch. Scaling factor a is used to fit the slope of the function F to the experimentally obtained slope. Sub-loops are automatically generated by the model as scaled versions of the saturation loop.

The AHDL source code of the model (see Appendix A) consists of three main sections: interface description, initialization routine, and model body. Typically, each instance of a ferroelectric capacitor in a circuit schematic can have different properties

[1] IMEC, Kapeldreef 75, B-3001 Leuven, Belgium. (www.imec.be)

such as capacitor size, coercive voltage, or remanent polarization. The interface description specifies model parameters that are passed by the circuit simulator to the model. The following 8 parameters are required for this model: capacitor area, capacitor thickness d, relative dielectric constant Er, saturation polarization Ps, remanent polarization Pr, positive coercive voltage Vcp, negative coercive voltage Vcn, and leakage resistance Rleak. Parameters d, Er and area are used to calculate the capacitance C_0 of the dielectric capacitor, while parameters R_{leak} and area are used to calculate the leakage resistance R_L:

$$C_0 = \frac{\varepsilon_0 \cdot Er}{d} \cdot area \qquad R_L = R_{leak} \cdot area$$

The initialization routine is located in the *initial{}* section and is invoked only once prior to the transient simulation and prepares the model for simulation. It calculates internal model variables from the external parameters and initializes the arrays that store the voltage history.

The body of the model is located in the *analog{}* section and is a function that uses the current voltage across the capacitor to calculate the resulting current through the capacitor. The model accounts for the history dependency of the ferroelectric capacitor by keeping track of the voltages that have been applied to the capacitor. It accomplishes this by maintaining a table of *turning points [27]*. A turning point is located where dV/dt changes sign. If, for example, a series of voltages –1.0 V, 0 V, 1.0 V, and 0.5 V is applied to the ferroelectric capacitor, 1.0 V is considered a turning point. To identify turning points, the algorithm always remembers if the voltage was increasing or decreasing previously. If the algorithm detects a change in direction, it considers the previous voltage as a turning point and stores its voltage and charge in the turning point array. Two 1-dimensional arrays, Vtp and Qtp, are used to store the voltage and charge of a turning point, while an index variable ix is used to remember the last position within the arrays. All entries with an even index (0, 2, 4, 6, …) are positive turning points, while all entries with an odd index (1, 3, 5, 7, …) are negative turning points. Arrays Vtp and Qtp have been dimensioned so that a maximum number of 50 turning points can be recorded per ferroelectric capacitor instance. This is more than

sufficient as for most transient simulation no more than about 10 turning points have to be stored at a time.

If the direction does not change, but the magnitude of the present voltage exceeds the voltage of existing turning points, the algorithm removes these turning points from the array. Once the algorithm has determined on which sub-loop the current voltage must be located, it uses the following formula to calculate the corresponding charge:

$$P\left(V_{fe}\right) = Q_{tp}\left[ix-1\right] - m_i \cdot \left(F\left(V_{tp}\left[ix-1\right]\right) - F\left(V_{fe}\right)\right)$$

$$\text{with:} \qquad m_i = \frac{Q_{tp}[ix-1] - Q_{tp}[ix-2]}{F\left(V_{tp}\left[ix-1\right]\right) - F\left(V_{tp}\left[ix-2\right]\right)}$$

The current through the capacitor is the derivative of the charge P and is automatically computed by the circuit simulator using an internal function called *dot()*:

$$I = dot(P)$$

In the following, an example walk-through is presented for the algorithm described above. It is assumed that the last applied voltage, which is stored in Vtp[ix], is 0 volts and the current voltage Vfe is 1.0 volt. Moreover, the contents of the turning point arrays Vtp and Qtp are as shown in Table 4, and the value of index variable ix is 2.

Table 4: Arrays Vtp and Qtp store voltage and charge of turning points

ix	V_{tp}	Q_{tp}
0	$+\infty$	P_S
2	0	P_r
4		

ix	V_{tp}	Q_{tp}
1	$-\infty$	$-P_S$
3		
5		

As the current voltage of 1.0 volt is larger than the previous voltage of 0 volts, the voltage increased since the last invocation of the algorithm. In addition, the even value of ix indicates that voltage was also increasing the time before. Therefore, the last voltage is not a turning point and no new sub-loop has been entered. Since the present

voltage is also smaller than the voltage of the last positive turning point Vtp[0], no turning point has to be removed from the table. Accordingly, mi and P can be calculated as follows:

$$m_i = \frac{Q_{tp}[1] - Q_{tp}[0]}{F\left(V_{tp}[1]\right) - F\left(V_{tp}[0]\right)} = \frac{-P_S - P_S}{F(-\infty) - F(+\infty)} = \frac{-P_S - P_S}{-1-1} = P_S$$

$$\begin{aligned} P(1.0\,V) &= Q_{tp}[1] - m_i \cdot \left(F\left(V_{tp}[1]\right) - F(1.0\,V)\right) \\ &= -P_S - P_S \cdot \left(F(-\infty) - F(1.0V)\right) \\ &= -P_S + P_S \cdot \left(1 + F(1.0V)\right) \\ &= P_S \cdot F(1.0V) \\ &= P_S \cdot \frac{2}{\pi} \cdot \arctan\left[a \cdot \left(1.0V - V_{cp}\right)\right] \end{aligned}$$

The calculated P belongs to the ascending branch of the saturation curve.

Summary

The presented implementation of a ferroelectric capacitor model is very compact, fast, and memory-efficient. It is very flexible, since it allows the choice of a different function F to achieve the best fit possible to experimental hysteresis loops. It accounts for the history dependency of the ferroelectric capacitor by memorizing turning points and produces scaled sub-loops that are derived from the chosen function F. In addition, it models leakage current and dielectric properties independent from the ferroelectric properties. The source code is expandable and new features can be added easily. Moreover, since the AHDL syntax is very similar to the syntax of the C language, the presented implementation can be ported over to other circuit simulators with only little effort.

B. Analysis of Memory Failure Due to Imprint

In this section, the ferroelectric capacitor model described in the previous section is used to approach the imprint phenomenon from a device point of view. Imprint in ferroelectric thin films is defined as the preference of one polarization state over the

other and is known to be a serious failure mechanism in ferroelectric memory cell capacitors. The impact of imprint on memory operation is investigated using circuit simulation of typical read/write operations on virgin and imprinted ferroelectric memory cell capacitors. This enables a more accurate lifetime prediction and better failure analysis than the common lifetime estimates, which are usually based on manual calculations of very simplified read/write operations.

B.1 Experimental

Imprint measurements have been performed on $SrBi_2Ta_2O_9$ (SBT) thin films, which were prepared by wet chemical solution deposition on silicon wafers. Top and bottom contacts are 100 nm thick platinum films that were deposited by sputtering. The SBT film thickness is 230 nm and a contact pad size of 0.0177 mm^2 is used. The SBT film is annealed for 60 min. at 780°C. After sputtering of the platinum top contact, a post-annealing process step of 30 min. at 800°C is very important for the ferroelectric characteristics of the samples. Electrical characterization, especially imprint measurements, were performed by the aixACCT TF Analyzer 2000 FE Module at test conditions of 100 Hz, 3.3 V, and 125°C.

B.2 Memory Operation and Imprint

After information has been written to a memory cell capacitor, it is stored until it is needed or overwritten. The duration of the storage interval can vary from only a few nanoseconds to days or even months. If, for example, a “1” is stored (state A, see Figure 39), the continuous positive polarization will lead to static imprint that causes the hysteresis loop to be shifted left. In a similar manner, a stored “0” (state B-1) will lead to static imprint that causes the hysteresis loop to be shifted right. Moreover, if a “0” is repeatedly rewritten (state B-2), the unipolar pulsing that is applied to the capacitor will lead to dynamic imprint, which again causes a right shift. In contrast, continuously rewriting “1”s causes no imprint because of the relatively balanced bipolar pulsing that is applied to both terminals of the capacitor.

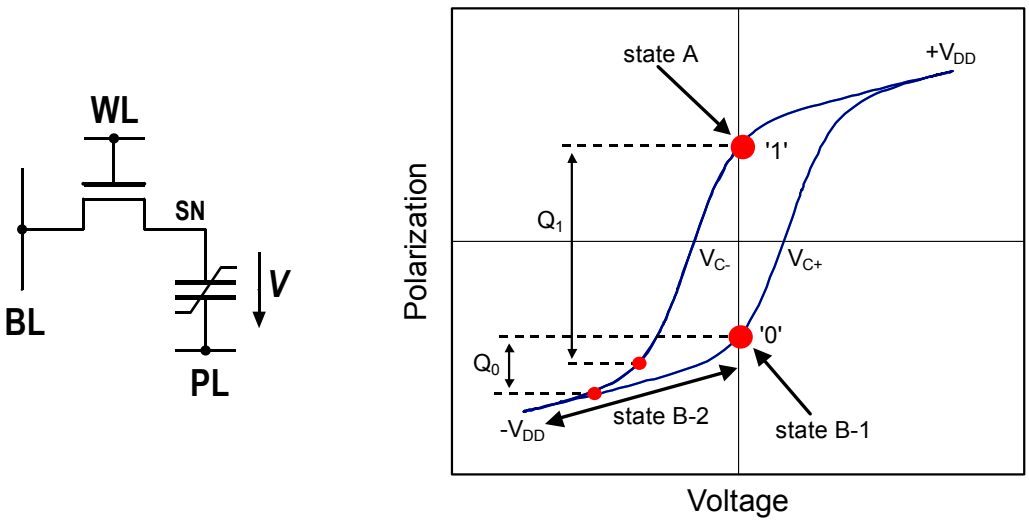


Figure 39: States in a ferroelectric capacitor that cause imprint

B.3 Failure Due to Imprint

Failure criteria

Imprint causes a variation in relevant ferroelectric capacitor properties (V_C, P_r). These variations affect the amount of charge (Q_0, Q_1) that will be dumped onto the bit lines during readout and, consequently, affect the bit-line voltage levels. In a 1T1C configuration, the voltage on bit-line is compared against a reference voltage V_{REF} that has been established on bit-line bar or vice versa. Usually, a minimum signal difference of V_S – typically about 70 mV – is required for stable operation [29]. This is the case if V_{BL0}, the bit-line voltage after reading a stored "0", is smaller than V_{REF}–V_S and V_{BL1}, the bit-line voltage after reading a stored "1", is larger than V_{REF}+V_S. However, if V_{BL0} would become larger than V_{REF}–V_S due to imprint, the "0" could be sensed as a "1". The same applies to a V_{BL1} that becomes smaller than V_{REF}+V_S – the "1" could be sensed as a "0". This results in the failure criterion:

$$V_{BL0} > V_{REF} - V_S \quad \wedge \quad V_{BL1} < V_{REF} + V_S$$

Possible sequences of operation

As presented earlier (Figure 39), two states A and B can cause imprint in a memory cell capacitor. Based on the two states, there are in principle four different sequences of operation that an imprinted memory cell might experience (see Table 5). We will find later, that out of these four possible sequences, only three are critical and can lead to bit errors according to the failure criterion that was derived in the previous paragraph.

Table 5: Sequences of memory operations

	Present state: A	Present state: B
Next operation: Read	READ-A	READ-B
Next operation: Write	OVERWRITE-A	OVERWRITE-B

READ-A

In the first sequence, called READ-A, state A was maintained for a certain time and is left by a read operation (see Figure 40).

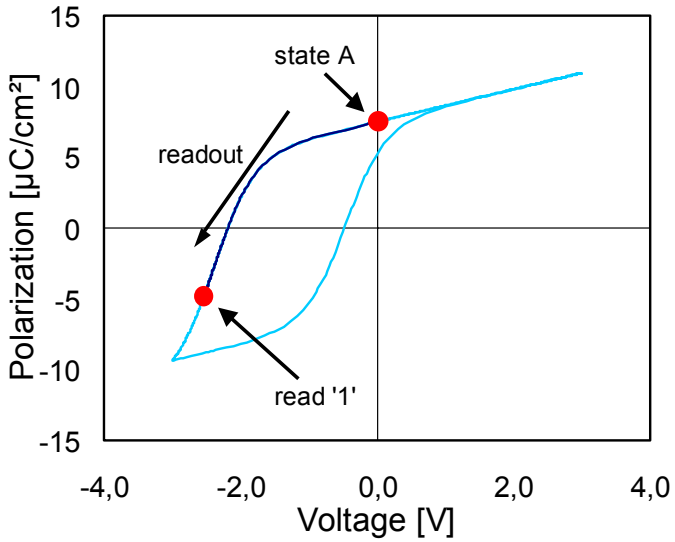


Figure 40: Simulation of imprint sequence "READ-A"

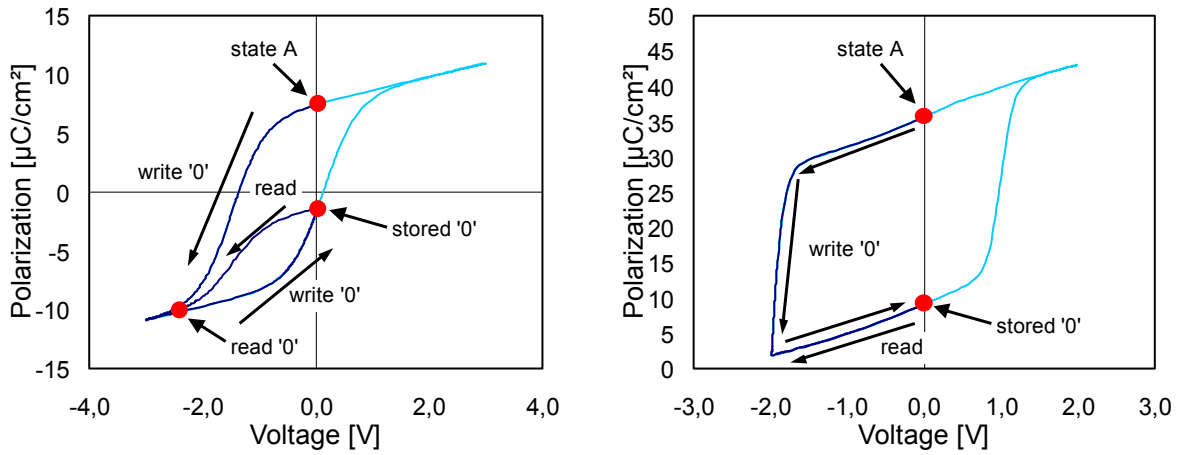
Due to imprint, the hysteresis loop has shifted to the left and, accordingly, the coercive voltages V_{C-} and V_{C+} have shifted, too. If the read voltage (=V_{DD}) is not large enough,

i.e. $V_{DD} < -V_{C-}$, to generate sufficient charge Q_1 during readout, V_{BL1} might be significantly reduced and might fall below the critical threshold of $V_{REF} + V_S$. Accordingly, this sequence can lead to bit errors.

OVERWRITE-A

In OVERWRITE-A, again state A was maintained for a certain time, but then it is overwritten by a "0" and finally read out. In the case where the write voltage (=V_{DD}) is significantly larger (~2x) than the shifted coercive voltage, the polarization can be completely reversed by applying $-V_{DD}$ to the ferroelectric capacitor. However, because the coercive voltage V_{C+} is also decreased due to imprint, the resulting Q_0 and V_{BL0} can be increased (Figure 41a). Consequently, this sequence can lead to bit errors, too.

On the other hand, if the shifted V_C is almost as large as V_{DD}, the polarization cannot be completely reversed by applying $-V_{DD}$ to the ferroelectric capacitor (Figure 41b). However, the "weak 0" will not necessarily generate an increased Q_0 because the subsequent read voltage (=V_{DD}) is not large enough to generate switching charge during readout. Therefore, V_{BL0} might not increase at all.



(a) for $V_C << V_{DD}$ *(b) for $V_C \sim V_{DD}$*

Figure 41: Simulation of imprint sequence "OVERWRITE-A"

READ-B

No bit errors will be generated by READ-B because the increased coercive voltages (right shift) can only cause a decrease of Q_0 and V_{BL0}, which would increase the bit-line differential.

OVERWRITE-B

In OVERWRITE-B, state B was maintained for a certain time and is left by writing a "1" followed by readout (Figure 42). Due to imprint, the coercive voltage V_{C+} will be increased and the switching charge Q_1 will be reduced. Consequently, V_{BL1} is decreased and this sequence may also lead to bit errors.

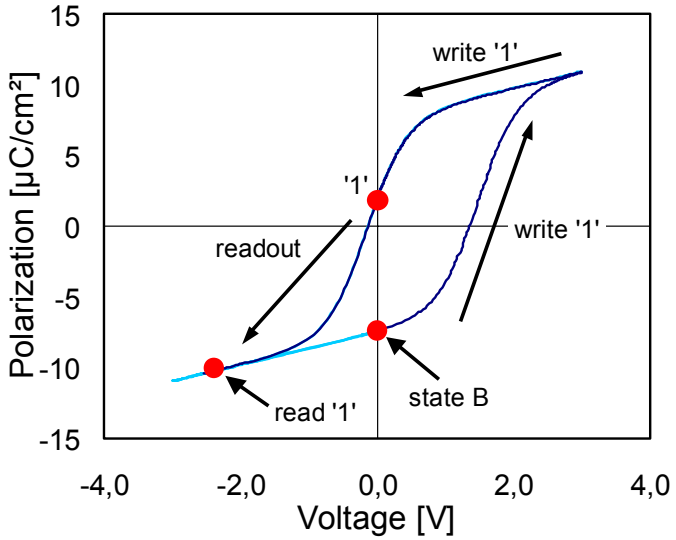


Figure 42: Simulation of imprint sequence "OVERWRITE-B"

B.4 Simulation

Simulation procedure

Prior to simulation, the ferroelectric capacitor model has to be fitted to the measured data of the virgin capacitor and various stages of the imprinted capacitor. This implies that a separate parameter set has to be created for each ΔV_C, which is defined as the relative shift ΔV_C of the average of both coercive voltages:

$$\Delta V_C = \frac{V_{C+} + V_{C-}}{2}$$

This can be done using a spreadsheet and a conventional fitting algorithm. Alternatively, the dependence of each model parameter (e.g. V_{C+}, V_{C-}, P_{r+}, P_{r-}, etc.) on ΔV_C could be approximated analytically and the corresponding equations directly included

in the model. Figure 43 shows how well the earlier presented ferroelectric capacitor model fits the experimental data.

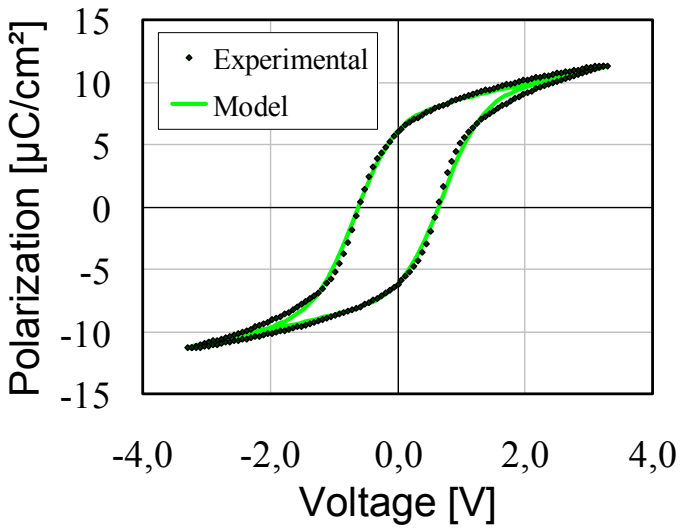


Figure 43: Measured and simulated hysteresis loop (SBT)

After fitting, a transient simulation with ΔV_C as the sweep parameter is used to evaluate the impact of ΔV_C on bit-line voltage levels. At the beginning of the transient simulation, a "1" is written into the capacitor using a regular write operation. Then the "1" is overwritten by a "0". To obtain V_{BL0}, the capacitor is read out and the resulting bit-line voltage is the V_{BL0} for the given ΔV_C. To obtain V_{BL1}, the "0" has to be again overwritten by a "1" and the bit-line voltage after the following read is V_{BL1}.

Simulation results

Figure 44 is a plot of the bit-line voltage after readout for "0" (V_{BL0}) and "1" (V_{BL1}) data, where the x-axis is ΔV_C. As discussed earlier, for proper operation V_{BL1} must be larger than the upper end of the reference voltage band and V_{BL0} must be smaller than the lower end. This is the case if the shift is smaller than about 1.05 V (value from Figure 44), otherwise, sequence READ-A can cause failure. There is no failure caused by OVERWRITE-A as the impact of the shift on V_{BL0} is very small thanks to the high coercive voltage relative to V_{DD}.

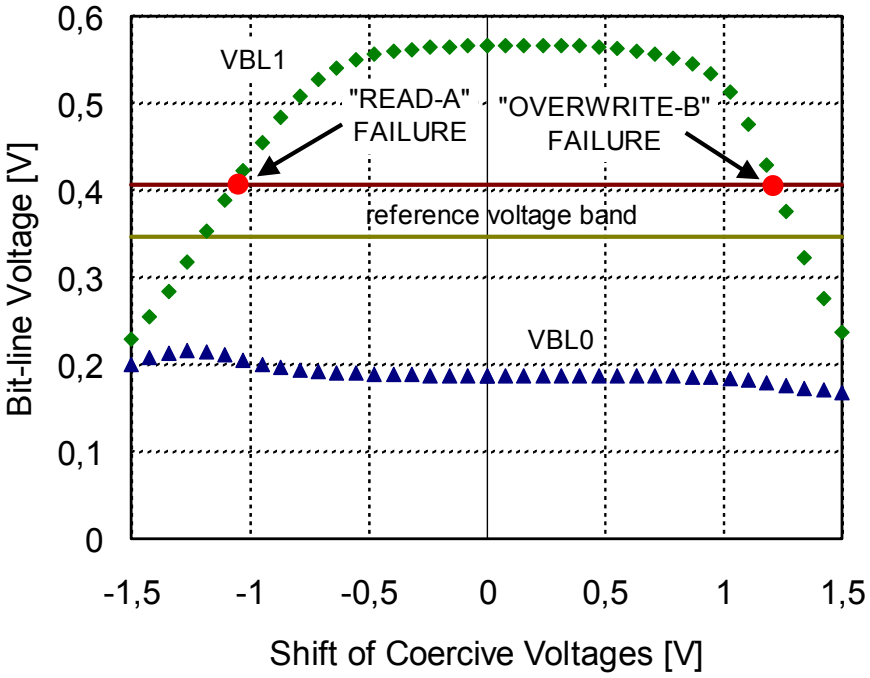


Figure 44: Bit-line voltage versus ΔV_C (for $V_C \sim V_{DD}$)

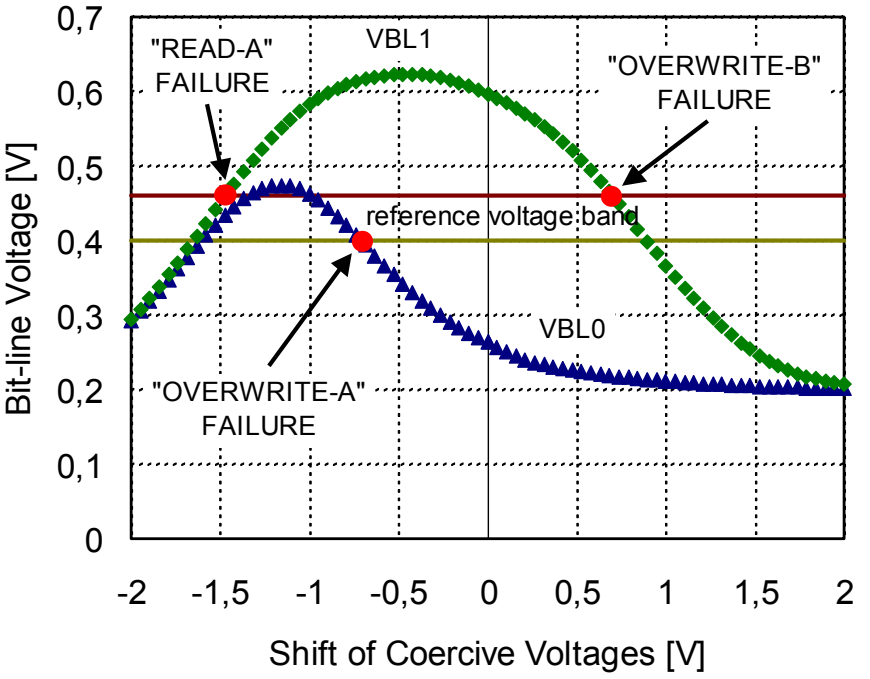


Figure 45: Bit-line voltage versus ΔV_C (for $V_C << V_{DD}$)

Note that a shift of up to ±0.5 V has only little impact on bit-line voltage levels, which shows that the memory cell is very robust against small variations of the coercive

voltage. In Figure 45, the same graph is presented for $V_C << V_{DD}$. Even a small coercive voltage shift can result in a negative V_{C+} or a positive V_{C-}. Consequently, the OVERWRITE-A failure is very likely to occur with a negative ΔV_C (= left shift).

Lifetime prediction

Once the dependence of bit-line voltage levels on ΔV_C has been obtained, lifetime limitations due to imprint can be identified in two steps. This will be illustrated for Figure 45. First, the maximum allowed shift for all failure cases has to be determined (Figure 46). Then the corresponding time for each shift is obtained by linear interpolation of the measured ΔV_C (see Figure 47). The earliest failure case gives the lifetime limit.

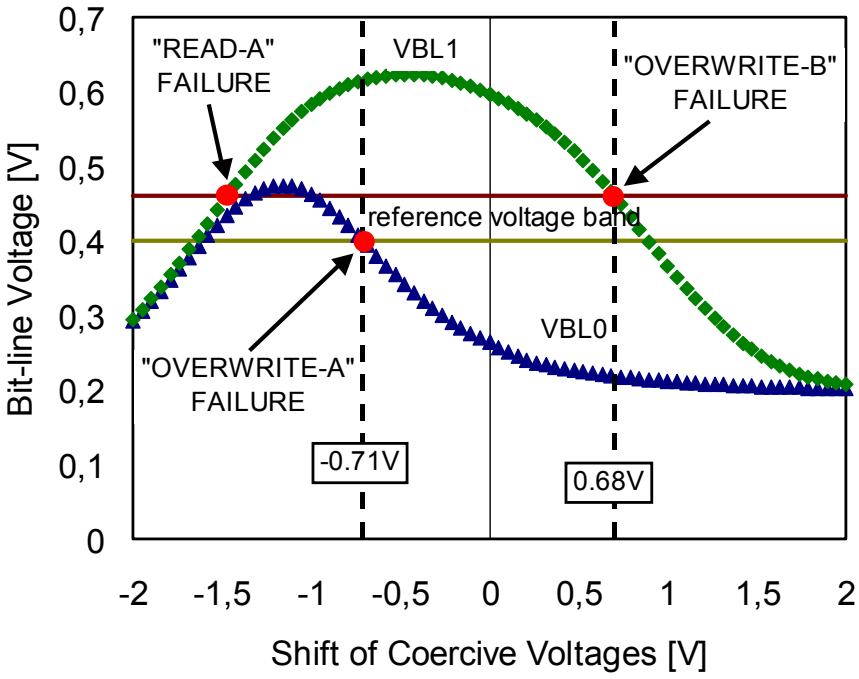


Figure 46: Maximum allowed negative and positive voltage shifts due to failure cases OVERWRITE-A and OVERWRITE-B

B.5 Conclusion

Lifetime limitations due to imprint can be easily identified using the ferroelectric capacitor model and the analysis described earlier. The accuracy of the simulation results strongly depends on the underlying experimental data and the fitting algorithm for the ferroelectric capacitor model. Based on the different results for the cases $V_C <<$

V_{DD} and $V_C \sim V_{DD}$ and the general fact that SBT exhibits smaller coercive voltages, it is evident that SBT is more susceptible to a small coercive voltage shift than PZT and especially more susceptible for failure case OVERWRITE-A. Furthermore, failure cases OVERWRITE-A and OVERWRITE-B could probably be overcome by circuit techniques in the near future because of their nature – they are write failures and it is not necessary to retrieve the information that was stored prior to imprint. If this can indeed be done, then READ-A failures will limit lifetime due to imprint.

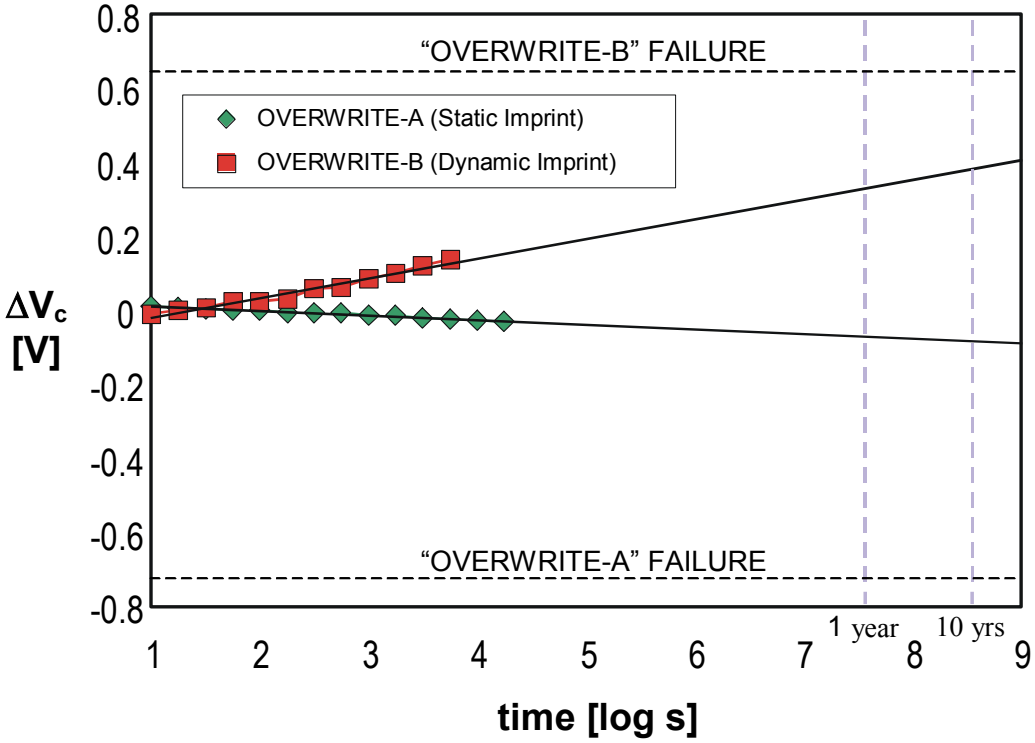


Figure 47: Imprint: Interpolation of measured ΔVC

V. AN INITIAL FERAM TEST CHIP

The first test chip created during the course of this doctorate was developed in cooperation with IMEC. Its main purpose was to serve as a ferroelectric technology demonstrator for a 0.35-µm process with PZT or SBT as the ferroelectric. The nominal operating voltage for this process is 3.3 V.

A. CHIP FEATURES

As presented in Chapter III, there are two fundamentally different architectures for ferroelectric memories today: the standard and the Chain FeRAM (CFRAM) architecture. In order to be able to experiment with the different memory architectures, the chip was partitioned into two memory blocks: one 256-kbit block consisting of conventional FeRAM and one 256-kbit block consisting of Chain FeRAM. To simplify matters, the two memory blocks share a large part of the I/O and control circuitry. The design also strives to maximize testing capabilities by allowing most internal timing signals to be controllable by external signals. In this way, almost any read and write sequence can be implemented during chip characterization. In addition, the reference voltage is supplied from the outside and is freely adjustable.

Table 6: Remanent charge for different capacitor sizes and P_r values

Capacitor	SBT			PZT		
size	5 µC/cm²	7 µC/cm²	10 µC/cm²	15 µC/cm²	20 µC/cm²	30 µC/cm²
0.64 µm²	-	-	-	-	128 fC	192 fC
0.81 µm²	-	-	-	121.5 fC	162 fC	243 fC
1 µm²	-	-	100 fC	150 fC	200 fC	300 fC
1.56 µm²	-	109.2 fC	156 fC	234 fC	312 fC	-
2.1 µm²	105 fC	147 fC	210 fC	315 fC	-	-
2.56 µm²	128 fC	179.2 fC	256 fC	-	-	-
3.6 µm²	180 fC	252 fC	-	-	-	-

Furthermore, the FeRAM can be switched between 2T2C or 1T1C operation via external control signals. The FeRAM array features a hierarchical bit-line architecture and seven different ferroelectric capacitor sizes (from 0.64 µm² to 3.6 µm² – see Table 6) to account for the different P_r values of SBT and PZT and manufacturing process variations.

B. Chip Architecture

The top level of the chip consists of the two memory blocks and a row decoder (Figure 48). Since the interface signals of both memory blocks are almost identical, they share most of the control signals and the address decoder. The address decoder receives a 12-bit address that is decoded into a segment and row address. There are a total of 8 segments (4 FeRAM and 4 CFRAM segments). Bits 11-9 are the address of the segment and bits 8-0 are the address of the accessed row. Two separate 4-bit data busses form one 8-bit data bus on the top level. The upper 4 bits (bits 7-4) of this data bus come from the FeRAM block and the lower 4 bits (bits 3-0) come from the CFRAM block.

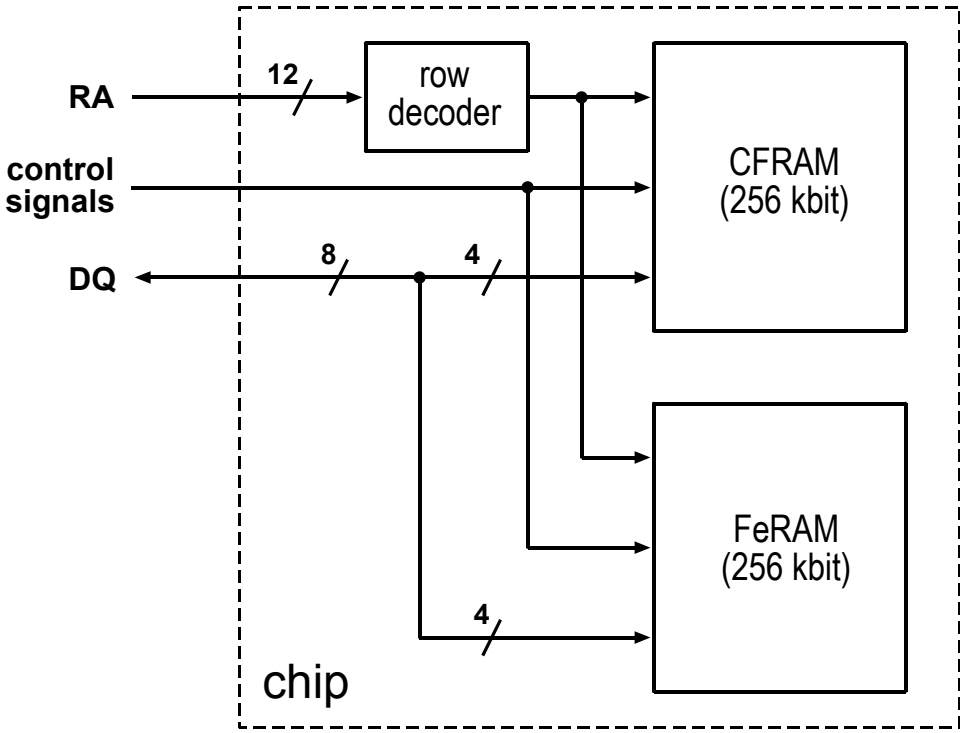


Figure 48: 0.35 µm chip block diagram

B.1 FeRAM Block

Bit cell

The schematic of the FeRAM bit cell is shown in Figure 49. It is very similar to the 2T2C bit cell described on page 29 except that is has two separate word lines (WLE, WLO) instead of only one common word-line. It is often used as a 1T1C cell for a folded bit-line architecture. This bit cell is suitable for 2T2C and 1T1C operation. However, since the plate-line (PL) is shared by both capacitors, the unselected capacitor will experience disturb pulses in 1T1C operation.

To obtain the same bit- and word-line pitch, the layout of the bit cell that contains the largest (A=3.6 μm²) of the seven different capacitors was created first. The layouts of the other six bit cells were then derived from the layout of this cell. Only the size of the top capacitor electrode was adjusted in order to obtain the different capacitor sizes. By maintaining the same bit- and word-line pitch, the same peripheral circuits (drivers, decoders, sense amplifiers, etc.) can be used for arrays of the different bit cells. Of course, this also allows mixing of different bit cells within the same array.

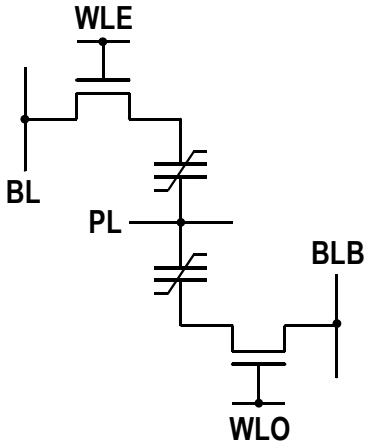


Figure 49: FeRAM block: Bit cell

The parasitic capacitances of the bit cell were calculated manually. These calculations enabled a reasonable estimate of the bit-line capacitance. The number of cells per bit-line was optimized for a remanent charge of ~160 fC using the technique presented on page 32ff. (Chapter III). This allows plenty of margin to account for a large range of P_r values (see Table 6 above).

Segment

A segment contains 32k bit cells organized as 512 rows by 64 columns. It employs a hierarchical bit-line architecture to reduce bit-line capacitance. Eight bit cells are connected to one sub-bit-line and each of the 64 sub-bit lines is connected to a bit-line via an n-channel transistor (see Figure 50). The structure of the segment is shown in Figure 51. Each of the 512 rows has a separate local plate-line LPL_X that is connected via an n-channel transistor to the global plate-line. This is a typical segmented plate-line architecture (see page 41ff). Note that the 512 LPL pass-gates represent a relatively large load to the global plate-line driver.

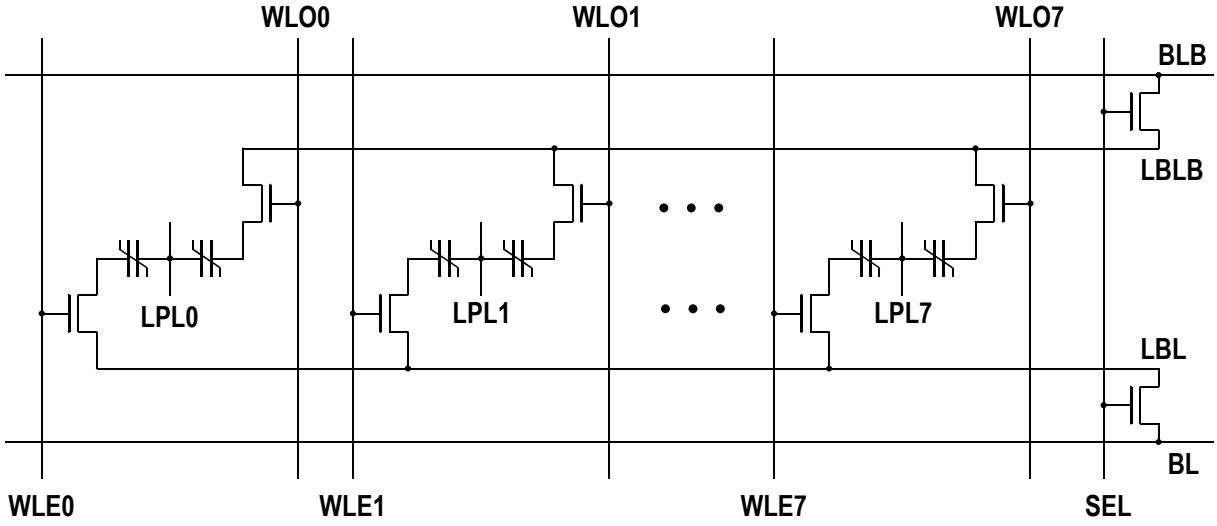


Figure 50: FeRAM block: Hierarchical bit-line

It is necessary to dimension the pass-gate properly to achieve a short plate-line rise/fall time. If the W/L ratio of the pass-gate is chosen too small, only a poor plate-line rise/fall time is achieved. On the other hand, if the pass-gates are chosen too large for the global plate-line driver, they are likely to excessively load down the driver. First worst-case estimations indicated that 64 active cells are a manageable number of cells for a plate-line driver. A follow-up simulation could confirm this assumption by displaying an acceptable trade-off between driver area and plate-line performance. Therefore, each segment comprises 64 bit-line pairs (BL/BL bar).

Since each row consists of two word lines (even and odd), there is a total of 1024 word lines. Each bit-line pair has its own pre-charge circuit and is connected to one

sense amplifier (see Figure 52). In 1T1C operation, the external reference voltage is generated on either the odd or even bit-line.

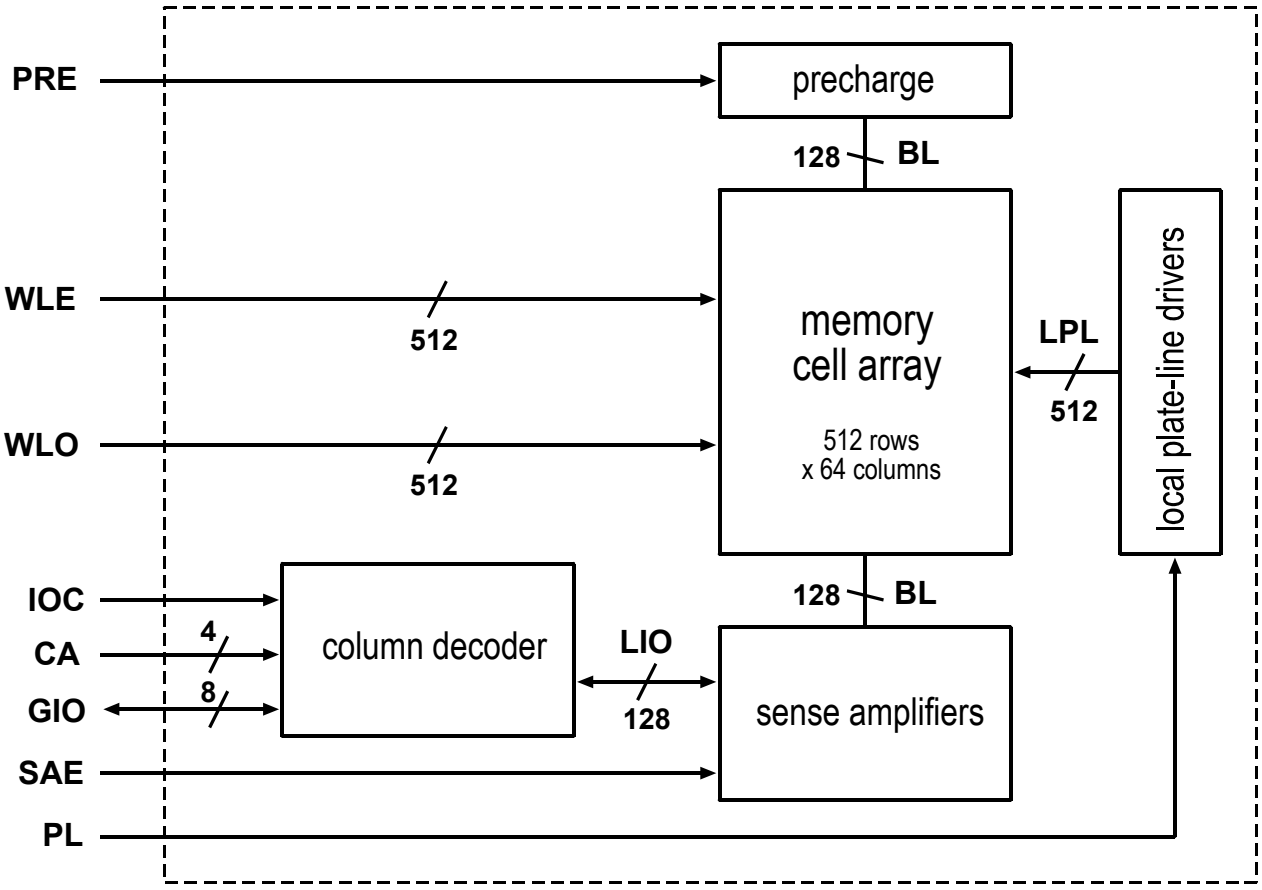


Figure 51: FeRAM block: 64-kbit segment

Hierarchical bit-line

With C_{acc} being the parasitic capacitance per word-line transistor and C_{wire} being the parasitic routing capacitance, the total capacitance for a non-hierarchical bit-line is approximately:

$$C_{BL} \approx C_{wire} + (rows - 1) \cdot C_{acc}$$

For example, 512 cells per bit-line (=512 rows), a routing capacitance of C_{wire}=234 fF, and a capacitance per word-line transistor of C_{acc}=2.13 fF together give a bit-line capacitance of:

$$C_{BL} \approx 234\,fF + 511 \cdot 2.13\,fF = 1.3\,pF$$

In contrast, the hierarchical bit-line with N=8 cells per sub-bit-line only has (rows/N)-1 transistors per bit-line plus (N-1) transistors per sub-bit-line that contribute to the total bit-line capacitance. Accordingly, the capacitance of the hierarchical bit-line is approximately:

$$C_{BL} \approx C_{wire} + \left(N + \frac{rows}{N} - 2 \right) \cdot C_{acc}$$

$$\Rightarrow C_{BL} \approx 245\,fF + 70 \cdot 2.13\,fF = 394\,fF$$

Note that the slight increase of C_{wire} to 245 fF is caused by the additional routing required for the sub-bit-line. In this example, employing a hierarchical bit-line architecture reduced bit-line capacitance substantially. This allows use of a smaller cell capacitor or selection of a larger number of cells per bit-line. On the other hand, since the bit-line and sub-bit-line cannot usually share the same metal layer due to the minimum column pitch, an extra metal layer is required for implementing a hierarchical bit-line. Only this requirement by itself makes a hierarchical bit-line architecture in many cases impractical. For this chip, the hierarchical bit-line was implemented for experimental purposes.

Sense amplifier

In addition to the well-known cross-coupled latch, the sense amplifier features two additional transmission gates that separate the internal sensing nodes IO and IOB from the bit lines. Operation is described next. Prior to sensing, the sense amplifier is disabled with control signal SAN(SAP) being held at $V_{SS}(V_{DD})$. At the same time, the transmission gates are closed with transmission gate control signal OEN(OEP) being held at $V_{SS}(V_{DD})$. After the signal has developed on both bit lines during readout, both transmission gates are opened by taking OEN(OEP) to $V_{DD}(V_{SS})$ in preparation for the sense operation. Now the sense amplifier is activated by taking SAN(SAP) to $V_{DD}(V_{SS})$. Once sense operation is complete, the transmission gates are closed again for the write-back operation.

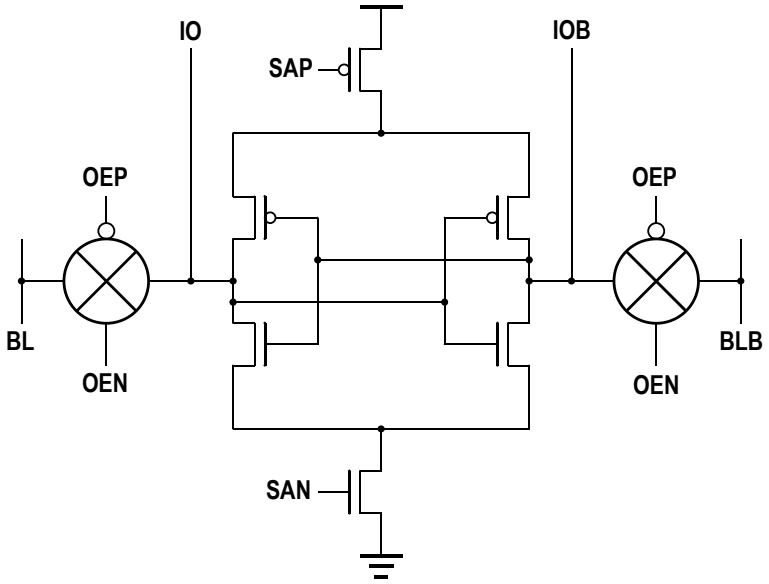


Figure 52: Sense amplifier

The internal sensing nodes IO and IOB are also connected to a multiplexer that connects them to a global I/O bus. The 64 IO/IOB pairs are demuxed to four global I/O line pairs (GIO). A local column decoder that receives a 4-bit address decides which set of four IO/IOB pairs is connected to the global I/O bus.

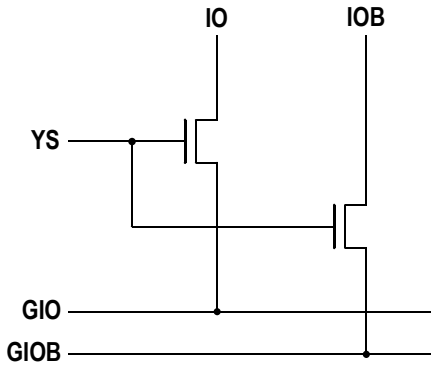


Figure 53: I/O multiplexer

Memory array

The memory array consists of 4 segments and the total memory capacity is 256-kbit. For greater area efficiency, all segments share the same 1024 word lines and the associated drivers and decoders. They also share the control signals. When a row within one segment is accessed, the word-line transistors of the same row in the other 3 segments are turned on as well. This is no problem because every segment has separate

bit- and plate-line drivers. Since both bit- and plate-line are kept at V_{SS} for these cells, the stored information is not destroyed.

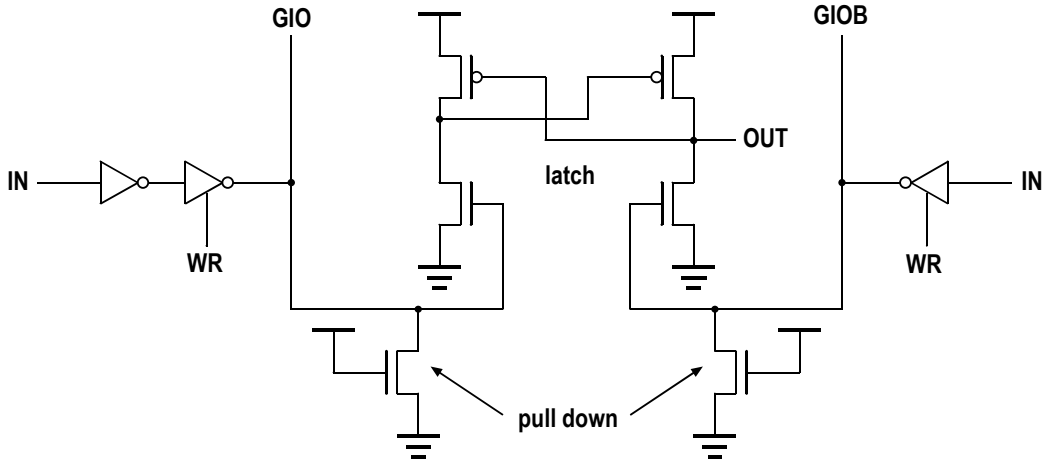


Figure 54: Receiver/driver for global I/O bus

The global I/O bus is connected to all segments. At its end, the global I/O bus is connected to an I/O interface circuit (see Figure 54), which splits the bus into two separate data busses – an input and an output data bus. The I/O circuit comprises a latch-type receiver and tri-state drivers. The tri-state drivers are enabled by the WR signal.

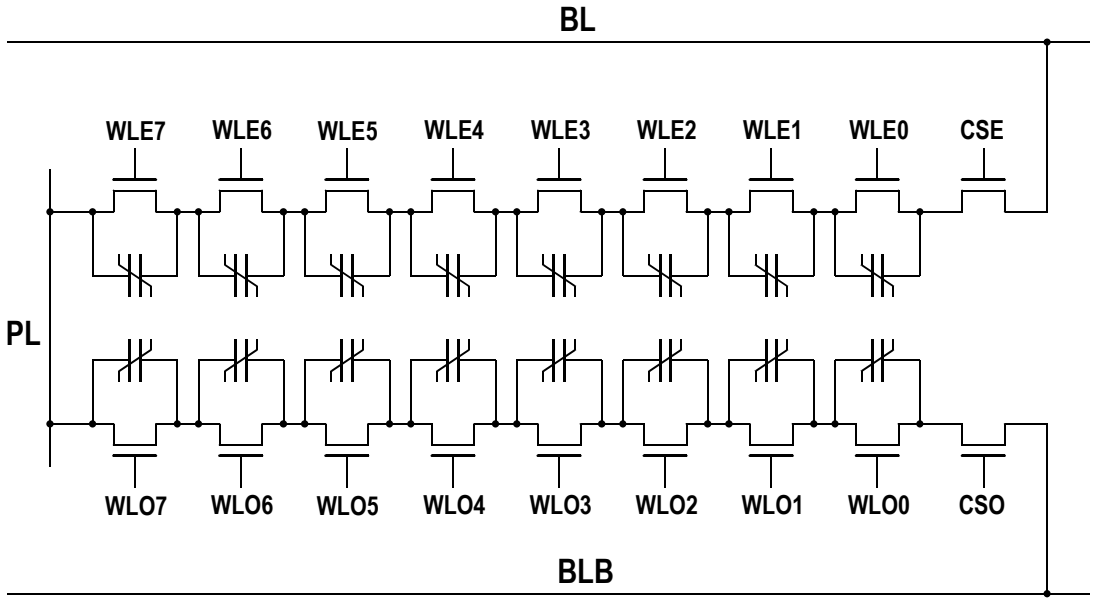


Figure 55: CFRAM block: Bit chain

B.2 Chain FeRAM Block[2]

Bit chain

Figure 55 is a schematic of the CFRAM bit chain pair. One chain is connected to even word lines while another one is connected to odd word lines. Both chains share a single plate-line. This allows the chain FeRAM to be operated in either 2T2C or 1T1C mode. In contrast to FeRAM, the cell capacitors in the unselected chain do not experience disturb pulses in 1T1C operation because they are shorted via their cell transistors. The ferroelectric capacitor size is 1.05 μm² for all capacitors in the chain.

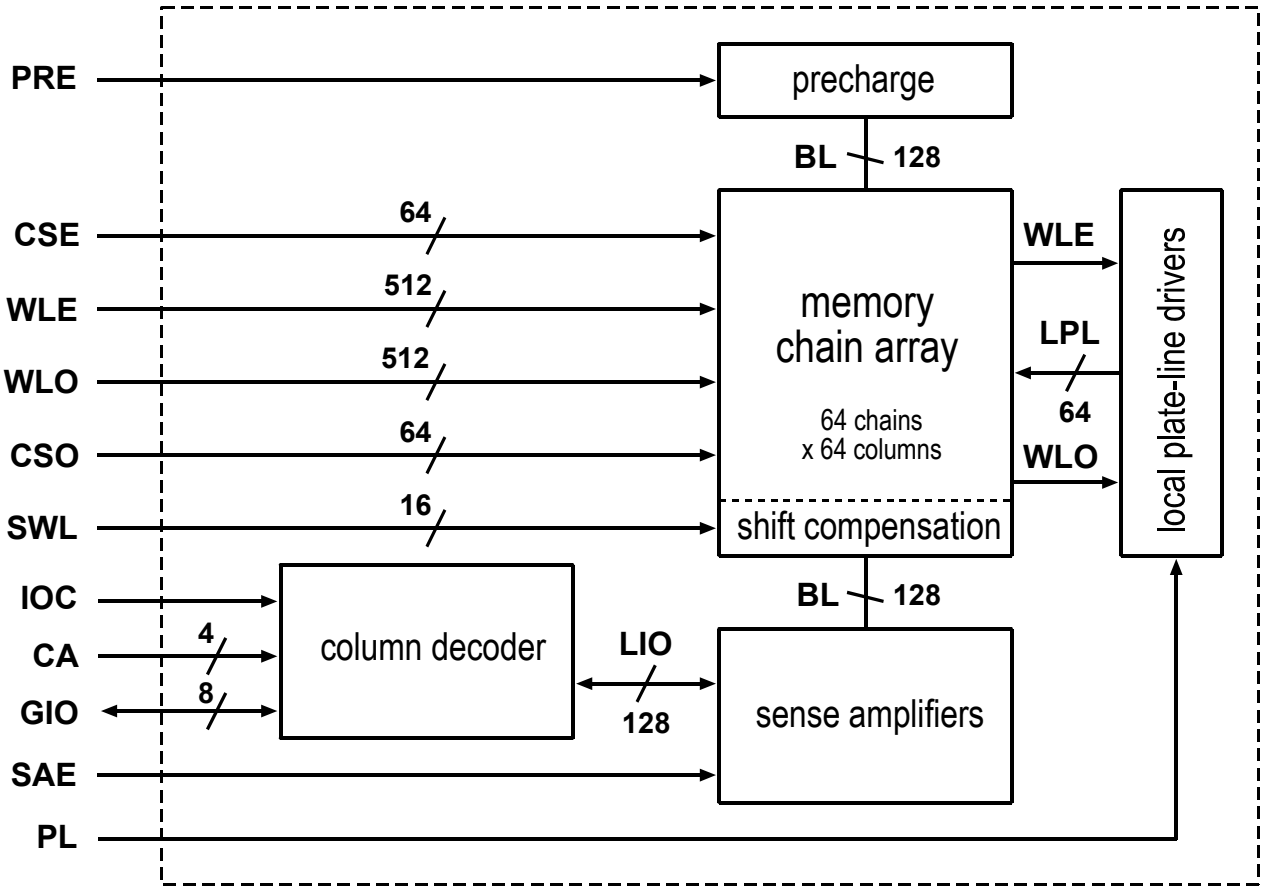


Figure 56: CFRAM block: 64-kbit segment

Segment

64 chains are connected to each bit-line. Similar to the FeRAM, a segment consists of 64 bit-line pairs. Each of the 64 chains has a separate local plate-line that is connected

[2] see Chapter IX for more information about Chain FeRAM

to a global plate-line. Since each chain has 16 word lines (even and odd), there are a total of 1024 word lines. Also each bit-line is connected to a readout voltage shift compensation circuit. The rest of the periphery is identical to the periphery for the FeRAM segment. A simplified schematic of the segment is shown in Figure 56. The segment has a total capacity of 64-kbit (64 x 64 x 16 bit).

Readout voltage shift

In CFRAM, the voltage that is established on the bit-line after readout depends on a memory cell's position within the chain because the effective bit-line capacitance is different for each cell position. This variation in bit-line voltage has been reported as readout voltage shift [30]. For this chip, a new circuit has been implemented in order to compensate the readout voltage shift. It comprises a replica chain without ferroelectric capacitors in which the order of the word lines is reversed. This chain adds a complementary amount of capacitance to the bit-line, depending on what cell position within the chain is accessed. It manages to successfully eliminate the readout voltage shift. More details about this circuit are presented in Chapter IX.

Memory array

The memory array architecture of the CFRAM is identical to that of the FeRAM described earlier. Therefore, both arrays share the same external interface signals. The only difference is that the CFRAM has two additional signals, namely CSE and CSO, that activate the block-selecting transistors shown in Figure 55.

C. CHIP LAYOUT

To reduce manufacturing process complexity, only two of five metal layers were available for layout. This circumstance made the layout of the chip sometimes troublesome. For example, the standard cell library that was available for the full 5-metal CMOS process had to be manually changed to use only two metal layers.

The layout for all leaf cells of the memory array and supporting structures was created manually. Parametric cells were used whenever appropriate. The layout of the CFRAM word-line driver/decoder, especially, was challenging since the word-line pitch was very tight and due to the availability of only two metal layers. Also, special attention had to be paid to the layout of the sense amplifier to ensure that the parasitic

loads were as balanced as possible. Any layout imbalance directly translates into an offset voltage in the sense amplifier. The layout of memory array itself was automatically generated using a memory compiler. This allowed for quick adjustment of the number of cells per column or row without having to hand-edit the layout. A place-and-route tool was used to perform the top-level cell placement and signal routing.

Layout verification

DRC

DRC (design rule check) was performed in two ways – during the artwork creation of single cells and prior to LVS for the full chip. Full-chip DRC is very time-consuming as a single run might take hours or even days depending on the chip size and/or method of DRC (hierarchical or flat). Running DRC while creating the artwork of each cell offers the advantage of catching possible design rule violations as early as possible thereby minimizing turn-around time. However, full-chip DRC is necessary to verify there are no design rule violations due to abutting or spacing of neighboring "DRC clean" cells (i.e. cells that passed DRC).

LVS

LVS (layout versus schematic) is used to verify that the artwork matches the schematic in every possible aspect (signal connections, transistor sizes, capacitances, signal names, etc.) and no errors have sneaked in during artwork creation. If there are errors in the artwork that impair the functionality of the chip that are not in the schematic, they cannot be found using circuit simulation of the schematic. LVS will reveal any differences between the artwork and schematic and thus allow the engineer to correct either the artwork or the schematic – whichever is wrong.

Full-chip LVS verification was performed using flattened netlists for layout and schematic. This process is also very time-consuming since the extraction step for the layout netlist alone (the netlist has to be generated from the artwork) takes about the same time as a full-chip DRC. Schematic netlist generation is usually very fast because only a format conversion step is required. After both netlists are available, they are compared against each other and any differences are flagged as errors or warn-

ings. After resolving detected discrepancies, the whole process has to be repeated until the design is “LVS clean”.

VI. REQUIREMENTS AND SPECIFICATION OF THE 4-MBIT TEST CHIP

The next three chapters describe work that was performed as part of a joint development program between Agilent Technologies and Texas Instruments. The planning and design phase as well as part of the testing of a 0.13-µm 4-Mbit embedded FeRAM test chip are presented. The first of the three chapters addresses the requirements and specification of the test chip.

A. GOALS AND REQUIREMENTS

The primary goal of this joint development program was to create a vehicle that could be used to evaluate and verify the effects of integrating the PZT ferroelectric process loop. The vehicle should provide feedback to process development and help to investigate memory failure modes caused by material issues such as imprint, fatigue, and retention loss. Some of the key requirements were to be able to measure the switching and non-switching polarization of each cell, and to be able to exercise the memory cells 10^{12} times in a very short time. It should also exhibit how well the ferroelectric process integrates with a standard CMOS process to ensure that it does not affect product reliability. Furthermore, it was desirable to be able look at different capacitor sizes. The same chip should be able to be operated as a regular memory with full external control over all major internal timing signals. It was not a priority to achieve high area efficiency, the fastest possible circuit operation, or production-ready circuits. The chip was intended to be a pure research and development vehicle and not a product prototype.

In the early planning phase, it turned out that this circumstance provided a perfect basis to explore new circuit ideas. It was obvious that the chip's capabilities would exceed what was commonly found on previous FeRAM test chips. For example, it was decided to include a dedicated circuitry that is capable of obtaining the full-chip charge distribution within a few milliseconds. A novel sense amplifier was devised to help to achieve this goal. Also implemented were an accelerated fatigue mode, intended to exercise a selectable group of 8k cells to 10^{13} cycles in less than a few days,

as well as other modes that measure the offset of the sense amplifiers, bit-line capacitance, or the degraded write voltage for "1"s.

B. SPECIFICATION

The chip is an all-level full custom layout utilizing Texas Instruments' 0.13µm CMOS 5-metal copper (Cu) technology with a nominal operating voltage of 1.2 volt. The chip is organized in 128 Kwords by 32 bits. It features three different ferroelectric capacitor sizes of 0.12 μm^2, 0.18 μm^2 and 0.25 μm^2. The die size is 4 mm x 4 mm with only about 2.3 mm x 1.9 mm occupied by the memory array itself (see Figure 57). It also features a 100-pin chip scale package with a 152-pad die. The voltage range that is generated by the on-chip reference voltage generator is externally adjustable.

B.1 Overall Chip Architecture

The chip's main components are the control logic, the memory array, the segment address decoder, the reference voltage generator, and the charge distribution circuit (see Figure 58).

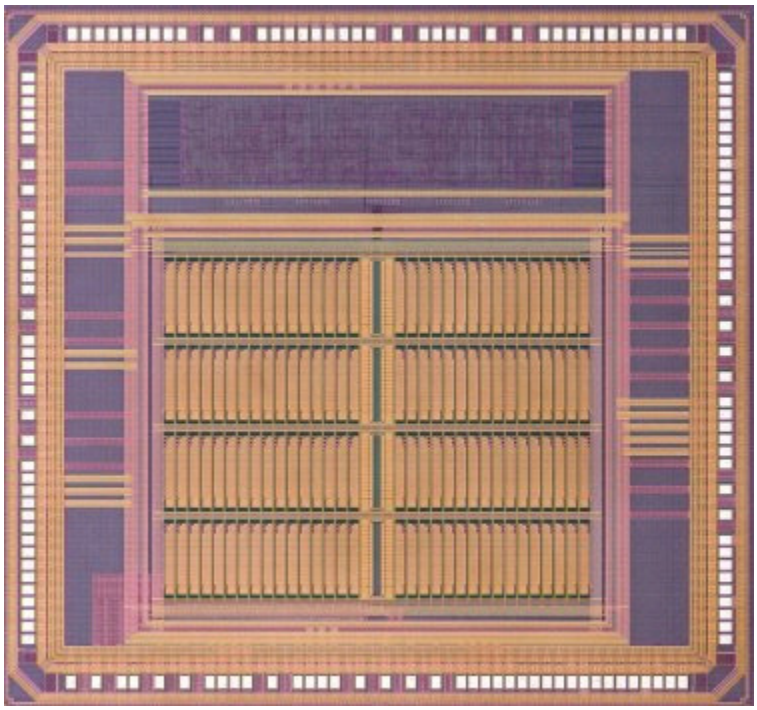


Figure 57: Die micrograph of the 0.13-µm 4-Mbit chip

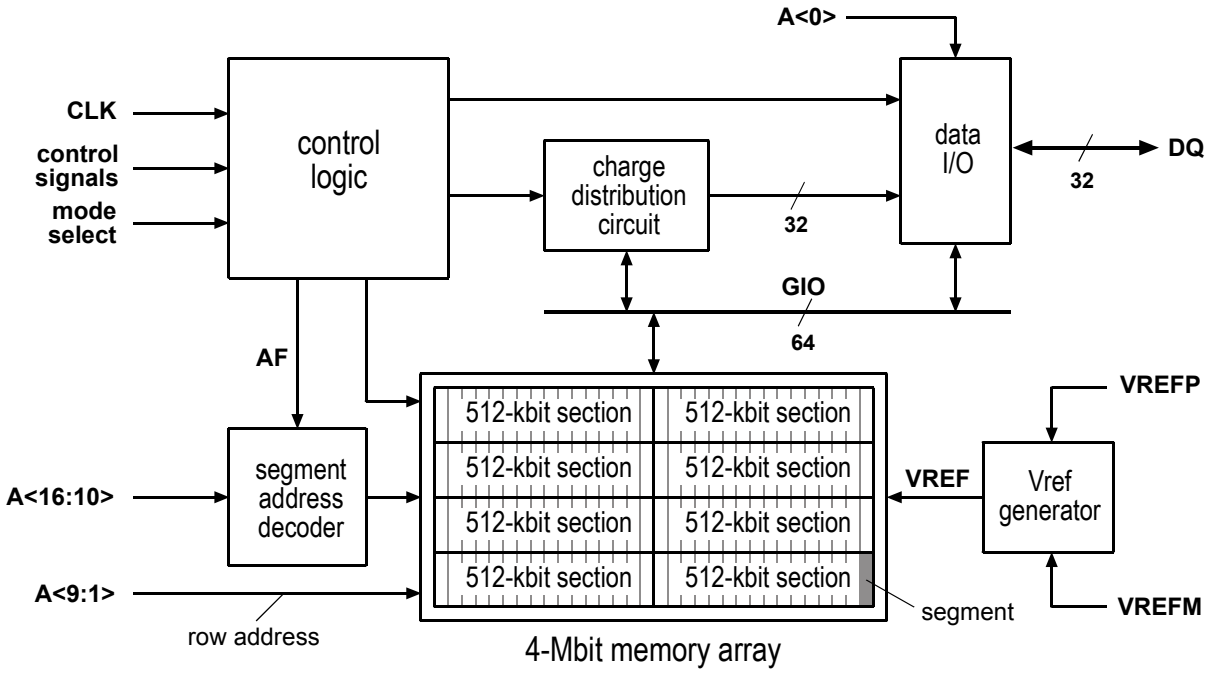


Figure 58: 0.13 µm chip logic diagram

Memory array

The 4-Mbit memory array is divided into 8 sections of 512 kbit. Each section is further divided into 16 segments of 64 kbit. During normal operation, only one segment is active at a time. The active segment is selected by the segment address decoder. Each section includes 512 word lines, which are continuous across the 16 segments in the section as shown in Figure 59.

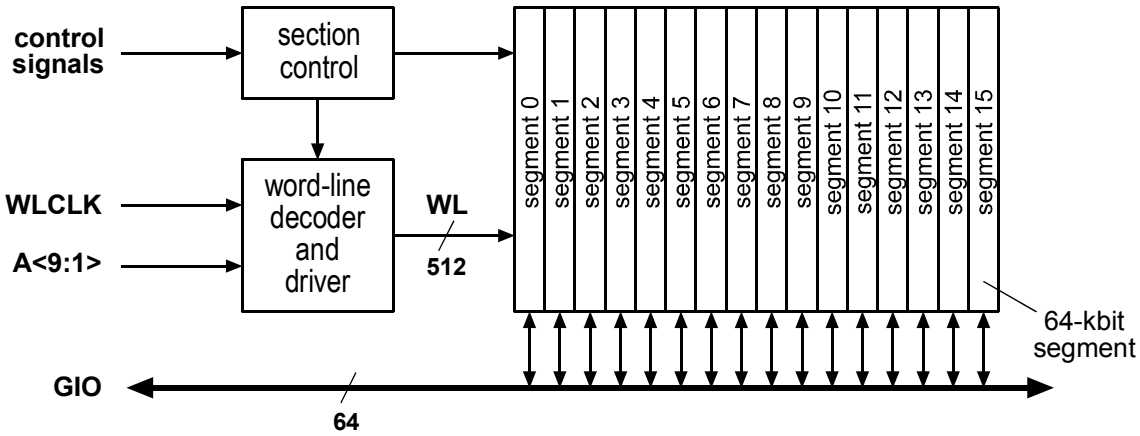


Figure 59: 512-kbit section

Row address 0 is set at the top of the segment and increments sequentially from top to bottom. The segment ordering is from the left side of the section with segment 0 being at the left-most side of each section. Each segment includes 64 bit lines and 16 plate-line groups (see Figure 60). Each plate-line group includes 32 rows that have a common local plate-line and associated plate-line driver. A normal read or write operation accesses 64 cells at a time. Data is brought in and out of the RAM through an external 32-bit bus (DQ) and an internal 64-bit bus (GIO).

Charge distribution circuit

The charge distribution circuit comprises a 7-bit counter that is connected to 64 7-bit registers, and multiplexers. Together with the reference voltage generator and the sense amplifier, the circuit allows the analog voltage of 64 bit lines to be measured simultaneously within only a few microseconds. A resolution of about ±5 mV is achieved.

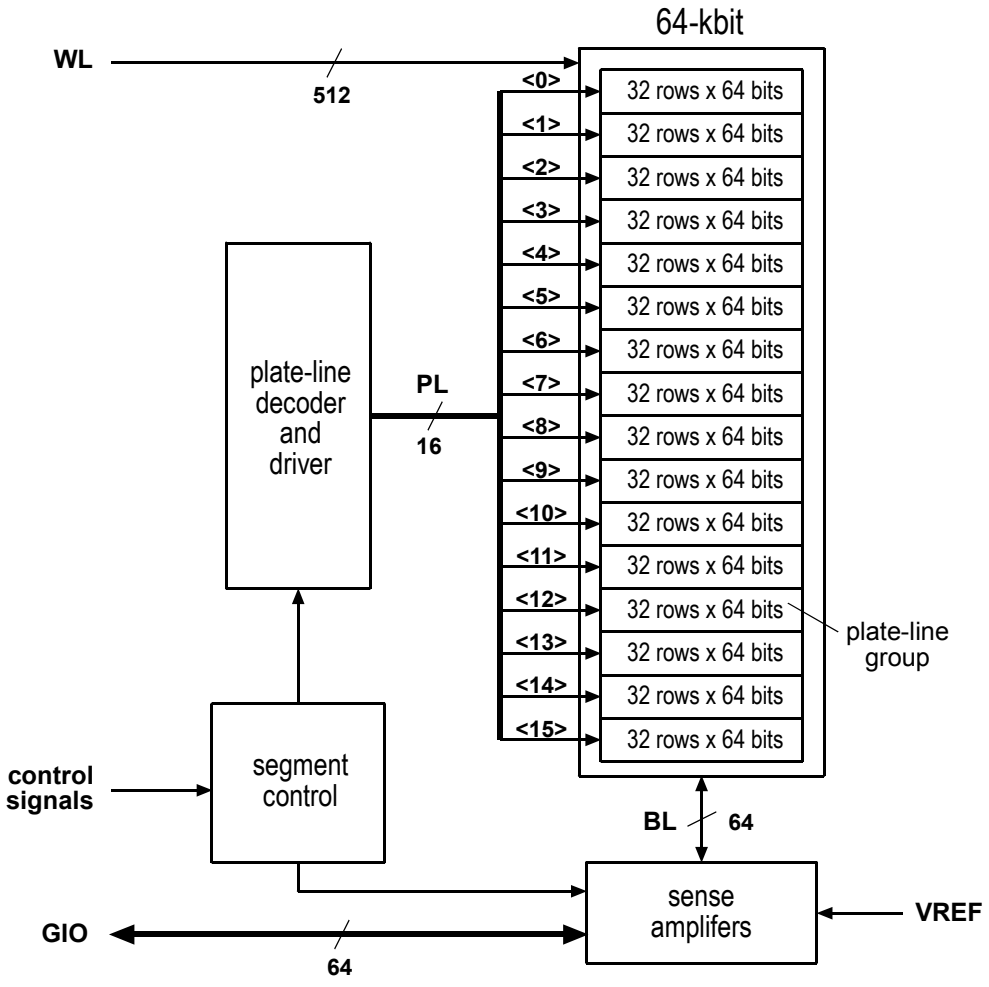


Figure 60: 64-kbit segment

Addressing scheme

Address bits A[16:14] determine which one of the 8 sections is active. Address bits A[13:10] select the segment within the active section. Address bits A[9:6] select the plate-line group within the selected segment. Each plate-line group is subdivided into 32 word lines. Address bits A[5:1] select the word-line to be accessed. Address bit A[0] selects between the upper/lower 32 bits of the 64-bit wide internal data bus.

Table 7: Address assignment

Address	Usage
A [16:14]	Section select. Selects 1 of 8 sections.
A [13:10]	Segment select. Selects 1 of 16 segments within selected section.
A [9:6]	Plate-line group select. Selects 1 of 16 plate-line groups.
A [5:1]	Word-line select. Selects 1 of 32 word lines within selected plate-line group.
A [0]	Word select. Selects the upper or lower 32 bits of the enabled column word.

Reference voltage generation

The on-chip reference voltage generator is able to generate a series of decreasing voltages from VREFP down to VREFM and is mainly used in combination with the charge distribution circuit. Its output VREF is connected to the reference voltage input of all sense amplifiers and is therefore heavily loaded. Despite this circumstance, the reference voltage generator is still capable of changing the output voltage within about 2-ns.

B.2 Modes of Chip Operation

Idle mode (IDLE)

At power-up and during standby, the chip should be in idle mode. In this mode, external control signals to the main controller are ignored and the main control unit outputs default values for all internal signals. This mode is also used as an in-between state when changing modes during the same test run.

RAM mode (RAME)

This mode allows operation of the chip as a regular memory with external control over the chip's internal timing signals. Adjustable timing of word-line, plate-line, and sense amplifier activation, plus pre-charging and write back of data is possible. For example, word-line control is established through the signals word-line clock WLCLK and word-line boost WLBST. Plate-line control is established through the signal PLCTRL. Sense amplifier operation is controlled through the signal SACTRL. Bit-line pre-charging is established through the signal bit-line low control BLLOCTRL. Write back of data is controlled through the signal WBCTRL.

Read operation

The read access is initiated by placing the WR and BLLOCTRL inputs logic-low at time t0 (see Figure 61). Activation of the word- and plate-line via WLCLK and PLCTRL at time t1 causes a current through the ferroelectric capacitors of the selected row to the bit lines. The activated plate-line forces all ferroelectric capacitors in the selected row into the same polarization state, and the amount of charge to each bit-line depends on whether the activated plate-line flipped the polarization state of the ferroelectric capacitor or not. The sense amplifiers are activated by taking the SACTRL input to logic-high at time t2. When sense operation is completed, data from the RAM at the selected address A[16:0] becomes available at the outputs DQ[31:0] at time t3. At the same time, write-back is initiated by taking WBCTRL and WLBST inputs to logic-high. At this time, bit-line voltages are restored to full levels and the word lines are boosted to V_{PP}. Write-back is completed by first taking PLCTRL input back to logic-low at time t4 followed by taking WBCTRL and WLBST inputs back to logic-low and taking BLLOCTRL input back to logic-high. Finally, the read cycle is completed by taking WLCLK back to logic-low at time t5.

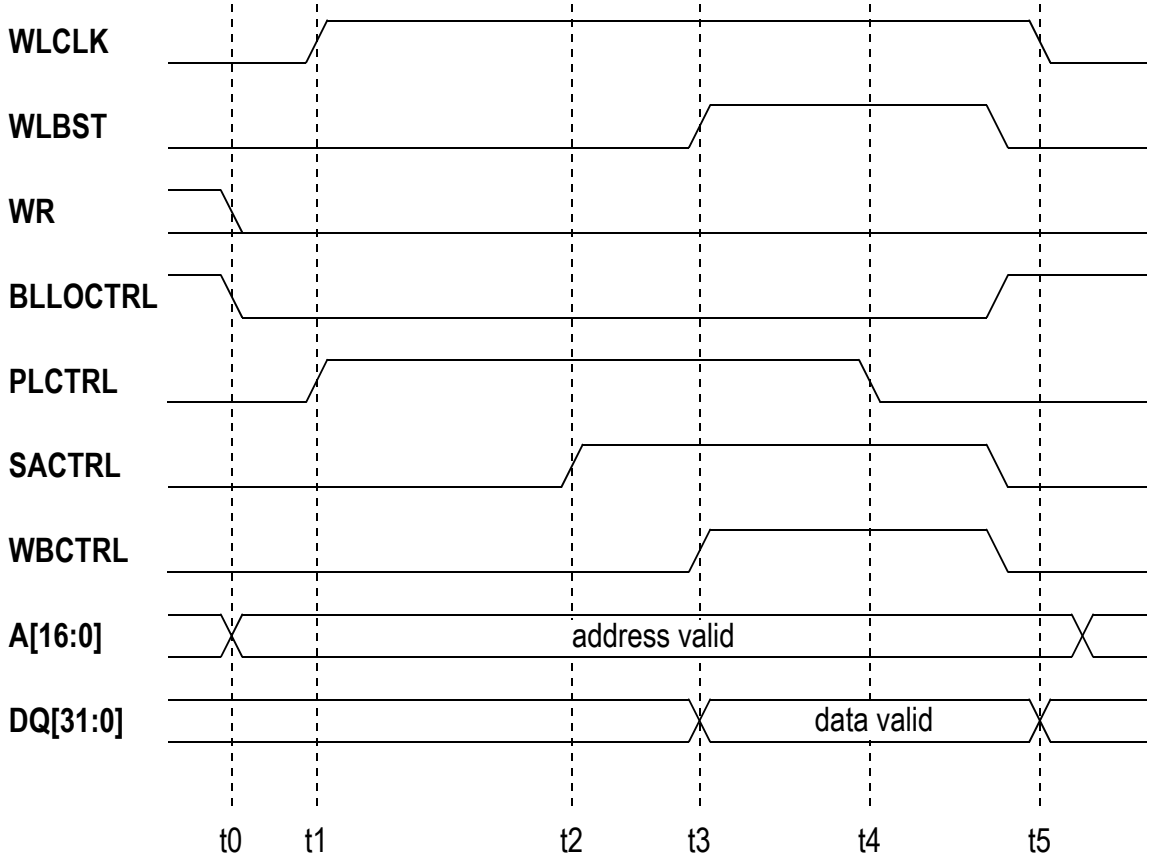


Figure 61: Timing diagram for a read operation

Write operation

The write access is initiated by placing WR input logic-high, applying the lower 32 bits of data to inputs DQ[31:0], applying address bits A[16:1], and setting address bit A[0] to logic-low at time t0 (see Figure 62 below). After address bit A[0] has been toggled to logic-high at time t1, the upper 32 bits of data are applied to inputs DQ[31:0]. When BLLOCTRL input is placed logic-low at time t2, the GIO bus that carries the data to be written starts to drive the bit lines of the selected segment. WLCLK and PLCTRL are taken to logic-high at time t3 to activate the word- and plate-line. Next, the write to the memory cells is accomplished by first taking WLBST to logic-high at time t4 followed by taking PLCTRL input back to logic-low. Taking WLBST input back to logic-low and taking BLLOCTRL input back to logic-high concludes the write access. Finally, the write cycle is completed by taking WLCLK back to logic-low at time t5.

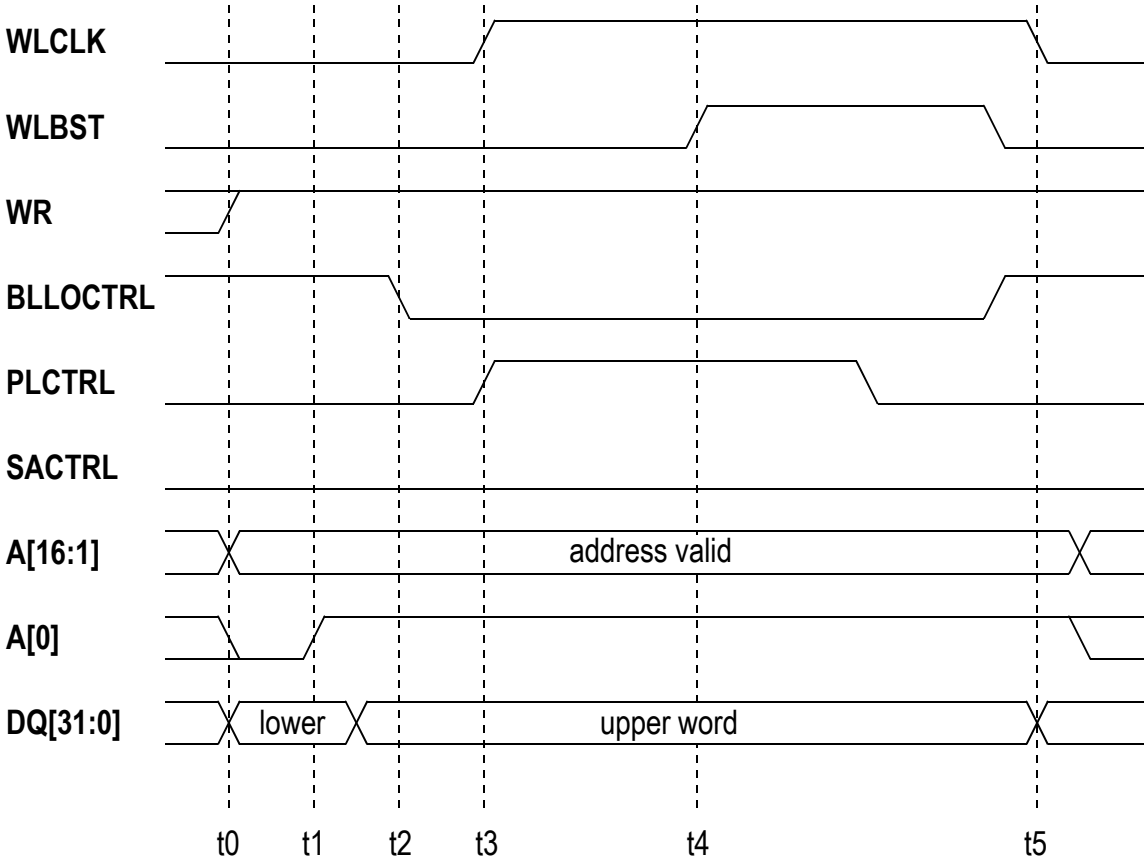


Figure 62: Timing diagram for a write operation

Charge distribution mode (QD)

Development, manufacture, and use of a semiconductor memory such as FeRAM generally requires testing that determines the characteristics of the chip and determines whether it is functioning properly. Generally, the bit-line voltage that results from reading a bit cell varies according to the performance of the particular cell being read. A chip's cell charge distribution is of greatest interest since it allows the optimum reference voltage to be determined as well as the manufacturing process to be monitored. Measuring the charge distribution can indicate whether a particular memory is defective or whether a repair operation is required.

The QD mode is a mode of operation that enables the quick characterization of a chip's memory cells. It measures the amount of charge dumped out onto the bit-line for each cell for both "0" and "1" state. Hereby, the QD mode makes use of extra circuitry to speed up the process of comparing a bit-line voltage against multiple reference voltages in order to determine a bit-line's analog value (see Chapter VII). A single meas-

urement is conducted for 64 capacitors in parallel and takes about 500 ns. A full-chip measurement allows obtaining the charge of all 4,194,304 memory cells within about 45 ms.

Accelerated fatigue modes (AF1PG & AF16PG)

A ferroelectric capacitor's ability to store non-volatile charge decreases with the number of polarization reversals. To characterize the magnitude of this effect as thoroughly as possible, memory capacitors must be exercised for as many cycles as time allows under various conditions. If the test indicates that the memory can be theoretically operated $>10^{16}$ times, a 10-year lifetime can be guaranteed. However, for a typical cycle time of 80 ns it takes about 2 days of test time to exercise a memory cell 10^{12} times. To reduce test time, two dedicated chip modes have been implemented. The accelerated fatigue modes AF1PG and AF16PG are significantly faster in exercising the ferroelectric capacitors for fatigue testing than regular write cycles (see page 95ff.).

Sense amplifier calibration mode (SACAL)

An ideal sense amplifier should output a "0" if the voltage applied to its first input is smaller than the voltage applied to its second input, and output a "1" if the voltage applied to its first input is larger than the voltage applied to its second input. If the same voltage is applied to both inputs, the sense amplifier should output a "0" or "1" in 50% of the time. However, process-related variations cause a mismatch between otherwise balanced transistors. Transistor threshold voltage mismatch, especially, is a serious problem in 0.13 µm technology and beyond. Any transistor mismatch is reflected in an offset voltage of the real sense amplifier, which requires one input to be larger at least by the same magnitude as the offset voltage before the sense amplifiers' output changes from "0" to "1" or vice versa. Since the offset voltage varies for each sense amplifier on the chip, the overall effect for a memory chip is a reduction of available signal margin. For a test chip, the main disadvantage would only be a loss of measurement accuracy.

In sense amplifier calibration mode, the offset voltage of each sense amplifier can be determined exactly. This is accomplished by precharging all bit lines to ground and comparing against a series of reference voltages. One input of each sense amplifier is always connected to a bit-line, while the other input is connected to the reference volt-

age. The comparison results enable one to exactly determine the offset voltage for each sense amplifier. The measured offset voltages can be used to improve the accuracy of other measurements or to characterize the sense amplifier itself.

Bit-line capacitance measure mode (BLMSR)

The mode provides a method for measuring the capacitance of the bit lines. Connected to each bit-line through pass transistors there is an array of 4 binary weighted thin oxide capacitors. These capacitors are individually selectable. This gives 2^4 possible combinations. The values of these large capacitors are known with a good amount of accuracy since they are deliberately large and square. The voltage applied across them is fixed. Thus, knowing C and V and that Q = C*V, known amounts of charge can be dumped onto the bit-line. Then, the charge distribution circuitry is used to measure the resulting voltage that has been established on the bit-line. With both the bit-line voltage and the amount of charge that has been dumped known, the capacitance of the bit-line can be calculated.

Storage node measure mode (SNMSR)

To prevent the degradation of the V_{DD} level from bit-line to storage node SN, a voltage higher than V_{DD} must be applied to the gate of the word line transistors. This technique is known as word-line boosting. The boosted word-line voltage is often referred To as V_{PP}. Since V_{PP} has to overcome the threshold voltage drop, it needs to be larger than $V_{DD}+V_{th}$. This can be relatively large compared to V_{DD} (e.g. for V_{DD} = 1.5V, V_{PP} > 2.3V) and could degrade or even destroy the gate oxide of the device. To overcome this problem, usually devices with a thicker gate oxide are used inside the memory array. These high-voltage devices require extra mask levels and this makes the solution expensive for embedded memories. As cost is a driving

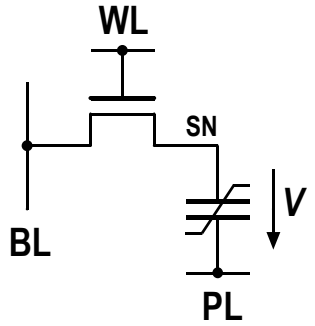


Figure 63: Storage node voltage SN

factor for embedded memories, ideally one would like to use the same standard logic transistors also for the array with no need for extra masks. However, because operating voltages are very small for future FeRAMs, this goal will be even harder to achieve

as V_{th} does not scale very well. Therefore even for operating voltages of 1.0 V, V_{PP} has to be almost 0.8 V larger. On the other hand, V_{PP} will be limited due to thin gate oxide failure mechanisms. The ability to exactly determine the required minimum V_{PP} by measurement is therefore desirable. The storage node measure mode allows the dependence of the degraded storage node voltage on V_{PP} to be determined.

VII. DESIGN AND IMPLEMENTATION OF THE 4-MBIT TEST CHIP

This chapter describes the key circuits and techniques of the 0.13-μm 4-Mbit embedded FeRAM test chip.

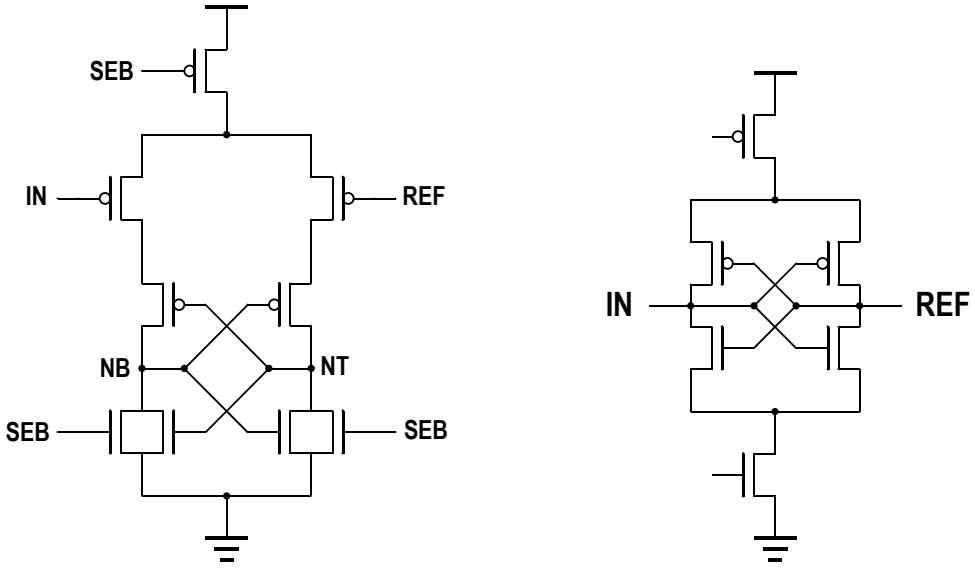
A. SENSE AMPLIFIER

The digital differential comparator, shown in Figure 64a, has been used in DRAMs as well as SRAMs. The top p-channel transistor has its source connected to power and its drain connected to the left and right leg. Each leg includes two serially connected p-channel transistors followed by two parallel-connected n-channel transistors. However, this circuit usually finds application in the data path external to the memory array itself, amplifying the data signal received from the memory array and passing it on to the output buffer. Since this circuit lacks any ability to write-back onto the input nodes, and since it is somewhat more complex than the traditional latching sense amplifier, shown in Figure 64b, it does not normally find use in the memory array itself. However, this comparator has distinct advantages over the latching sense amplifier of Figure 64b. The primary advantage it possesses is its speed when amplifying a small differential input signal. Since the output nodes, or latching nodes of the comparator, nodes NT and NB, are typically lightly loaded capacitively, they are very quickly driven to their respective output logic levels. For most memory applications, the input node (or nodes) is heavily loaded capacitively. By separating the lightly loaded output from the heavily loaded input, the comparator is able to quickly amplify the small differential input signal and establish a large differential output signal. There are several characteristics of ferroelectric memories that make the application of a comparator attractive.

Disadvantages of the conventional sense amplifier

Typically, in DRAMs the bit lines are pre-charged to a mid-V_{DD} level prior to the time of sensing. This level is generally high enough that the n-channel cross-coupled devices in the sense amplifier are able to amplify the difference signal. Since the mobility

of n-channel transistors is typically twice to three times that of p-channel transistors, most of the amplification occurs from the n-channel pair rather than the p-channel counterparts, even for simultaneous activation of both pairs. In the FeRAM case, however, the bit lines are pre-charged to ground, at least for the higher coercive voltage ferroelectric materials. N-channel sensing and amplification is consequently not possible since the resultant bit-line voltages are at, or below, the n-channel threshold voltages of the process. If a traditional latching sense amplifier were to be used, the sensing and amplification would be done exclusively by the p-channel cross-coupled pair, thus reducing the sensitivity of the amplifier and the speed of amplification.



(a) comparator *(b) latching sense amplifier*

Figure 64: Sense amplifiers

An additional delay in accessing the data from the memory array results when a traditional latching sense amplifier is used with ferroelectric storage cells. When the stored polarization of the cell is restored, at the time that charge is required by the storage cell, this charge is pulled from the bit-line, temporarily reducing the difference voltage between the bit lines. Since an adequate voltage separation between bit-line and bit-line bar must occur before they can be switched to the output data bus, additional time is required to charge and discharge the bit-line. An additional delay is thus introduced while the sense amplifier increases the difference voltage after the polarization charge in the cell has been removed from the restoring bit-line.

A.1 Comparator with Write-back

The comparator with write-back sense amplifier, illustrated in Figure 65, solves the issues discussed above by separating the heavily loaded bit-line input capacitance from the lightly loaded output capacitance. In addition to the electrical connectivity shown in Figure 64, another n-channel transistor connects nodes N1 and N2 and is used for equalization.

During operation, the comparator compares the voltage on the bit-line BL to the voltage on the reference input REF. The lack of write-back capability inherent in the standalone comparator is overcome by the addition of a tri-state write-back driver tied to the bit-line. It receives the comparator's output bar signal NB and is enabled by write-back signals WB and WBB. These additional complementary control signals are asserted after the comparator has been activated and after NT and NB have been driven to their full logic levels. Simultaneous to the write-back restore on the bit-line, data can be accessed via the sense amplifier output enable signal SOE and started on its path to the chip's data output buffer, irrespective of the time required to accomplish the write-back. The write-back circuitry must be tri-stated during the sensing operation since, otherwise, it will destroy the signal intended to be sensed and amplified.

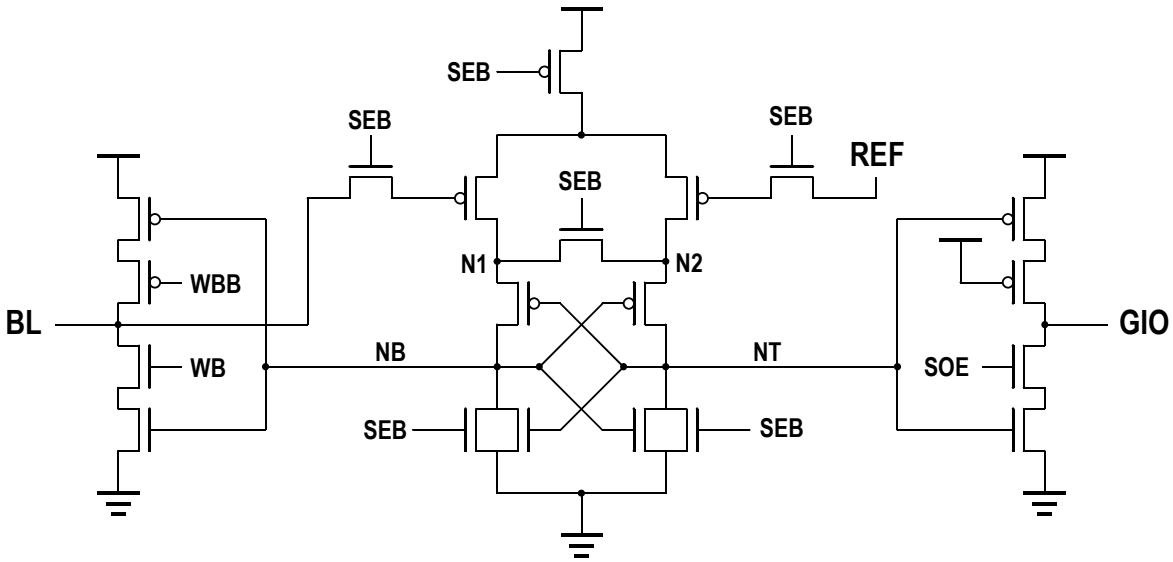


Figure 65: Schematic of sensing circuitry

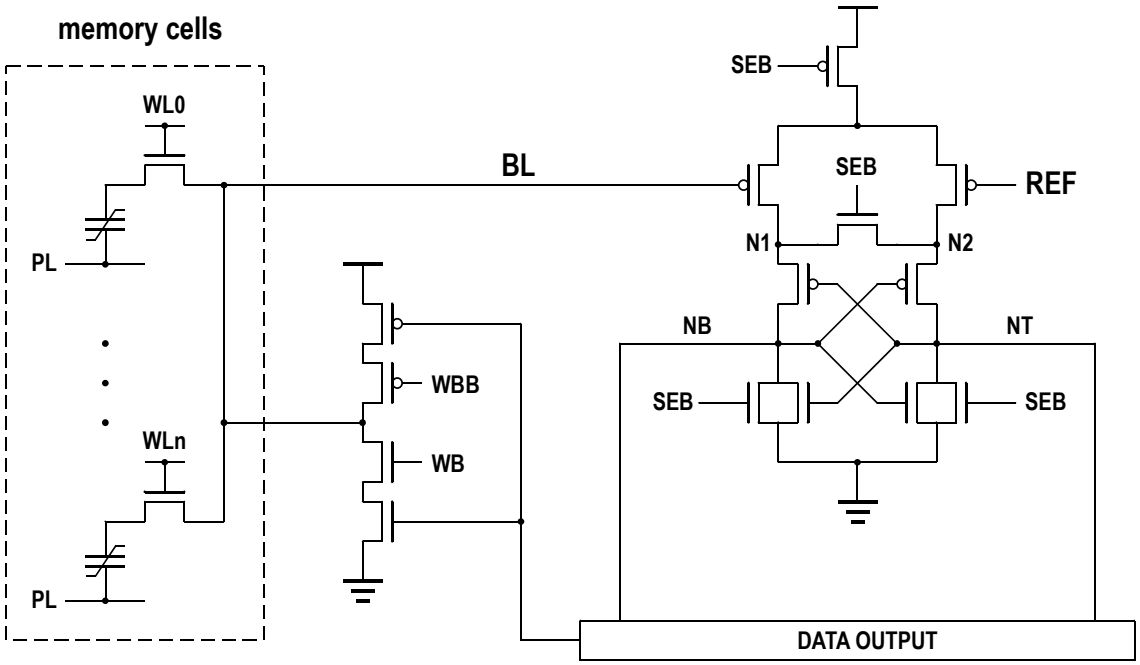
Another significant advantage of this sense amplifier is that no longer is there any need to balance the capacitance on the bit-line input and reference input nodes, as in the traditional latching sense amplifier. The reference input can be the traditional corresponding bit-line complement or can be developed in any fashion desired and distributed to the sense amplifiers in the array. This gives the designer maximum flexibility in creating the reference voltage to the sense amplifier and offers many options that circumvent the reference voltage issues widely discussed in the FeRAM literature.

A.2 Multiple-Comparison Operation

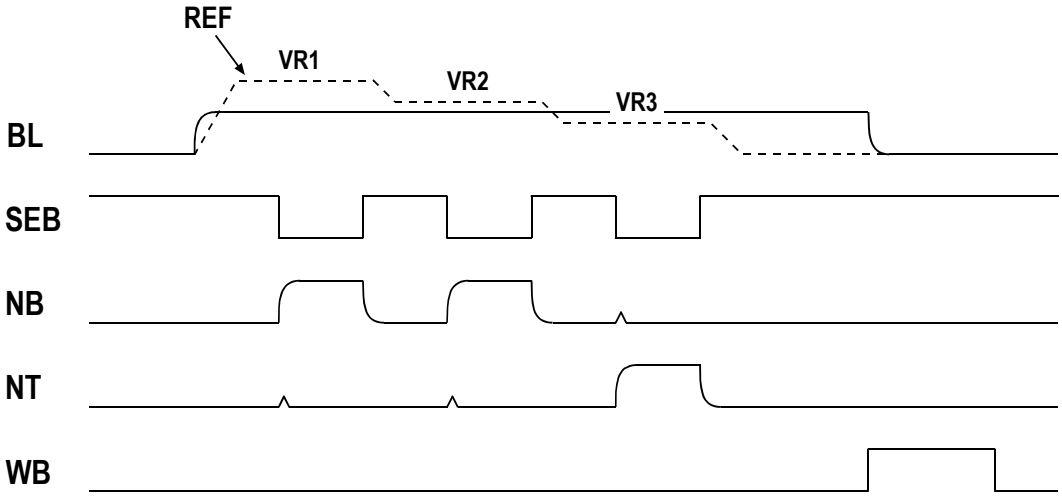
Since the sense amplifier does not disturb the bit-line voltage, the bit-line voltage can be maintained while a fast series of sensing operations is performed with different reference voltages. One non-destructive sensing operation can be completed within 1-2 ns and is therefore quick enough to make a multiple-comparison operation practical. A multiple-comparison operation could characterize the performance of a ferroelectric memory cell, or FeRAM devices that store multiple bits or levels of information in a single memory cell could use a multiple-comparison operation to quickly read a multi-bit value from the cell. Figure 66a shows a portion of an FeRAM to illustrate the multiple-comparison operation. It contains a column of memory cells, the comparator-type sense amplifier plus write-back driver, and a data output circuit. One input of the sense amplifier is connected to the bit-line while the other one receives a reference voltage.

The data output circuit receives the complementary output signals NT and NB and is also intended to store the results from the comparisons of the bit-line voltage to the series of reference voltage levels. The data output circuit could contain a latch or flip-flop for storing the comparison results. The write-back circuit is controlled by the data output circuit. Figure 66b shows a timing diagram for the multiple-comparison operation. To prepare the sense amplifier for operation, sense amplifier enable bar signal SEB is initially deactivated (V_{DD}) to shut off power in the sense amplifier, to ground nodes NT and NB, and to equalize voltages on nodes N1 and N2. To begin the multi-comparison operation, the word and plate-line voltages are raised to V_{DD}. The activated word-line turns on the access transistor and electrically connects the selected

ferroelectric capacitor to the bit-line. Reference voltage VR1 is applied to the REF input of the sense amplifier. Then, the sense amplifier is activated by driving enable signal SEB to V_{SS}. Complementary output signals NT and NB represent the comparison result that indicates whether the bit-line voltage is greater or less than VR1. This result is temporarily stored in the data output circuit.



(a) schematic



(b) timing diagram

Figure 66: Schematic and timing diagram to illustrate multiple-comparisons operation

Without changing the bit-line voltage, the multi-comparison operation deactivates enable signal SEB and changes the reference voltage to the next level VR2 before reactivating SEB for a second comparison. The second comparison result indicates whether the bit-line voltage is greater or less than the second level VR2, and the data output circuit also stores the result of the second comparison. Further comparisons can be conducted in the same fashion. In particular, the multi-comparison operation deactivates SEB and changes the reference voltage to the next level without changing the bit-line voltage and then reactivates SEB for the next comparison. Each comparison indicates whether the bit-line voltage is greater or less than a corresponding level of reference voltage and the data output circuit temporarily stores the result of the comparisons. After the comparisons to all of the desired reference voltage levels VR1, VR2, …, the enable signal SEB is deactivated, and write-back signal WB is activated. The data output circuit controls the voltage that the write-back driver forces on the bit-line so that the original data value is rewritten to the accessed bit cell.

Possible applications of the multiple-comparison operation

The comparison results can be used for a variety of purposes in different applications. An on-chip bit error detection circuit could use the comparison results to determine whether the selected memory cell provides a signal voltage that is large enough for reliable operation. For example, reference voltage level VR1 could be chosen to be slightly larger than V_{BL1}, the expected bit-line voltage for a "1", while the other reference voltage levels VR2 and VR3 are lower than V_{BL1} but higher than V_{BL0}, the expected bit-line voltage for a "0". A multiple-comparison operation that finds a bit-line voltage lower than level VR1 but higher than level VR2 or VR3 indicates that the selected bit cell was storing a "1" but did not provide the expected bit-line voltage. An error signal could then be generated, or the bit cell could be replaced with a redundant bit cell to prevent data errors.

Another use of the multiple-comparison operation is to allow storage of more than one binary value in one memory cell. If, for example, the memory cell had polarization states of different polarization magnitude and corresponding to different data values, the levels VR1, VR2, … of reference voltage can be the boundaries of ranges for bit-line voltages corresponding to the different polarization states. The comparison results in-

dicate a voltage range for the bit-line voltage of the selected bit cell and therefore indicate the stored data value. The multiple-comparison operations as described above allow on-chip bit failure prediction, detection, and correction and provide a very fast and efficient way to capture charge distributions as described next.

B. Measurement of Cell Charge Distribution

A recently published article [31] presented a technique to measure the charge distribution for a 4-Mbit FeRAM (see Figure 67). This measurement included writing a "0" or "1" to the memory cells, charging the reference bit lines to one of a series of voltage levels, reading out the charge from the memory cells to the bit lines, and comparing the voltage of each bit-line to the voltage of the corresponding reference bit-line using a sense amplifier. For each reference voltage used in measuring the distribution, the process must repeat setting the reference voltage and reading out the charge from the memory cells. Further, the entire series is repeated for each data "0" and "1". The charge distribution measurement for all cells in a 4-Mbit FeRAM can in fact take several minutes. The time required for the measurement may be acceptable during development of an integrated circuit when a relatively small number of devices are tested, but in production, such a lengthy test time can reduce manufacturing throughput and increase costs. For on-chip testing, the lengthy test time is unacceptable.

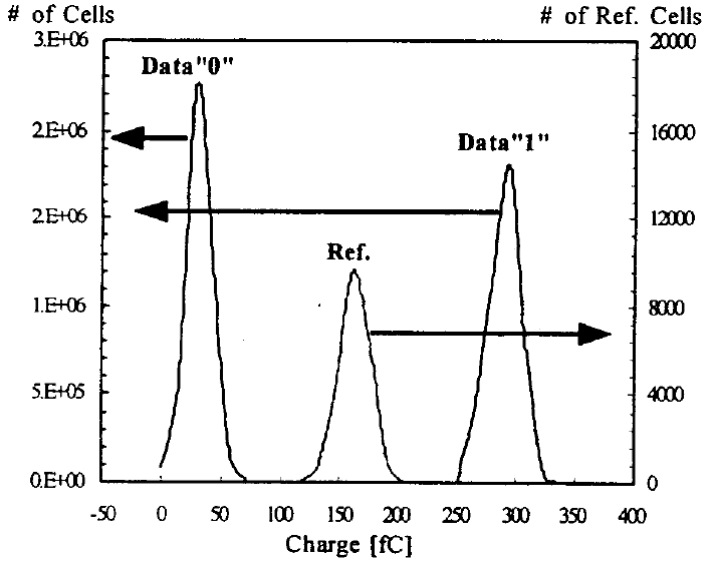


Figure 67: Example FeRAM memory cell charge distribution [31]

Another problem is that each readout operation generally writes-back or refreshes the data value stored in the memory cell. This can lead to inconsistency in the distribution measurement when the charge from a ferroelectric capacitor varies for each readout operation. The repeated readout operations that refresh the stored state can mask retention problems or time-dependent changes in delivered charge since the charge read out almost always corresponds to a freshly written or refreshed data value.

B.1 High-Speed Charge Distribution Measurement

In the following, a new technique for charge distribution (QD) measurement is presented. It employs the earlier introduced comparator-type sense amplifier that compares the bit-line voltage to a series of reference voltages as described in the multiple-comparisons operation. The same bit-line voltage is retained throughout the series of comparisons, which avoids delays, measurement inconsistencies, and endurance problems that arise in conventional processes that read a memory cell repeatedly for a series of comparisons. The process consists of:

(a) Reading out the charge from a row of cells to the corresponding bit lines.

(b) Biasing the reference line to the first/next voltage from a series of reference voltages.

(c) Comparing the voltage on the reference line against the voltages on the bit lines.

(d) Keeping the charge/voltage on the bit lines constant while repeating steps (b) and (c) until the reference line has been biased to a last reference voltage of the series.

Figure 68 shows the portion of the chip implementing the charge distribution measurement. It contains a set of memory cells, a sense amplifier, a reference voltage generator, and a charge distribution circuit. The control logic operates the reference voltage generator and directs the sense amplifiers to measure bit-line voltages. It causes the reference voltage generator to sequentially supply to the sense amplifiers a series of reference voltages, while the bit-line voltage remains constant. For each of the reference voltages, the sense amplifier generates a binary output signal indicating whether or not the bit-line voltage is greater than that reference voltage while leaving

the bit-line voltage undisturbed. The reference voltages decrease in equal steps so that the comparison result from the sense amplifier changes when the reference voltage is about equal to the bit-line voltage. Accordingly, the distribution measurement can be performed quickly and accurately without repeated write and readout operations that could modify the performance of the FeRAM cell.

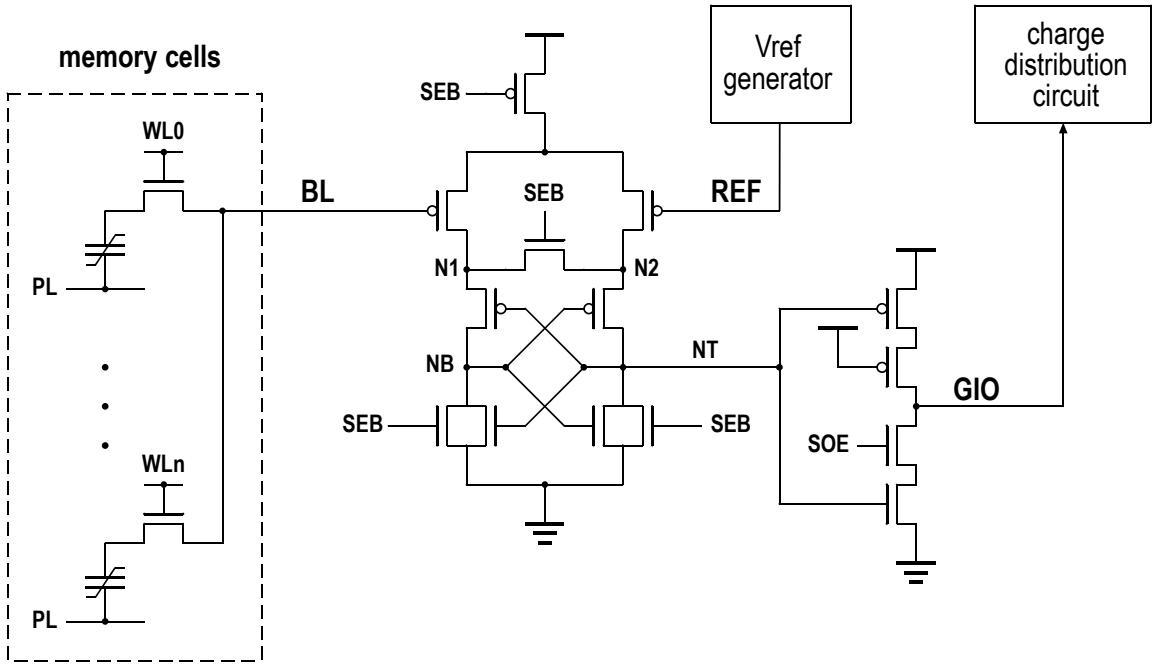


Figure 68: Circuitry associated with QD measurement

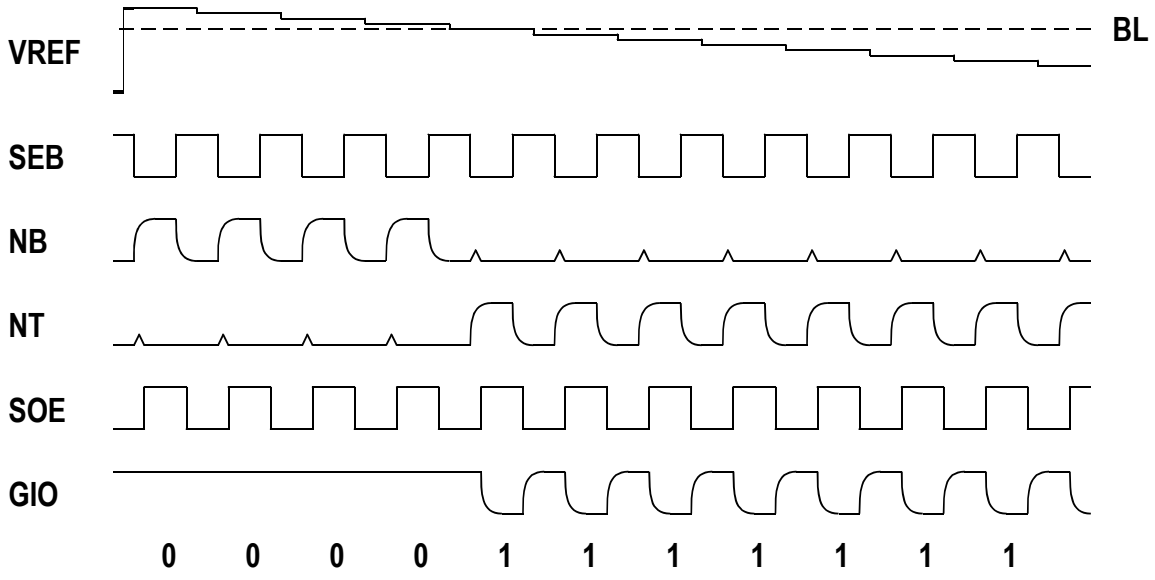


Figure 69: Timing diagrams for selected signals in QD-Mode

Figure 69 shows timing diagrams for selected signals during execution of QD mode. Reference voltage V_{REF} steps through a series of voltage levels. The voltage on the bit-line has been established during read out of the cell and remains constant while being measured. Sense amplifier enable bar signal SEB is activated (low) in a series of intervals during which the reference voltage remains constant. When signal SEB is active, the sense amplifier compares voltages BL and V_{REF}. Depending on whether BL or V_{REF} is at a higher voltage, node voltage NB or NT rises to supply voltage V_{DD}, and the other node voltage NT or NB settles back to zero volts. Sense amplifier output enable signal SOE is activated (high) a short delay after signal SEB. The delay is sufficient for node voltages NT and NB to separate to full levels indicating the results of the comparison of voltages BL and V_{REF}. As a result, the sense amplifier's output driver either leaves the GIO bus line at the precharged level (V_{DD}), indicating the bit-line voltage is greater than the reference voltage, or pulls the GIO bus line down indicating bit-line voltage is less than the reference voltage. During the series of intervals when SOE is activated, each GIO bus line indicates a series of binary values indicating the results of the voltage comparisons. For 100 different reference voltages, the GIO bus line serially provides 100 bits of data representing different comparison results. The appropriate range of voltage levels depends on the properties of the ferroelectric capacitors and particularly on the expected range of bit-line voltages. In this example, V_{REF} ranges from 600 mV down to 100 mV in 100 steps of about 5 mV.

In Figure 69, the reference voltage starts at the upper limit of the voltage range and steps down, but V_{REF} could also increase in steps from the lower voltage limit or change in any desired pattern. For the case where V_{REF} steps down monotonically, the output bit stream associated with a bit-line has a value of "1" until V_{REF} falls below BL. Thereafter, the bit stream is expected to have the opposite value of "0". The measurement of bit-line voltages is conducted simultaneously for 64 bit lines resulting in the output of 64 parallel bit streams. The number of parallel bit streams is limited by the number of sense amplifiers that can be operated in parallel and by the width of the internal data bus.

B.2 On-Chip Compression of Distribution Data

The bit streams associated with a set of parallel comparisons could be directly output for external analysis. However, a full-chip measurement of the charge distribution for a large FeRAM would require the transfer of a significant amount of data. In particular, for 100 different reference voltages and separate "0" and "1" cell signals, a row of 64 cells already provides 128 (=2x64) 100-bit values. The full 4-Mbit chip measurement would provide over 8 million 100-bit values, which is equivalent to 100 MB of data. Nevertheless, since the series of reference voltages is a monotonic decreasing series of voltages, the resulting output bit stream for a single bit-line only changes value once (from "1" to "0") when the reference voltage crosses the bit-line voltage. This characteristic is exploited by a compressor circuit that compresses the output bit stream and therefore reduces the time required to output the test data.

Figure 70 is a block diagram of the compressor circuit that compresses the bit streams characterizing the charge distribution. It includes a 7-bit counter and a set of 64 7-bit registers connected to the counter and the GIO bus, and an output multiplexer. The counter is reset when a new row of memory cells is selected for measurements. At the same time, the reference voltage is set to its initial value VREFP. Each time the counter increments its count signal CNT, the reference voltage generator changes the reference voltage. The value of signal CNT is thus synchronized with changes in the reference voltage V_{REF} and indicates the reference voltage level corresponding to the current comparison results. During the simultaneous measurements of bit-line voltages, each GIO bus line corresponds to a different bit-line and provides a series of binary values indicating whether the voltage of the bit-line or the reference line is the greater. During a sense amplifier offset measurement, each GIO bus line indicates whether the reference voltage is smaller than the voltage offset needed to trip the corresponding sense amplifier. Each of the 64 7-bit registers receives count signal CNT as a data input signal. Each register in the set latches the value of count signal CNT if the register is enabled when the count signal CNT changes. The GIO bus lines 0 to 63 act as the enable signals for the respective registers 0 to 63, and a value of "1" enables the corresponding register to latch the new count value while a value of "0" disables latching in the corresponding register.

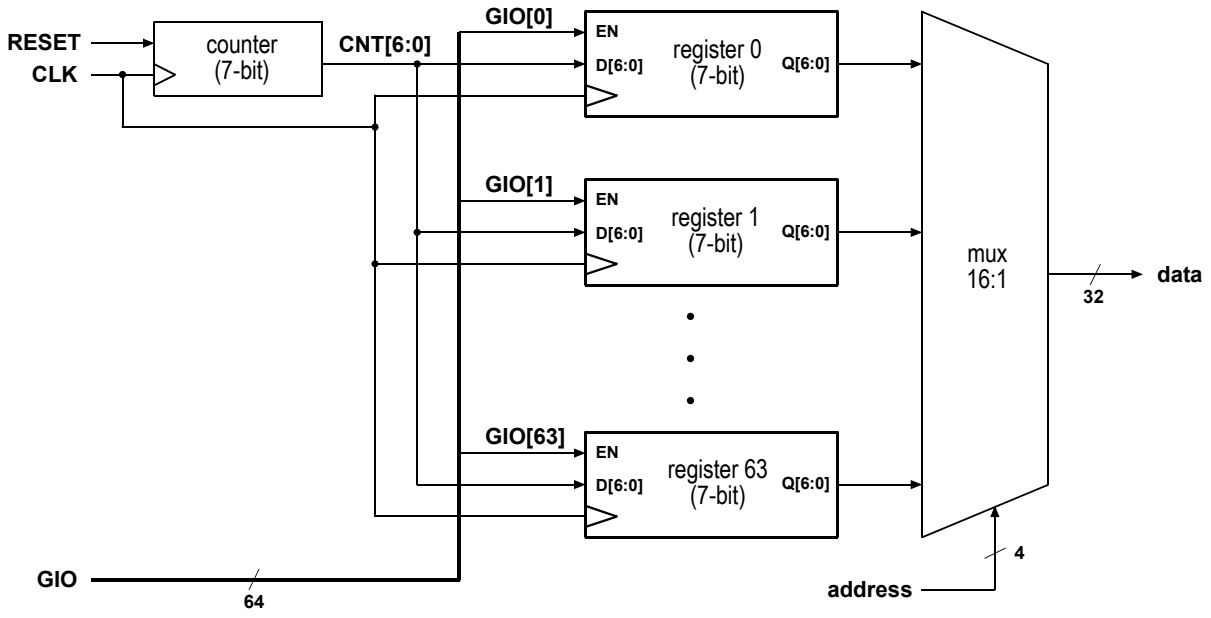


Figure 70: Compression circuit that processes the bit streams from a charge distribution measurement

The count value retained in a register after completion of a series of comparisons will be equal to the count corresponding to the *last* comparison for which the voltage of reference line was *greater* than the voltage of the bit-line signal. Accordingly, for a bit-line voltage measurement, the stored value indicates the approximate bit-line voltage read out of a memory cell, and for a sense amplifier offset measurement, the stored value indicates the offset voltage required to trip the sense amplifier. The compression circuit thus reduces the 100 bits associated with the testing to 7 bits. Consequently, the amount of data to be transferred for a full 4-Mbit chip measurement is reduced from 100 MB to only 7 MB.

At the end of a series of comparisons, each register stores a 7-bit value representing a measured voltage. The multiplexer selects output signals from a subset of registers. For example, four 7-bit measurement values from a group of four registers can be output via a 32-bit data path. Accordingly, the bit-line measurements for 64 cells require 16 output cycles through the multiplexer, instead of 200 output cycles, which would be required to output the uncompressed data.

Technique to determine accuracy of measurement

For the case where the reference voltage consistently steps down, ideal operation of the sense amplifier will provide a bit stream associated with each measurement that

has one binary value (e.g., "1") until the reference voltage falls below the voltage of the bit-line. Thereafter, the bit stream is expected to have the other binary value (e.g., "0"). This ideal bit stream can be represented, without loss of information, by a value indicating when the reference voltage crosses the bit-line voltage. However, noise or variations in the circuitry may cause the output of a sense amplifier to alternate between "0" and "1" when bit and reference line have approximately the same voltage. Variations in sensing performance can be quantified merely by using an alternative precharge scheme for the global I/O bus in addition to the existing one. In this scheme, the GIO bus is precharged only once for the entire series of comparisons that measure the bit-line voltage instead after every comparison. For *precharge once*, the count value retained in a register after completion of a series of comparisons will be equal to the count corresponding to the *first* comparison for which the voltage of reference line was *smaller* than the voltage of the bit-line signal. A bit-line voltage can be measured once with the conventional precharge scheme and a second time with the *precharge once* scheme. A difference in the two measurement values indicates the amount of variation in the performance of sensing operations.

Advantages over existing techniques

The time required for the measurement is significantly reduced because the 100 comparisons require only one readout for each memory cell. For example, if the time required for each comparison is about 5 ns, 100 comparisons can be performed within about 0.5 μs. In contrast, conventional measurements of the cell charge require reading out the memory cell each time the bit-line voltage is compared to a new reference voltage. Consequently, the readout time adds directly to time required for each comparison. Assuming a cycle time of 60 ns will give 6 μs.

The presented measurement technique not only provides fast distribution measurements but can also test data retention. Avoiding repeated readout followed by write-back or refresh operations allows an accurate measurement of the effect aging has on the charge distribution. In particular, the readout charge from each ferroelectric capacitor is the charge after aging, and the measurement determines the charge distribution as aged since data was written into the FeRAM array, not the charge distribution of data as refreshed during the measurement. An exemplary test process

could consist of the steps: (a) write data to the memory cells; (b) bake the FeRAM or otherwise age the stored data; and (c) perform the charge distribution measurement. Since charge is read out of a ferroelectric capacitor only once and measured without refreshing the stored data, the charge distribution measurement accurately reflects the aging of stored data. The aged distribution can be compared to a distribution measured immediately after writing data values in the FeRAM.

C. PLATE-LINE ARCHITECTURE

One issue in the design of a FeRAM is the selection of a plate-line architecture, as described in Chapter III. As one possibility, the *global plate-line architecture* may be employed. However, despite the fact that it is most area efficient, the mediocre plate-line performance due to the very high capacitive load of the plate-line is often unacceptable. The *segmented plate-line architecture* solves this problem by isolating the global plate-line driver from the local plate lines via pass gates. This reduces the capacitive load on the global plate-line driver so that the operational speed of the memory array may be increased. A disadvantage is that the n-channel transistor pass gate requires its gate voltage to be boosted above V_{DD} in order to overcome the transistor threshold voltage and avoid voltage degradation. Since ultra low voltage operation is a primary requirement for future embedded memories, circuits that require over-voltage become increasingly difficult to implement.

Another plate-line architecture is represented in Figure 71. In the *local plate-line driver architecture*, each row of a ferroelectric memory array has a separate plate-line driver. This architecture is suited for low voltage operation and achieves high performance. Its disadvantage is that the large number of drivers reduces the array efficiency of the memory to a point where it becomes unattractive for FeRAM products. This architecture usually finds application in FeRAM process bring-up vehicles that are used to characterize the ferroelectric memory cells and capacitors. The range of different plate-line architectures allows a FeRAM memory designer to select an architecture on a basis of a variety of factors, including available chip real estate and the target supply voltage (e.g., low voltage application). However, each architecture also has disadvantages. What is needed is a plate-line architecture, which enables a low

voltage, area efficient implementation, but which also may be operated at a relatively fast speed.

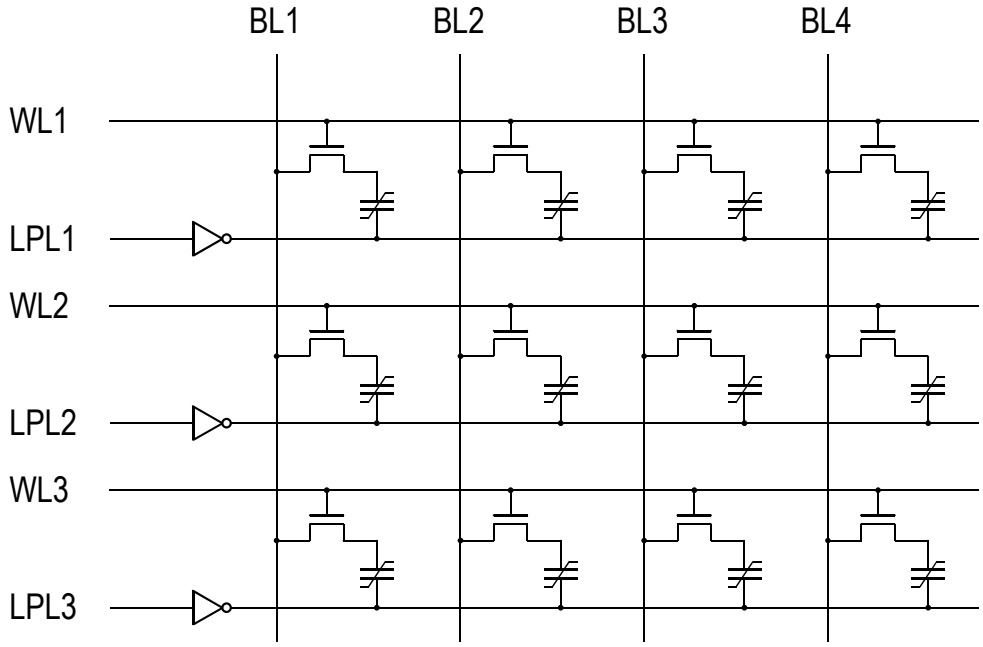


Figure 71: Local plate-line driver architecture

Grouped plate-line architecture

In the following, a new plate-line architecture is proposed. In this architecture, a subset of rows within the same array shares the same local plate-line. These rows form a *plate-line group* (see Figure 72). Each group has its own plate-line driver and decoder. This could be a NAND gate plus inverter. There are multiple groups per array. If, for example, an array would have 512 rows and a plate-line group consisted of 32 rows, then there would be 16 (=512/32) plate-line groups in total. The capacitive load per plate-line is largely reduced because a single plate-line driver only needs to drive one active row and a relatively small number of inactive rows (e.g., 31 rows). Since the load of an inactive row is approximately one percent of the load of the active row, the total capacitive load upon a particular CMOS driver is manageable. In addition to the CMOS driver, each plate-line group may be operatively associated with a separate decoder, while the overall circuitry is maintained within the area-related limitations of

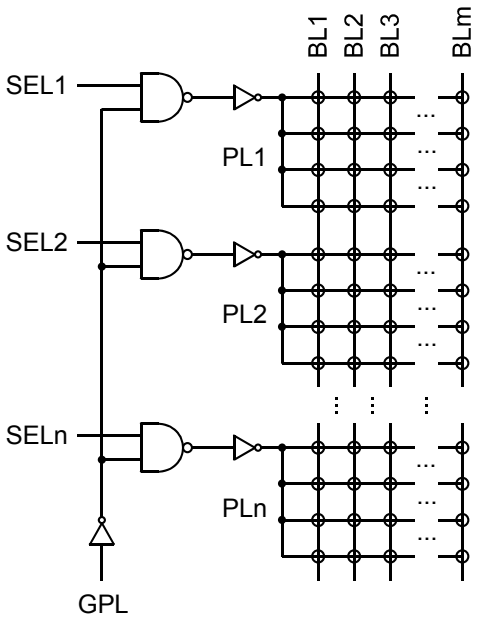


Figure 72: Schematic of grouped plate-line architecture

the integrated circuit chip. The advantage of using CMOS drivers, as compared to pass-gates, is that a boosted gate voltage is not required. Consequently, low voltage application of the architecture is possible. The pitch of the word lines is 0.6 µm in our process. By associating 32 rows with a single driver, sufficient height is available to lay out a practical inverter driver and the small decoder. Consequently, the 32 word lines associated with each plate group driver have a total height of 32 × 0.6 = 19.2 µm. To adequately drive 64 columns requires a simple p-channel driver width within the range of 40 µm to 60 µm, depending on the desired speed for operating the memory

Table 8: Comparison of plate-line architectures (V_{DD} = 1.5 V, 512 cells/bit-line, 64 cells/plate-line)

	Non-Driven	Driven Plate-Line		
Architecture	NDP	Global	Segmented	Grouped
Number of PL drivers	n/a	1	1	16
Load per driver	n/a	1 active and 511 unselected rows	1 active row and 511 pass-gates	1 active and 31 unselected rows
FeCAP voltage	0.75 V	1.5 V	1.5 V	1.5 V
Area efficiency	Best	Good	Average	Average
Speed	Fastest	Slow	Average	Fast
Disturbed cells	None	32768	None	1984
Floating PLs	None	None	511	None
Special requirement	Storage node refresh	No	Word-line boosting	No

array. The plate group driver width, including the NAND gate decoding, can be less than 10 μm. Consequently, the overhead for the plate group drivers used with a column pitch of 1.8 μm is less than 8% (i.e., 10 μm/(1.8 μm × 64)).

A disadvantage of this architecture is that the capacitors of inactive rows experience disturb pulses when the plate-line is pulsed, but this problem is common to all shared plate-line architectures (e.g. global plate-line, common plate-line, etc.). A qualitative comparison of the existing plate-line architectures with the grouped plate-line architecture is shown in Table 8.

D. FATIGUE TESTING

An FeRAM generally includes an array of memory cells where each memory cell contains at least one ferroelectric capacitor. Repeated reading and writing, which flips the polarization state of the ferroelectric capacitor, can fatigue the ferroelectric material and change the properties of the cell (see page 26ff.). The resulting fatigue may eventually lead to a sensing errors.

One way to predict when a particular bit cell may fail is to measure the cell charge before and after performing a series of read/write operations on the cell. A measured change in cell charge can then be extrapolated to the desired minimum required lifetime to project whether the cell will still be operable. If the extrapolation indicates that the cell will fail before reaching the desired minimum life, the FeRAM may have a latent defect and must be replaced. The minimum number of read or write cycles before a failure must be large (e.g., on the order of 10^{15} cycles or more) to provide a memory device with a commercially viable life. The large number of cycles before failure under normal operating conditions can make fatigue testing very time consuming. Extrapolation to 10^{15} read/write cycles, for example, might reasonably require a test to actually perform 10^{12} read/write cycles. A fatigue operation using conventional write operations with a cycle time of 80 ns requires about 2 days of test time to exercise a memory cell 10^{12} times:

10^{12} fatigue cycles: 10^{12} x 80 ns x 2 = 44.4 hours

Note that in order to switch the polarization inside the ferroelectric capacitors, each fatigue cycle requires two clock cycles. In the first clock cycle, a “0” is written to the

capacitors, whereas a "1" is written in the second clock cycle. This two-clock cycle sequence is repeated until the desired number of fatigue cycles is reached. However, performing 10^{12} read/write operations on every memory cell in a reasonably-sized FeRAM (e.g., in a 4-Mbit FeRAM) would literally require days, making such testing impractical for production testing and at least bothersome for a test chip. Extrapolation can be based on a smaller number of read/write cycles to reduce the testing time, but reduction of the number of cycles reduces the accuracy of the test.

Accelerated fatigue operation

During normal RAM operation, the segment address decoder activates only one of the 128 segments at a time. For accelerated fatigue (AF) operation, an override signal generated by the control logic causes the segment address decoder to activate all 128 segments simultaneously. AF operation then rapidly cycles the selected capacitors between the two polarization states by continuously pulsing the plate-line and driving the bit lines to a voltage that is complementary to the voltage of the plate-line. Large drivers connected to each bit-line help to quickly charge/discharge the heavily loaded bit lines.

AF1PG mode

AF1PG mode selects and activates one word-line in every segment and cycles all plate- and bit lines between levels that flip the polarizations states of the selected capacitors in all segments.

Consequently, AF1PG mode is capable of simultaneously exercising 8,192 cell capacitors:

128 segments x 1 row x 64 capacitors = 8,192 capacitors

Since each segment has independent drive circuits for word-, bit-, and plate lines, they do not need to be increased in size to provide the necessary current to mimic write and read operations during AF1PG operation.

AF16PG mode

In contrast to AF1PG mode, AF16PG mode selects and activates a word-line in every plate-line group of every segment. Consequently, each segment has now 16 simultane-

ously activated word lines instead of only one. With multiple word lines active in each segment, AF16PG operation cycles all plate lines and all bit lines to levels that flip the polarizations states of selected capacitors. AF16PG mode simultaneously exercises a large number of capacitors, but does not require larger drive circuits. In particular, each word-line driver still drives a single word-line, and each plate-line driver drives a single plate-line. Accordingly, the associated drive circuits that have the sizes required for normal write and read access are sufficient for AF16PG mode. However, the 16 times larger overall chip power consumption may require the AF16PG mode to operate at a lower frequency than AF1PG mode.

Advantages of AF operation

The accelerated fatigue operation described above has several advantages over prior methods for fatiguing ferroelectric capacitors. Generally, to accurately characterize fatigue effects during development of a FeRAM, the memory cells are exercised for as many cycles as time allows under various conditions. However, a fatigue operation using conventional write operations with a cycle time of 80 ns requires about 2 days of test time to exercise a memory cell 10^{12} times, or about 6 months of test time for 10^{14} cycles. In addition, the number of bit cells that can be simultaneously exercised using a conventional write operation using externally-supplied data generally depends on the data path width and is often less than 128 bit cells. Accordingly, fatigue testing of a significant portion of an average size memory using normal accesses (e.g., write operations) requires too much time to be practical. In contrast, AF operation activates 128 segments and exercises a selectable subset of rows within each segment. AF1PG operation exercises one row within each segment and allows simultaneous exercise of 8,192 memory cells. AF16PG operation exercises 16 rows within each segment and allows simultaneous exercise of 131,072 (i.e., 128 segments x 16 rows x 64) capacitors. The accelerated fatigue operations still need $2x10^{12}$ clock cycles to exercise the capacitors 10^{12} times. However, the exercise sequence is much simpler than in a normal write operation (see Figure 73). Since the sequence to be written is already known ("01010…"), the accelerated fatigue operation does not require data to be brought in from outside the array. The accelerated fatigue operation can thus be a closed internal operation, which eliminates much delay.

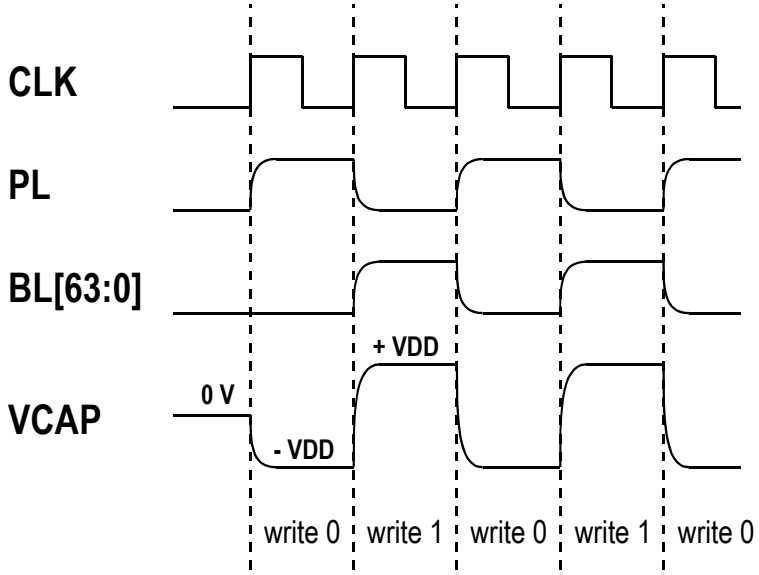


Figure 73: Timing diagram for selected signals during AF1PG mode

Another advantage is that the word lines do not have to be pulsed and can be held at V_{PP} because the same rows are repeatedly accessed. The clock period for the accelerated fatigue operations may thus be much smaller than the clock period for a normal memory access. A clock period of 10 ns, for example, allows AF1PG mode to exercise 8,192 memory cells 10^{12} times in only 5½ hours or 10^{14} times in 23 days:

10^{12} fatigue cycles: 10^{12} x 10 ns x 2 = 5.5 hours
10^{14} fatigue cycles: 10^{14} x 10 ns x 2 = 23 days

The accelerated fatigue operation significantly reduces the time required to test the effects of fatigue on a ferroelectric memory. The shorter time makes fatigue testing more practical during development and/or production of FeRAM.

E. Design Verification

Spice simulation of array and adjacent periphery

In the beginning, a very simple circuit that consisted of a single ferroelectric capacitor connected via an n-channel transistor to a bit-line was used to run initial bit cell simulations. The bit-line capacitance was modeled by a single capacitor and all signal driv-

ers for word-, plate-, and bit-line were modeled by ideal voltage sources. These simulations helped to determine the maximum number of cells per bit-line for a minimum required signal voltage and a given capacitor size. They also helped with the initial optimization process of the bit cell design.

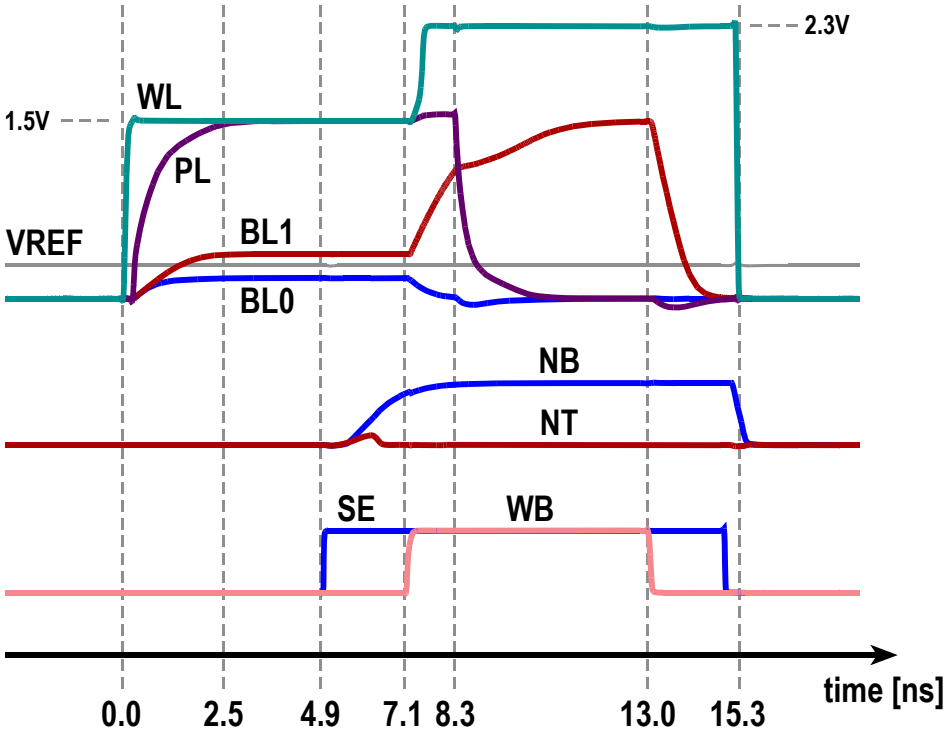


Figure 74: Simulated RAME read cycle

This circuit has been continuously extended during the design phase and provided valuable feedback for many important design decisions. The ideal signal drivers have been replaced with CMOS drivers and multiple bit cells have been connected to the same plate-line in order to account for the load of the plate-line drivers. Drivers and decoders have been added. A sense amplifier has been designed and added to the circuit. Finally, the simulation deck consisted of a full memory array including all required peripheral circuitry. It allowed both a functional and a timing verification of the memory array. However, at the time when more features (= transistors) were added to the chip, the SPICE simulation was too slow to thoroughly verify all the different modes of operation in a reasonable amount of time. At this stage of the design, only the use of a newer hierarchical circuit simulator such as HSIM™ for example would have allowed running a full-chip analog simulation within an acceptable time

frame. Unfortunately, a hierarchical simulator was not available at this time. This fact motivated the decision to implement a digital full-chip Verilog model in parallel for further verification tasks.

Full-chip Verilog simulation

The full-chip Verilog model included the full netlist of the control logic, charge distribution circuit, row address decoder, and additional I/O circuitry; a behavioral model of the memory array was also included. The memory array was modeled as a self-contained module that expects the correct sequence of input signals to generate the correct output signals. This significantly reduced the run time for full-chip simulations, which in turn allowed thorough verification of each of the 6 different modes of chip operation.

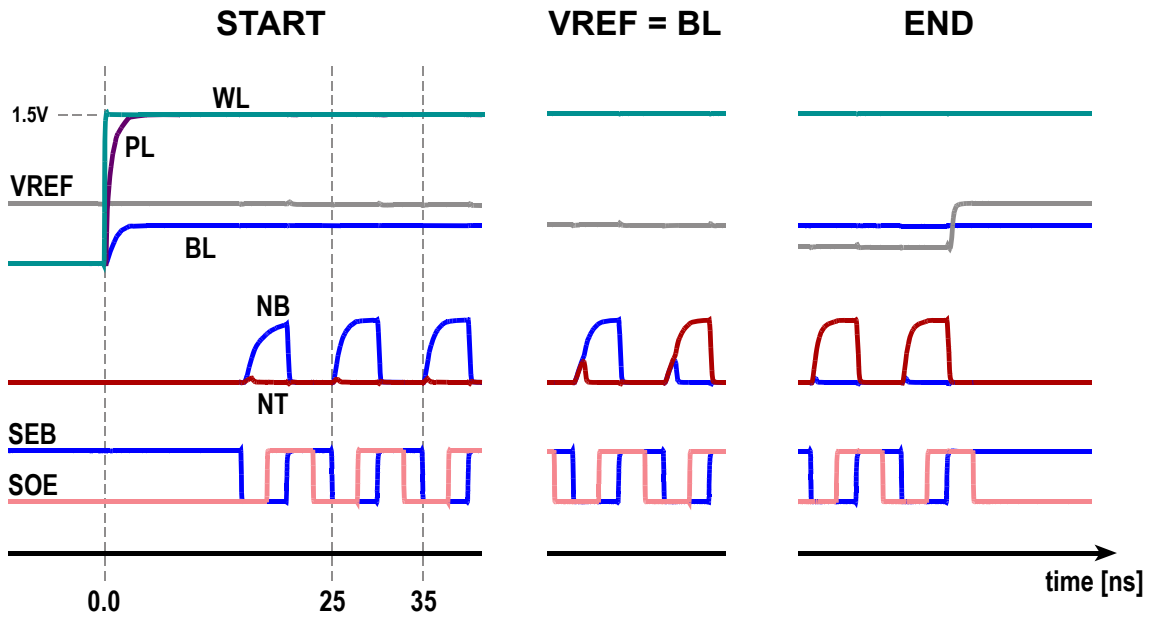


Figure 75: Simulated waveforms of charge distribution measurement obtained from mixed-mode simulation

Full-chip mixed-mode simulation

To further improve the design verification possibilities, a mixed-mode simulation deck was composed out of the Verilog and SPICE simulation decks. For this, the chip was partitioned into two parts: a digital and an analog part. The analog part includes most circuits of the memory array except a few decoders. The digital part includes all the

remaining peripheral circuits. The run time for full-chip simulation significantly increased, but was still acceptable. It was no replacement for the full-chip Verilog simulation, but it filled the need to be able to look at internal memory array signals while simulating complex operations, such as a charge distribution measurement for example (see Figure 75).

VIII. TESTING AND CHARACTERIZATION OF THE 4-MBIT TEST CHIP

The 0.13-µm copper interconnect FeRAM test chip was intended to be an FeRAM process bring-up vehicle. It has built-in functions like sense amplifier characterization, bitline capacitance measurement, accelerated fatigue operation, and charge distribution measurement. In this chapter, the test goals and the test plan are described, and some of the first test results are presented as well.

A. TEST GOALS AND STRATEGY

Two versions of the chip have been manufactured: one with and one without ferroelectric capacitors. The one without ferroelectric capacitors has shorts instead of capacitors. This version is especially suited to verify the chip design without having to deal with the influence of the ferroelectric process on the underlying CMOS process. Since the chip makes use of many new circuits, a primary test goal is to verify the basic functionality of as many aspects as possible. This goal is achieved by providing a series of tests that systematically test the functionality of the logic of the chip. By organizing the tests in this way, any problems should be easier to diagnose when they are first encountered. In addition, failure-binning results should then be more meaningful. For the functional verification, nominal timing is used because timing characterization is not a priority. Later, extended characterization tests that may only be applied to a small sample of the test chips can be executed. The results of such tests are useful to determine the accuracy of measurements made later, as well as to get some direct measurements of key parameters of the 0.13-µm process.

Once the basic functionality of the test chip has been validated, the suite of tests that characterize the process in view of FeRAM applications can be run. The primary goal of these tests is to gather as much data about the FeRAM bit cell as possible under various operating and stress conditions. Of great interest are, for example, charge distribution measurements of virgin as well as fatigued chips.

B. TEST PLAN

Functional verification

As described earlier, the main goal of these tests is to verify the chip's design using the chip version without ferroelectric capacitors. In addition, they can be used to determine which die with ferroelectric capacitors qualify for further characterization. The basic test suite includes an open and shorts test, a global I/O (GIO) bus test, and a gross sense amplifier functionality test. Since the GIO bus is the main internal data bus that is shared by all segments and the charge distribution circuit, its functional verification is of great importance. The GIO bus test therefore checks for shorts between all possible pairs of lines, stuck-at-faults affecting any individual lines, and stuck-opens affecting all driver transistors along the signal path. It specifically verifies that the GIO bus lines can hold "0"s or "1"s dynamically for at least a few microseconds, that the GIO bus lines can be pre-charged to "1"s, and that a "1" can be walked through a field of "0"s and vice versa. The gross sense amplifier functionality test verifies that the decoders address and enable the sense amplifiers properly and that the sense amplifiers can switch high and low and then force lines in the GIO bus low.

The extended test suite tries to evaluate and characterize the key circuits. It includes a sense amplifier offset voltage measurement, a bit-line capacitance measurement, a write-"1" voltage measurement and an overall circuit performance evaluation. The sense amplifier offset voltage measurement determines the individual offset voltages due to the imbalances of each sense amplifier. This information may be used to improve the accuracy of all other measurements that are affected by the sense amplifiers' offset voltage, like for example bit-line capacitance measurements, or bit cell charge measurements. The performance evaluation is used to determine the maximum operating frequency of the peripheral circuits such as pre-charge, sense amplifier, data I/O paths, etc.

The deliverables of all these tests are a wafer map that shows passing dies and failing bins, a distribution of sense amplifier offset voltages, a distribution of bit-line capacitance, Shmoo plots of V_{DD} versus frequency, setup and hold time margins for address, data, and control signals in with respect to the clock.

Bit cell characterization

The goal of this set of tests is to measure the charge that is delivered by each bit cell under various operating conditions. For example, the retention test evaluates the data retention capability of the bit cell. It writes a known data pattern to the bit cells, bakes the wafer at high temperature, and reads the bit cells out using the chip's QD mode (see Chapter VII.B above) after several days or weeks. The fatigue test exercises the memory cells up to 10^{12} or 10^{14} cycles and measures the cell charge at regular intervals during the test. All tests make either use of the chip's QD mode or a series of regular read operations combined with external stepping of the reference voltage in order to determine the charge that is delivered by the bit cells. For retention and imprint measurement, the use of QD mode is required since the write-back phase of the read operations would affect the measurement.

C. TEST RESULTS

C.1 Dependency of Write-"1" Voltage on V_{PP}

The write-"1" voltage measurement tries to measure the effective voltage that is applied to the ferroelectric capacitor when writing a "1". To write a "1" into the capacitor, the bit-line must be driven to V_{DD} and the plate-line held at V_{SS}. If the word-line voltage V_{PP} is too low, a voltage drop across the word-line transistor degrades the V_{DD} level resulting in a write-"1" voltage that is lower than V_{DD}. The optimal V_{PP} is just large enough to achieve zero degradation of V_{DD}, but not larger in order to maximize the transistors gate-oxide lifetime. The ability to determine the required V_{PP} by measurement helps to guarantee a write-"1" voltage of full V_{DD} during later bit cell characterization and memory operation. A single measurement consists of: (1) writing "1"s to the dielectric capacitor array; (2) dumping their charge onto the bit lines; and (3) measuring the resulting bit-line voltage. The above measurement was performed for different values of V_{PP} (Figure 76). The optimal V_{PP} is located where the bit-line voltage starts to saturate – in this case at about 2.3 volts.

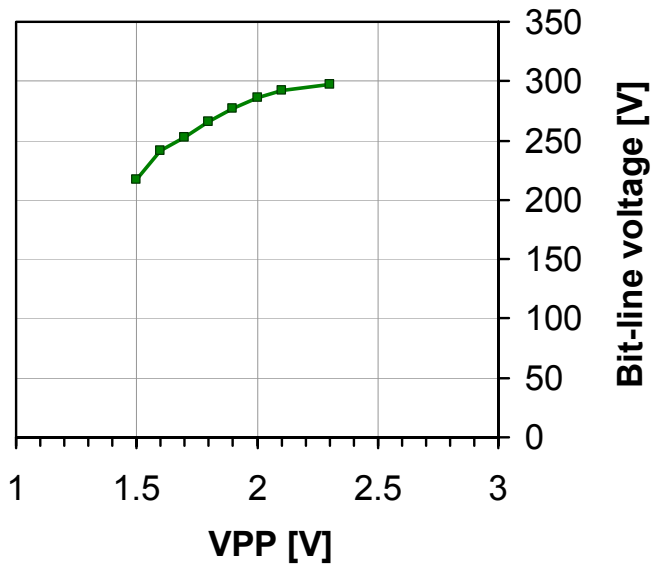


Figure 76: Measured bit-line voltage versus V_{PP}

C.2 Sense Amplifier Offset Voltage

Figure 77 shows one of the first measured offset voltage distributions. It was obtained from a chip without ferroelectric capacitors. Note that the center of the distribution is located at about 108 mV and not at 0 volt.

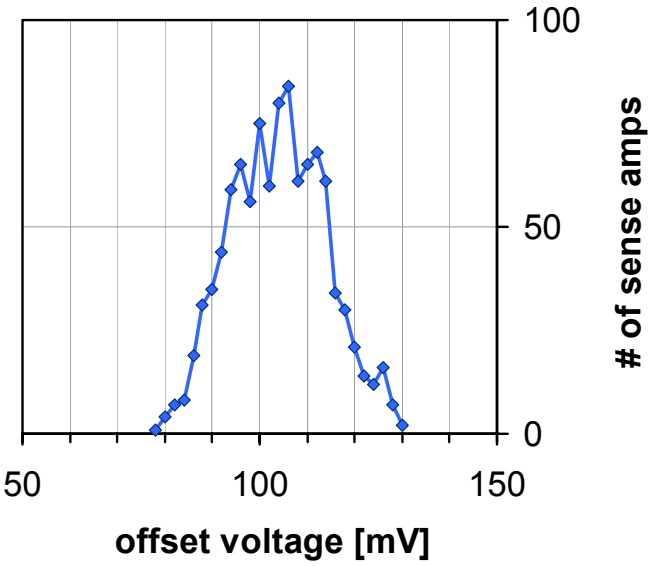


Figure 77: Measured sense amplifier offset voltage distribution

This was unexpected since parasitic extraction should have revealed significant imbalances in the layout. Therefore, the origin of the large offset was unclear. On the other hand, the offset did not represent a problem or even degrade the functionality of the chip and was therefore not further investigated at first. Instead, the measured offset voltage was subtracted from all affected measurements. After more chips had been characterized, it turned out that the vast majority had an offset voltage distribution that was centered at around 110 mV. The parasitic extraction for the sense amplifier was repeated and revealed a small capacitive imbalance of less than 180 aF that is responsible for the 110 mV offset (see Figure 78 below).

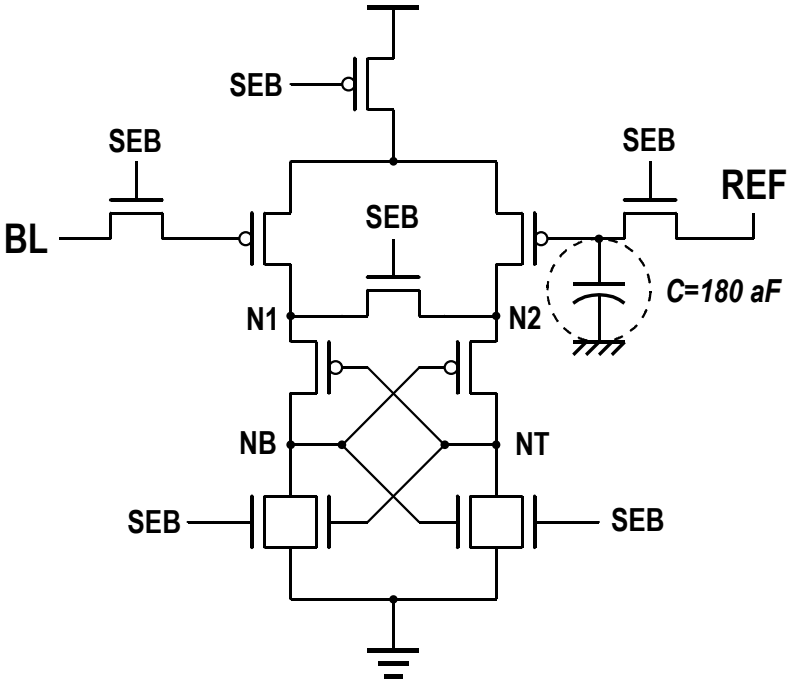


Figure 78: Imbalance responsible for 110 mV offset

C.3 Performance of Measurement Circuitry

For performance characterization of the measurement circuitry, the chip is operated in SACAL mode. The sense amplifier offset voltage measurement (SACAL) mode is more suitable for this purpose than the bit cell charge distribution measurement (QD mode). This is because both modes exercise the same circuitry, but SACAL mode is not subject to bit cell charge variations since the bit lines are held at ground.

The time required for a single charge distribution measurement (1 row) is determined by the time that each comparison between bit-line and the current reference voltage takes. Each comparison consists of: (1) activating the sense amplifiers; (2) driving the comparison result onto the GIO bus; (3) latching the count from the QD counter into the enabled QD registers; (4) stepping the reference voltage and incrementing the QD counter. The external signal SACTRL is used in the SACAL, BLMSR, SNMSR, and QD modes as the clock signal for the sense amplifiers. One cycle of SACTRL corresponds to one comparison. Consequently, the time allowed for each comparison is equivalent to the cycle time of SACTRL. In order to determine the minimum time required for a single comparison, the cycle time of SACTRL was varied between 3 and 5 ns for different operating voltages. The results are summarized in the Shmoo plot below (Figure 79).

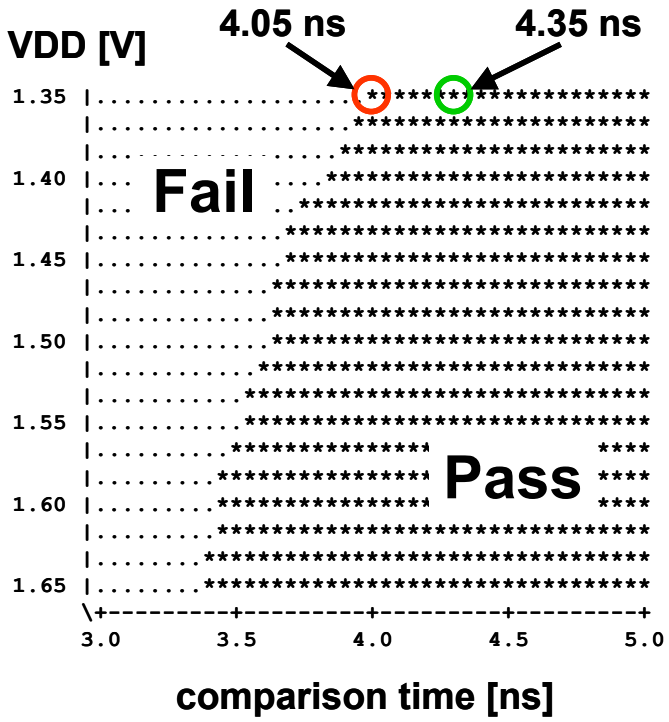


Figure 79: Shmoo plot of comparison time versus operating voltage

In Figure 79, "fail" represents the case where the measurement did not produce meaningful results at all while "pass" indicates a successful measurement. However, "pass" does not imply that the measurement results are accurate. It is important how far the

"pass" is from the border between pass and fail. To illustrate this, the measured distribution for a comparison time of 4.05 ns and 4.35 ns at V_{DD}=1.35V are compared against each other (Figure 80). Obviously, the distribution that was obtained for 4.05 ns is slightly shifted to the left compared to one obtained for 4.35 ns. This makes sense since 4.05 ns is right at the border between pass and fail in Figure 79, while 4.35 ns is a little bit further apart. Therefore, the data obtained at 4.35 ns is likely to be more accurate. For reference, the distribution that has been obtained for a very long comparison time of 20 ns is shown, too. It was found that a comparison time of about 5 ns is sufficient, because it produced almost identical results as for 20 ns.

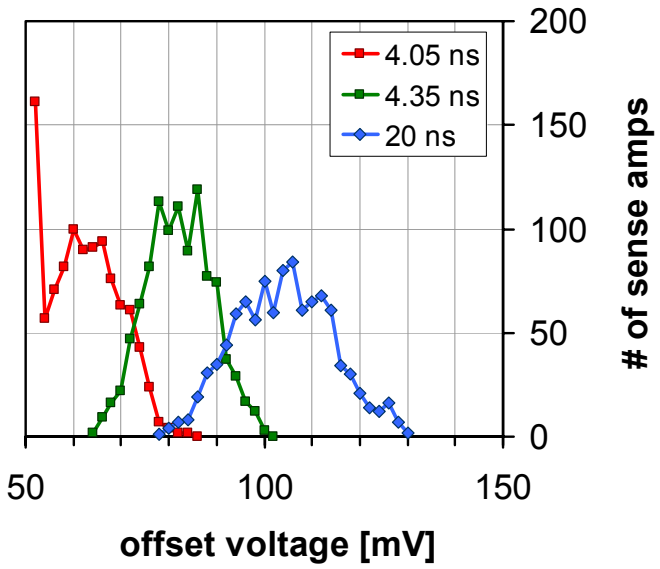


Figure 80: Offset voltage distribution for different cycle times

The comparison time of 5 ns allows one to obtain the cell charge distribution for the full 4-Mbit chip in about 45 ms.

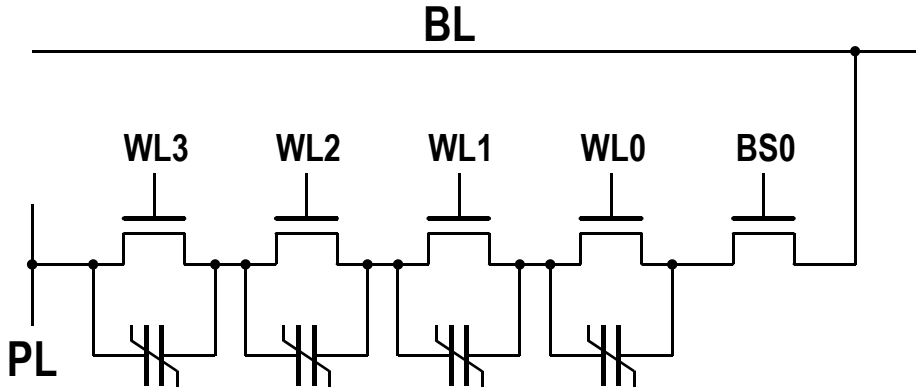
C.4 Conclusion

The presented results show that despite some minor problems, the test chip achieves the goals that were set in the beginning. In particular, the sense amplifier offset measurement is very powerful as it allows variations in the sense amplifiers to be subtracted from the bit cell measurements as well as characterizing the sense amplifiers themselves. This capability may be attractive for SRAM, too. The speed and accuracy

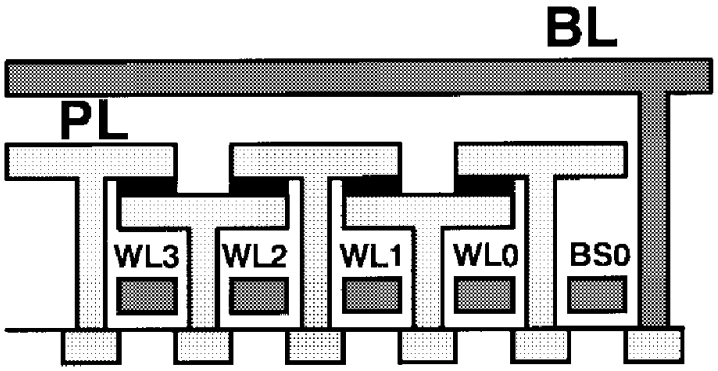
of the charge distribution measurement exceeded expectations and allows obtaining the bit cell charge distribution of the full chip in only 45 ms.

IX. CHAIN FERROELECTRIC MEMORY

Chain ferroelectric random access memory (CFRAM) has been proposed as a new architecture for ferroelectric memories that promises many advantages over the conventional 1T1C architecture [2,16,30,32-35]. In contrast to the 1T1C cell, the transistor and capacitor of the CFRAM cell are connected in parallel instead of in series.



(a) circuit diagram



(b) cross section

Figure 81: Chain FeRAM memory cell block concept [16]

Multiple cells are then lined up in a chain that is connected via a block-select transistor to the bit-line (Figure 81). During operation, all unselected capacitors in a chain are short-circuited through the parallel-connected cell transistors. This is possible since the ferroelectric capacitor retains its remanent polarization when its plates are shorted. The turned-on transistors provide a path from the plate-line to the bottom

plate of the selected capacitor and from the bit-line to the capacitor's top plate, thus permitting completely random cell access.

Reduced cell size seems to be a major advantage of CFRAM over conventional FeRAMs. However, the serial connection of cells produces several new issues that have to be taken into account. This chapter investigates the most important issues such as readout voltage shift, readout delay, and increased parasitic capacitance. In addition, new ideas that try to overcome these issues are presented. At the end, CFRAM is compared to FeRAM with respect to area, performance, reliability, and lifetime.

A. Important Issues for Chain FeRAM

A.1 Readout Delay and Word-line Voltage

In ferroelectric memories, it is crucial to have a fast moving plate-line to achieve good access/cycle times. It is therefore common practice to use very large drivers and a metal plate-line. However, in CFRAM the bottom electrode (BE) of the selected capacitor is not directly connected to the plate-line but is instead connected through a resistive path formed by the transistors in-between.

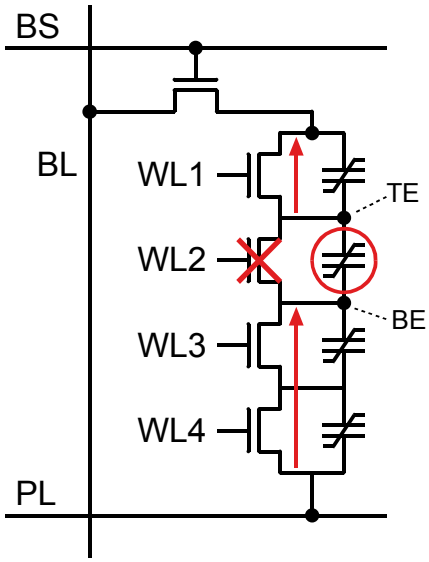


Figure 82: Signal delays in CFRAM chain (cell 2 is selected)

The same applies to the top electrode (TE) of the selected capacitor (see Figure 82). It is very desirable that the ohmic resistance of these paths be as low as possible to en-

able high-speed and low-noise operation for CFRAM. The on-resistance of minimum-width sub-micron MOS transistors is in the magnitude of several kilo-ohms and is voltage-dependent. Consequently, during readout the BE signal is always delayed with respect to the PL signal and the BL signal is delayed with respect to the TE signal. The second delay has already been reported as the readout delay [30]. The other one will be called the plate-line delay and their sum will be called the total readout delay. In CFRAM, typically 8 or more transistors are connected in series. During operation, most of them are subject to the body effect and consequently have a significantly higher threshold voltage. A rise from 0.6 to 1.1 volts or more can be observed for V_{th} during read operations, as shown in Figure 83. In this example, all threshold voltages V_{th1} to V_{th7} are increased except V_{th0}, which is the threshold voltage of the selected cell's transistor. In contrast to conventional FeRAM, the body effect in CFRAM has a much stronger influence on memory speed and, in fact, imposes a lower bound on the boosted voltage V_{PP} and access times, as presented next.

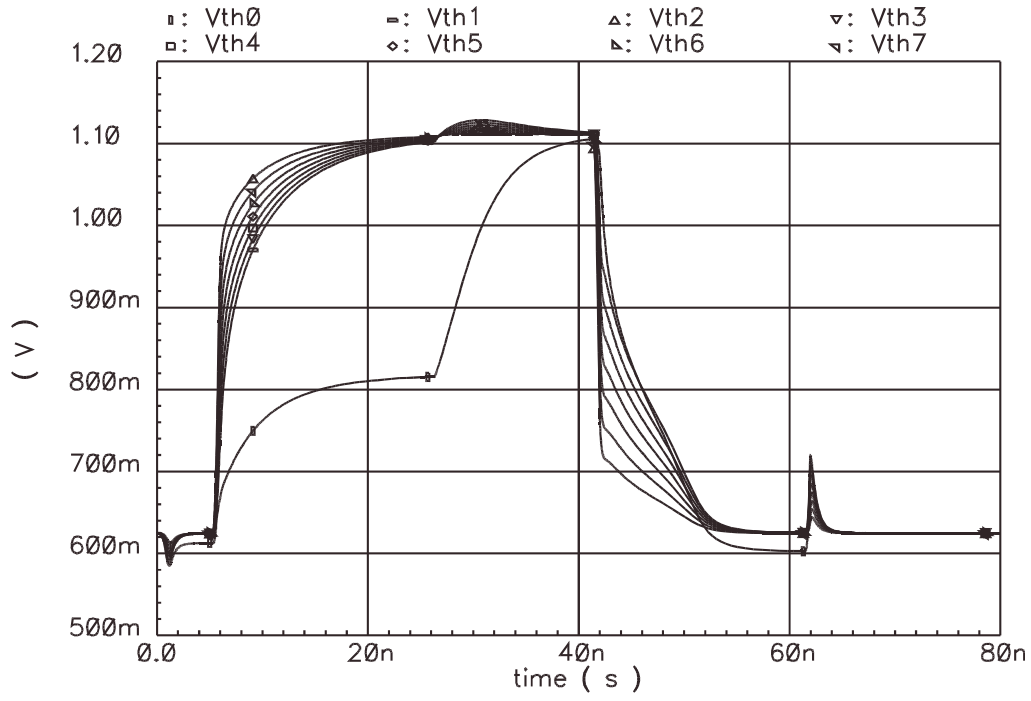


Figure 83: The body effect causes V_{th} to rise during operation. The cell nearest to the bit-line is accessed. ($N = 8$, $V_{DD} = 3.3$ V)

During readout the plate-line delay is much larger than the readout delay because the transistors between BE and the plate-line are strongly affected by the body effect

while the ones between the bit-line and TE are not. The delay reaches a maximum for the cell nearest to the bit-line. The impact of chain length N and word-line voltage V_{PP} on this delay is remarkably strong, as Figure 84 displays. For example, setting N = 32 and V_{PP} = 4.5 V results in a plate-line delay of 69 ns, which is unacceptable. In contrast, N = 8 and V_{PP} = 5.0 V appear to be suitable for fast operation assuming 0.35-µm CMOS and a capacitor size of 1 µm².

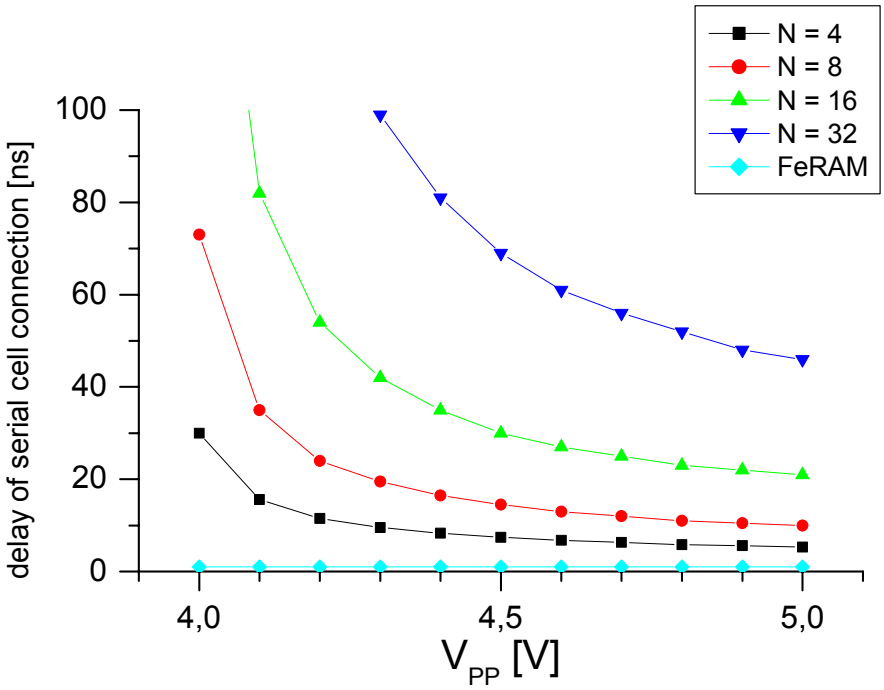


Figure 84: Plate-line delay versus V_{PP} for different chain lengths

A.2 Bit- and Plate-Line Capacitance

Reduced bit-line capacitance

The design of a ferroelectric memory array requires that the ratio of storage to bit-line capacitance being adjusted close to its optimum value, where the signal voltage is maximized while maintaining sufficiently high storage density. For a given cell configuration, this goal is commonly achieved by adjusting the number of cells per bit-line because the parasitic capacitances of each word-line transistor contribute to total bit-line capacitance. For conventional FeRAM, the bit-line capacitance is found to be:

$$C_{BL} = C_{wire} + (rows - 1) \cdot C_{acc}$$

C_{acc} is the parasitic capacitance per word-line transistor. In CFRAM, mainly the block-selecting transistors contribute to bit-line capacitance. To obtain the number of block-selecting transistors that are connected to one bit-line, the number of rows has to be divided by the chain length N. Consequently; bit-line capacitance is greatly reduced for Chain FeRAM. Another difference is that every cell transistor in the active chain, which lies between the bit-line and the selected cell, contributes to the bit-line capacitance. Hence, the bit-line capacitance is different for each cell in a chain and reaches a maximum for the cell farthest away from the bit-line. The effective bit-line capacitance when this cell is accessed can be approximated as:

$$C^*_{BL} = \underbrace{C_{wire} + \left(\frac{rows}{N} - 1\right) \cdot C_{acc}}_{C_{BL}} + \underbrace{C_{ser} \cdot N}_{C_P}$$

C_{ser} is the parasitic capacitance per turned on chain transistor. The value for N that minimizes the bit-line capacitance for a given number of rows is found to be:

$$\Rightarrow N_{OPT} = \sqrt{rows \cdot \frac{C_{acc}}{C_{ser}}}$$

Figure 85 is a plot of C_{BL} versus chain length for different numbers of rows, showing that the minimum value of C^*_{BL} moves to the right with an increasing number of rows. The comparable bit-line capacitance of a conventional FeRAM can be found at N = 1.

Increased plate-line capacitance

Just as the transistors between the bit-line and the selected cell contribute to bit-line capacitance, the transistors between the selected cell and the plate-line contribute to the plate-line capacitance C_{PL}:

$$C_{PL} = C_{wire} + columns \cdot \left(C_S + (N - 1) \cdot C_{ser}\right)$$

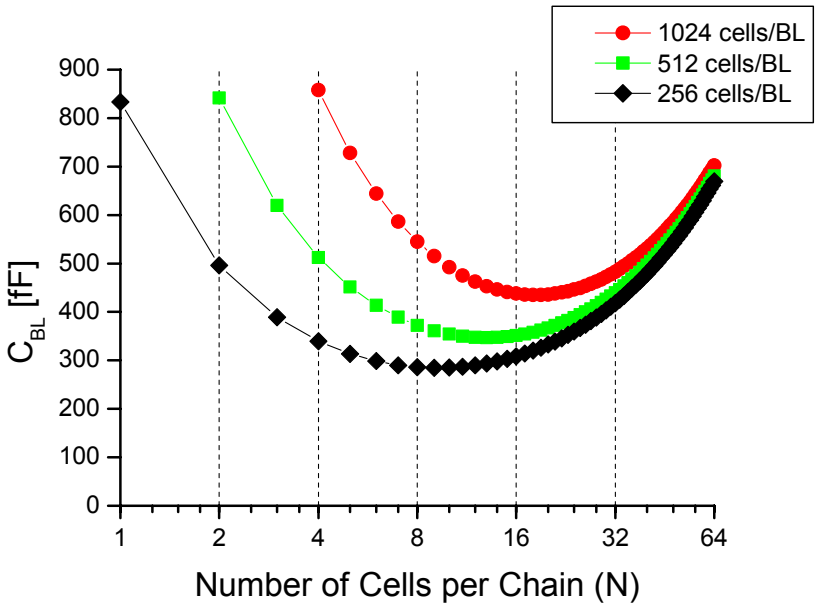


Figure 85: Bit-line capacitance versus chain length N

In present FeRAMs, the capacitor size is very large, typically about $25F^2$, where F is the minimum feature size. Therefore, cell storage contributes a relatively large fraction of the total C_{PL} for conventional FeRAMs:

$$\frac{C_S \cdot columns}{C_{PL}} \approx 97\%$$

On the other hand, in CFRAM, with increasing chain length the parasitic capacitance component occupies an increasingly large part of the total C_{PL}, as shown in Figure 86. This result implies that new techniques are required to mitigate this effect.

Early plate-line scheme

A possible way to weaken the impact of the increased plate-line capacitance on the plate-line rise and fall times is to allow the plate-line to be raised right after the falling edge of WL and before the rising edge of BS, the block select signal, and not simultaneously with BS. In this way, the parasitic capacitances are already partially charged when BS goes high and the charging current is more equally distributed over time. This will be called the early plate-line scheme.

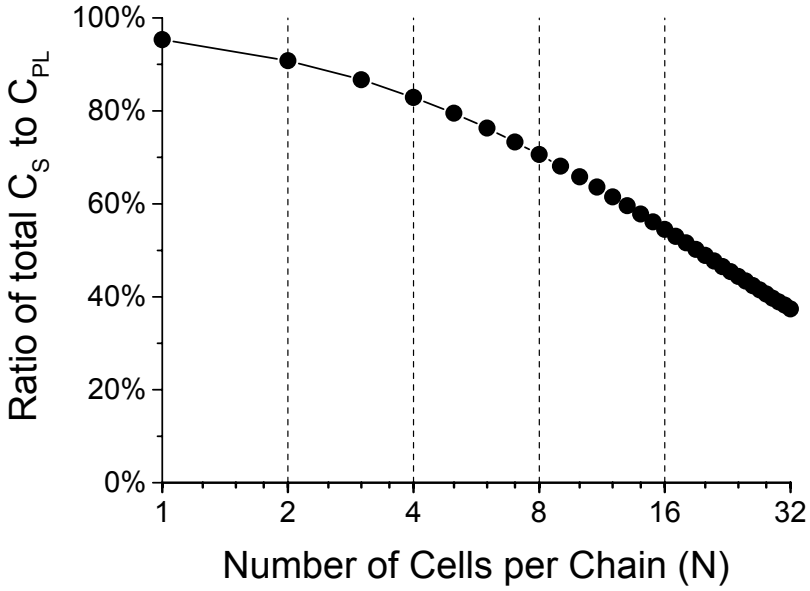


Figure 86: Ratio of total storage capacitance to plate-line capacitance versus chain length

Figure 87 presents a comparison of the new scheme with a conventional scheme concerning bit-line and cell plate-line rise times. The cell plate-line is the net connected to the bottom electrode of the selected ferroelectric capacitor. The bit-line nearest cell is accessed (worst case). This scheme achieves 6 ns faster readout of the cell signal.

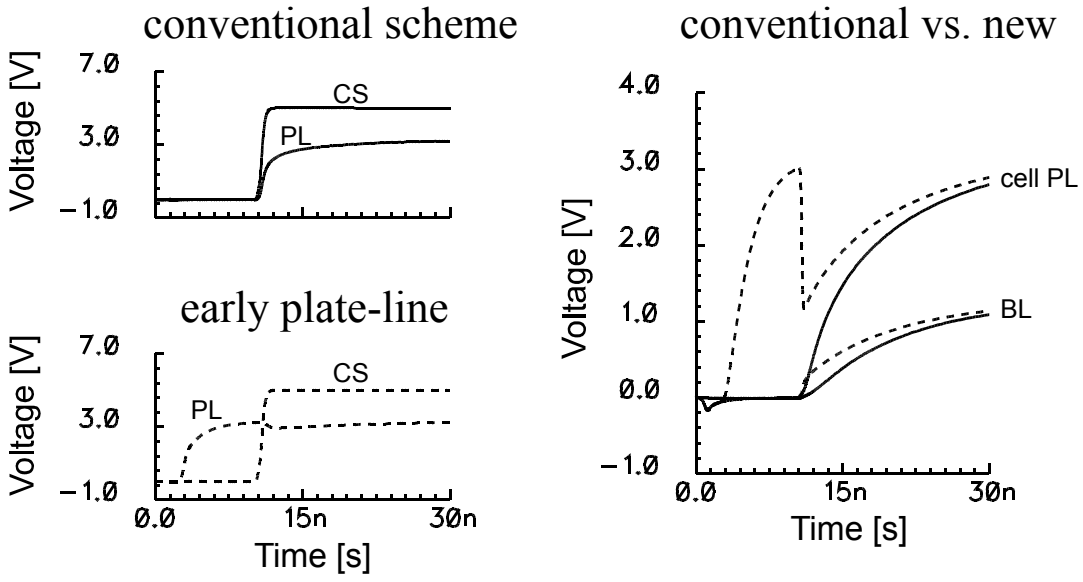


Figure 87: Early plate-line compared to conventional plate-line scheme. The nearest cell to the bit-line is accessed (worst case). (N=16)

In a reported non-driven half-V_{DD} plate-line scheme [17], although the plate-line delay is eliminated, only half-V_{DD} bias to cell capacitor requires relatively high-V_{DD} operation. Moreover, in a reported driven plate-line scheme [30], although the low voltage operation is suitable thanks to full V_{DD} bias to capacitor, a plate-line delay occurs. On the other hand, the proposed early plate-line scheme satisfies both a full V_{DD} bias to capacitor and the elimination of plate-line delay.

A.3 Readout Voltage Shift

In CFRAM, the bit-line voltage after readout depends on cell position because the effective bit-line capacitance is different for each cell (see above). Starting from the cell nearest to the bit-line, the bit-line voltage starts to shift for cells further away from the bit-line. This variation is called readout voltage shift (Figure 88).

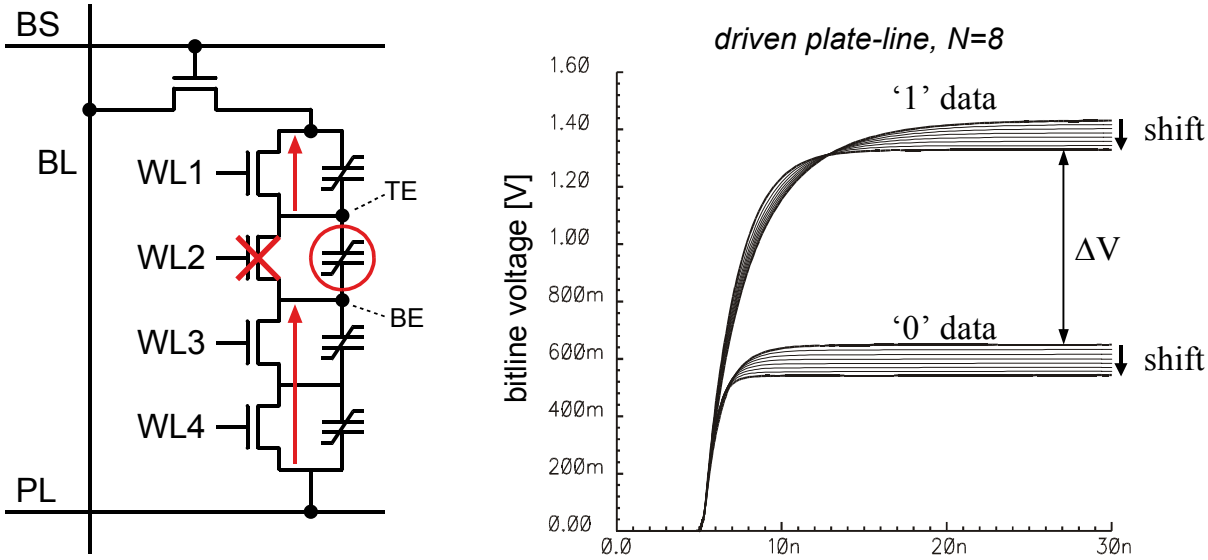


Figure 88: Readout voltage shift observed for CFRAM

Analysis

The effective bit-line capacitance C^*_{BL} is composed of C_P and C_{BL}. The first capacitance represents the parasitic capacitances of the chain transistors (unique to CFRAM) while C_{BL} includes all the other capacitances (as in conventional FeRAM). The value of C_P depends on a cell's position within the chain. Figure 89a presents a simplified circuit to model the influence of C_P on bit-line voltage V_{BL} for a non-driven plate-line scheme. C_S is the ferroelectric storage capacitor.

Prior to read access, bit-line is at ground and the internal storage node SN is at $V_{DD}/2$ due to word-line WL0 being at high and block-select BS at a low level. Thus, C_P is initially pre-charged to $V_{DD}/2$. After WL0 and BS change logic levels, nodes SN and BL are shorted and nodes SN and PL are only connected through C_S. C_P 's charge is redistributed and as node SN falls towards ground, additional charge Q_S leaves C_S and redistributes on C_P and C_{BL}. Finally, nodes SN and BL are both at V_{BL}, which can be calculated under the assumption that the charge of nodes SN and BL is only redistributed:

$$Q_{SN,initial} + Q_{BL,initial} = Q_{SN=BL,final}$$

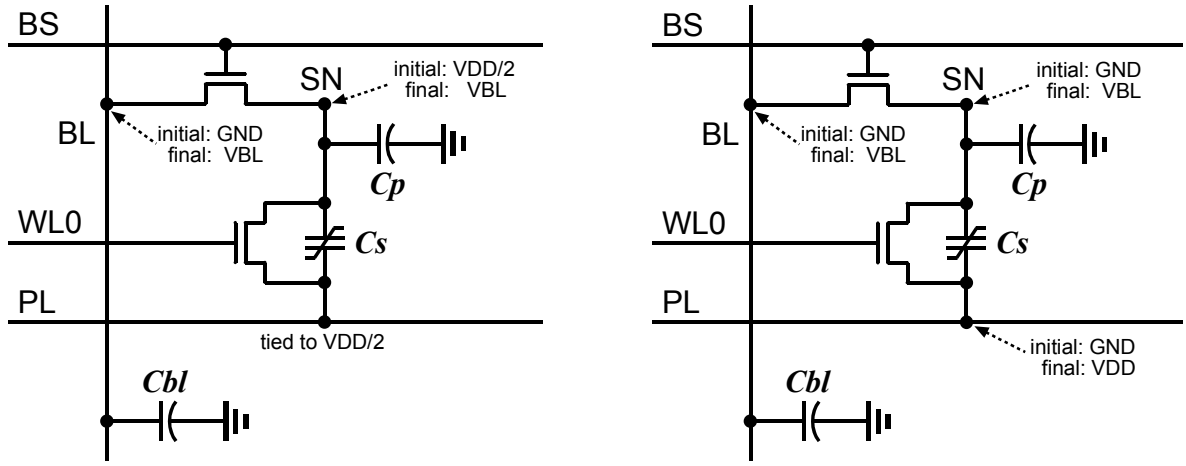
$$\tfrac{1}{2}V_{DD} \cdot C_P + 0 = V_{BL} \cdot \left(C_{BL} + C_P\right) - Q_S$$

$$\Rightarrow V_{BL} = \frac{\tfrac{1}{2}V_{DD} \cdot C_P + Q_S}{C_{BL} + C_P}$$

Since $V_{DD}/2$ is larger than V_{BL}, a larger C_P will increase V_{BL} leading to a positive shift of readout signal voltages, as:

$$\Leftrightarrow V_{BL} \cdot C_{BL} - Q_S = \underbrace{\left(\tfrac{1}{2}V_{DD} - V_{BL}\right)}_{>0} \cdot C_P$$

A possible solution to reduce the loss of signal voltage is to generate the reference voltage by dummy cells that are also arranged in a chain [30]. The reference voltage then shifts in a manner similar to the readout voltage and stays well in the center of V_{BL0} and V_{BL1} (see Figure 90). However, the storage cell to bit-line capacitance ratio still depends on the cell position and may be difficult to optimize. Furthermore, in contrast to earlier assumptions, a voltage shift also occurs for a driven plate-line scheme.



(a) non-driven *(b) driven plate-line scheme*

Figure 89: Simplified circuit models for read operation

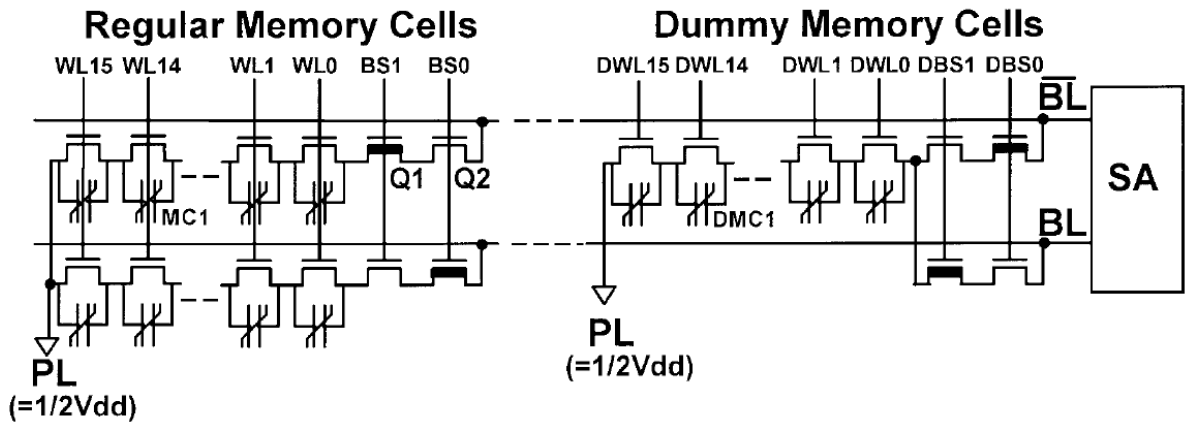


Figure 90: Proposed chain arrangement of reference capacitors to compensate for readout voltage shift [30]

In Figure 89b, the same circuit model is used again for a driven plate-line scheme. The only difference is that this time the storage node and bit-line both start at ground because the plate-line starts at ground and rises to V_{DD}. Hence, C_P is initially discharged.

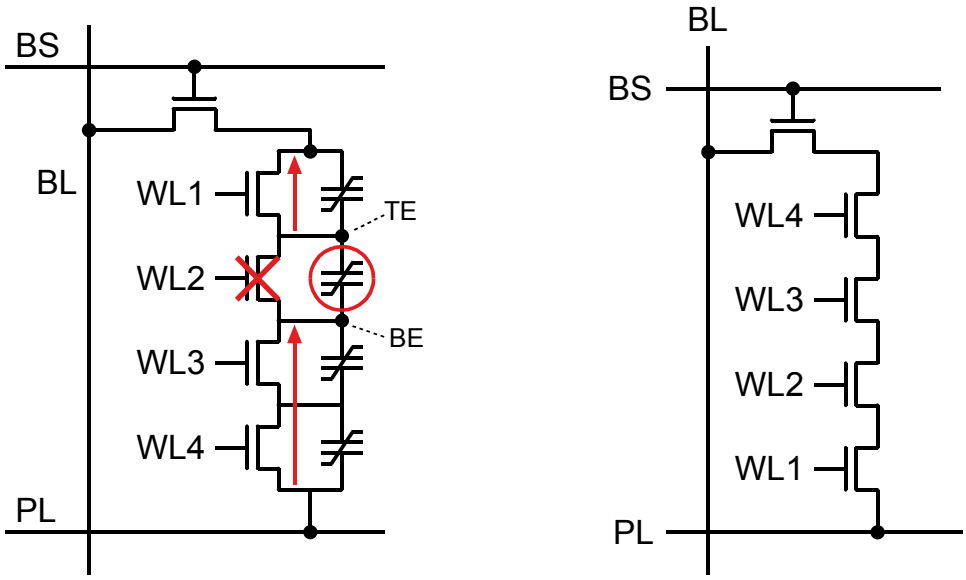
$$0 = V_{BL} \cdot \left(C_{BL} + C_P \right) - Q_S$$

$$\Rightarrow V_{BL} = \frac{Q_S}{C_{BL} + C_P}$$

Now a larger C_P will decrease V_{BL} leading to a negative voltage shift. Since the read-out voltage shift reduces the signal margin, a technique to compensate for this effect is highly desirable.

Compensation technique

The readout voltage shift is caused by variation in bit-line capacitance. In order to make the bit-line capacitance independent of cell position, a new compensation circuit is introduced. The circuit consists of a replica chain without ferroelectric capacitors in which the order of the word lines is reversed (see Figure 91). The new circuit adds a complementary amount of capacitance to the bit-line, depending on which cell position within the chain is accessed. Its functionality is described next.



(a) memory cell chain *(b) replica chain*

Figure 91: Technique to make bit-line capacitance independent of cell position

For example, when cell 2 is accessed, transistor 2 has to be turned off by applying a low voltage to WL2. Then C_{BL} is only increased by the channel capacitance of transistor 1:

$$C_{BL}^{*} = C_{BL} + C_{ser}$$

With the replica chain, C_{BL} is also increased by transistors 3 and 4 of the replica chain:

$$C_{BL}^{*} = C_{BL} + 3 \cdot C_{ser}$$

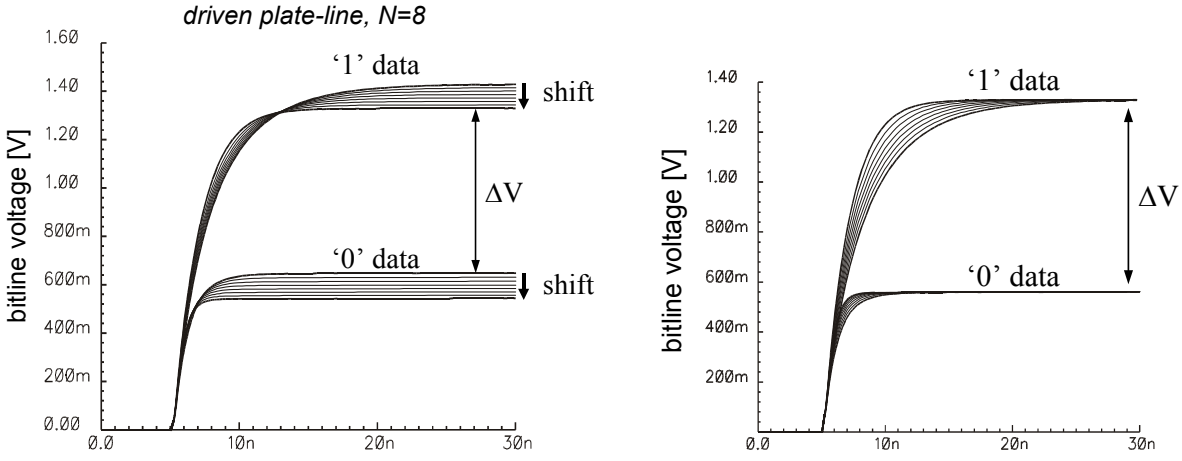
When cell 3 is selected, transistor 3 is turned off and C_{BL} is increased by the channel capacitance of transistors 1 and 2:

$$C_{BL}^{*} = C_{BL} + 2 \cdot C_{ser}$$

With the replica chain, C_{BL} is also increased by transistor 4 of the replica chain:

$$C_{BL}^{*} = C_{BL} + 3 \cdot C_{ser}$$

This results in the same effective bit-line capacitance for cell 3 as for cell 2. Due to the complementary contribution of capacitance, C^{*}_{BL} becomes independent of cell position. Consequently, the readout voltage shift is eliminated. Figure 92 presents a simulation of a readout operation with and without the replica chain connected to the bit-line. This technique is applicable for driven and non-driven plate-line architectures.



(a) without replica chain *(b) with replica chain*

Figure 92: Transient simulation of readout operation

A.4 Chain Length

The area required for the memory array is mainly controlled by two parameters: the chain length N and width W of the cell transistors. By increasing the chain length

from 2 to 16, a substantial reduction in average cell size can be observed as the layout becomes more and more efficient. A further increase of chain length is not worthwhile because the cell size is already very close to the smallest possible value. Moreover, the more cells that are connected in series, the more transistors there are between the selected cell and the bit-/plate-line. Consequently, C^*_{BL} (for $N > N_{OPT}$), C_{PL} and path resistance grow with chain length resulting in larger RC time constants. Especially the bit-line nearest cell experiences the largest penalty (worst case) in readout times. Wider cell transistors could reduce R_{ON} but also would increase C^*_{BL} and C_{PL}. Furthermore, the variation and amplitude of the readout voltage shift increases with chain length. Fortunately, noise across unselected cells during operation is reduced. Because of its strong (negative) impact on performance, it is advisable to dimension N as small as possible. Acceptable performance is achieved for 4, 8, or 16 cells per chain.

B. Comparison to Standard FeRAM and Conclusion

Area

Chain FeRAM has the potential to be superior to conventional FeRAM with respect to average cell size because of its innovative cell configuration. It reduces the average number of signal lines per row from two to almost one (WL - similar to DRAM) and, therefore, removes the connectivity drawback required for bipolar polarization of ferroelectric materials compared to unipolar paraelectric polarization. In addition, fewer transistors are connected to each bit-line in a CFRAM because all cells in a chain share the same block-selecting transistor. This reduces the total bit-line capacitance and enables a larger number of rows or again smaller cell sizes.

In practice, the CFRAM architecture only enables a more compact cell layout under certain circumstances. For example, the benefit of sharing the diffusion regions of adjacent chain transistors decreases if the minimum cell size is limited by a large ferroelectric capacitor and not by the transistor. Moreover, since the CFRAM chain requires more local interconnect than the FeRAM cell, the area advantage of using a CFRAM architecture instead of a FeRAM architecture is very dependent on the ferroelectric manufacturing process. For example, the proposed T-structure for the chain capacitors, i.e. the alternatively sharing of top- and bottom-electrode between adjacent

chain capacitors (see Figure 81b above), may not be used in a 1-mask capacitor stack etch process. The 1-mask etch requires the capacitor's top- and bottom-electrodes to be identical, thus, employing the T-structure is not possible.

Performance

For a FeRAM array that is surrounded by an ideal periphery, the rise/fall time of the plate-line could become infinitely small. Readout to the bit-line could then be completed in nanoseconds, as ferroelectric switching is expected to occur in the sub-nanosecond region. For a CFRAM array, the rise/fall time of the plate-line would be the same, but this would not necessarily result in a faster readout because the plate-line is different from the cell plate-line (CPL) in a CFRAM. The CPL signal is delayed with respect to the PL signal due to the resistive path through the turned on transistors in-between (Figure 82). It was also shown that the impact of the body effect on this delay is very strong and, in fact, introduces a lower bound for the boosted voltage V_{PP}, making simultaneous low-voltage/high-speed operation difficult to achieve. The boosted voltage V_{PP} is directly related to the ohmic resistance of the turned-on cell transistors. When chosen too small, some transistors that should be turned on will stay off for a certain period during operation because of the body effect. A plot of V_{th} versus time during read access was presented earlier to give a qualitative impression. In Figure 83, all transistors except one suffer body effect because the BL nearest cell is accessed. Therefore, it is necessary to dimension V_{PP} large enough. Besides, the impact of chain length on plate-line capacitance is substantial and could be disadvantageous for the performance of future CFRAMs. To mitigate this effect and improve performance, the early plate-line scheme was proposed. Nevertheless, it seems that the standard FeRAM architecture is the better choice for high-speed ferroelectric memory.

Lifetime

As described the CFRAM concept requires all unselected cell transistors to be turned on. Thus, their gate voltage needs to be boosted all the time (or at least during read and write operation). In addition, the boosted voltage has to be large enough to achieve the design target performance but at least $V_{CC}+V_{th}$. Over-voltage in turn degrades the lifetime of the transistors gate oxide. This could result in a serious problem for future low voltage CFRAM memories or at least in a trade-off of lifetime versus performance.

Reliability

Two reliability issues are the readout voltage shift and noise across unselected cells. The readout voltage shift was investigated in detail and a technique for compensation was presented. The latter noise is caused by the voltage drop across turned-on cell transistors that appears during read and write operations. Fortunately, its amplitude becomes smaller with increased chain length as it is shared across the cells of a chain.

X. SCALING FERAM INTO THE GIGABIT GENERATION

A. OPERATING VOLTAGE SCALING

In the past years, the operating voltage of FeRAMs scaled roughly with $F^{1/3}$. However, it is expected to scale with F in the coming FeRAM generations similar to DRAM and SRAM. Operating voltages for a 0.1 µm FeRAM should be between 1.0 and 1.2 Volt (Figure 93).

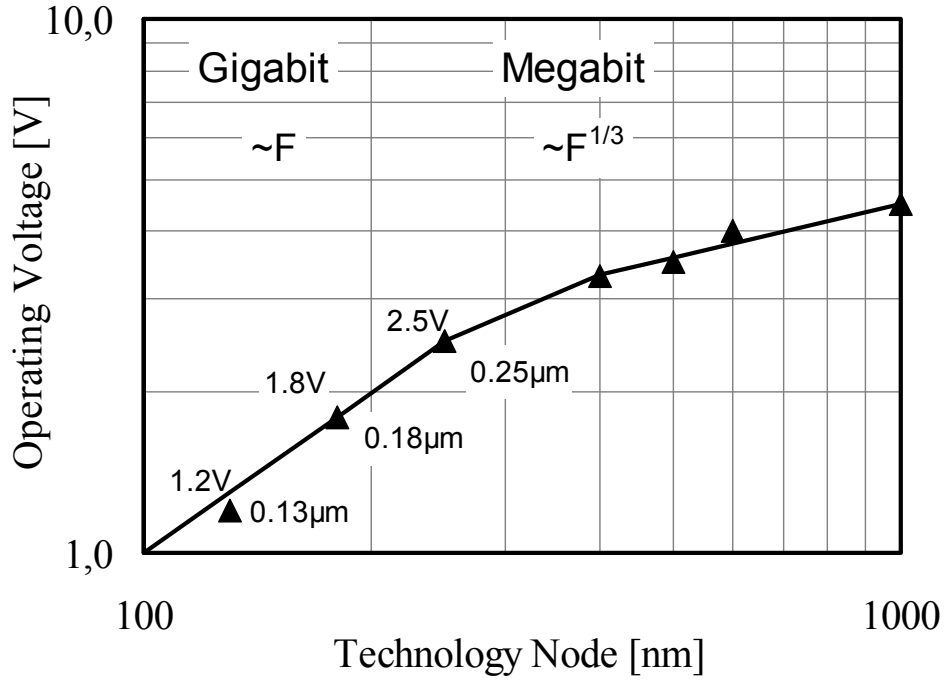


Figure 93: Operating voltage versus technology node

B. SCALING OF CELL CHARGE

As the minimum cell size is proportional to F^2, the area available for the planar ferroelectric capacitor is forced to be proportional to F^2, too. This results in a switched charge Q_{SW} that is also proportional to F^2 (see Figure 94).

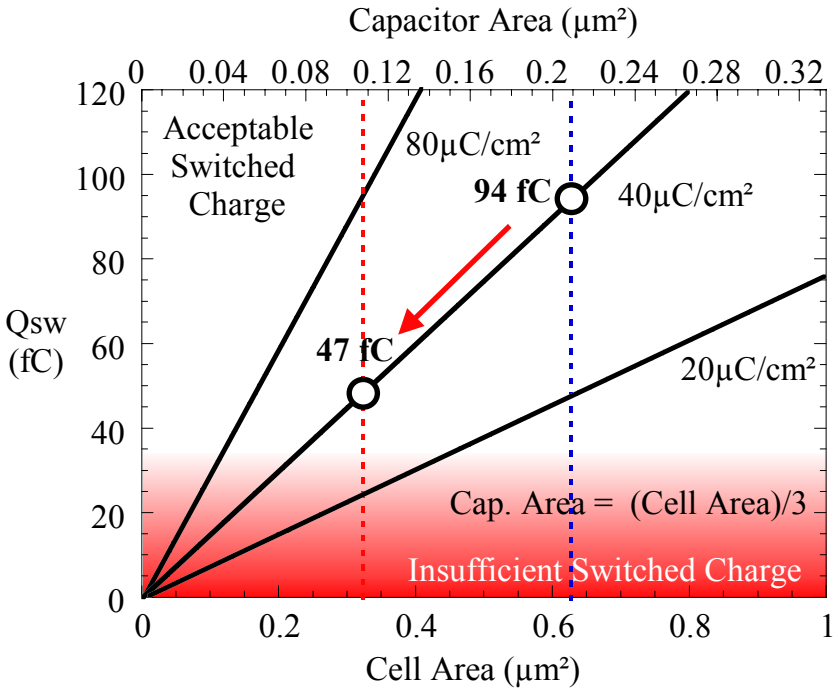


Figure 94: Planar ferroelectric capacitor scaling

The non-switching charge also depends on the film thickness t_{eff} and cell voltage V_{DD}:

$$Q_{NS} \propto \frac{A \cdot V_{DD}}{t_{eff}} \quad \rightarrow \quad Q_{NS} \propto F^2 \cdot \frac{V_{DD}}{t_{eff}}$$

Clearly, Q_{SW} cannot continue to scale with F^2 forever as the sensing signal V_S also decreases in proportion with Q_{SW}:

$$V_S \propto \frac{Q_{SW}}{C_{BL}} \qquad (C_{BL} >> C_S)$$

In previous years, V_S decreased from about 600 mV to 200 mV. However, the required minimum sensing signal for a cross-coupled latch sense amplifier has remained constant at about 70 mV. Q_{SW} can only continue to scale with F^2 if bit-line capacitance also scales with F^2. However, it will be shown next that this is not the case.

C. Bit-line Capacitance Scaling

Figure 95 is a graph of the typical bit-line capacitance versus technology node assuming 256 cells per bit-line and a bit-line that runs in METAL1. In the past, bit-line ca-

pacitance decreased with minimum feature size as indicated by the triangular data points that have been reported in literature. In this region – the Megabit era – bit-line capacitance is proportional to F.

However, this trend is unlikely to continue into the Gigabit era, as FeRAM is very similar to DRAM regarding this aspect and C_{BL} in DRAMs scales only with $F^{2/3}$ in this region [36]. Therefore, it is assumed that the bit-line capacitance in FeRAMs will also only scale with $F^{2/3}$. In Figure 95, the extrapolated bit-line capacitance values for 0.13 µm, 0.18 µm and 0.25 µm technologies are marked with dots.

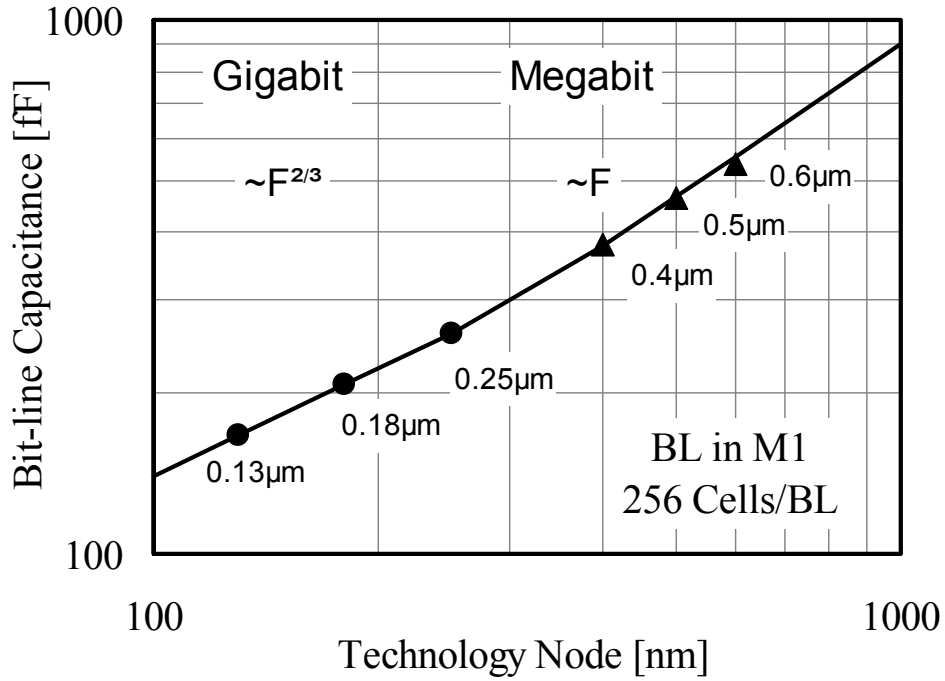


Figure 95: Scaling of bit-line capacitance

As discussed earlier, the switched charge currently scales with F^2, but bit-line capacitance does not. As V_S will stop scaling when its minimum value is reached, Q_{SW} will be forced to scale in proportion to C_{BL}. Thus,

$$Q_{SW} \propto V_S \cdot C_{BL} \qquad \xrightarrow{V_S \propto 1} \qquad Q_{SW} \propto F^{2/3} \text{ (!)}$$

When this happens, the area available for the ferroelectric capacitor will still have to scale with F^2 because cell size needs to scale with F^2. The use of 3D ferroelectric capacitor structures might be one possibility to meet these requirements.

D. Summary of Scaling Trends

Table 9 compares the scaling trends of cell parameters for FeRAM to DRAM. At a first glance, scaling trends in the Megabit era seem to be quite different between FeRAM and DRAM. This is because FeRAM, unlike DRAM, still has a lot of unused scaling potential. For instance, the sensing signal in DRAMs cannot be scaled anymore because it is already at its minimum value. DRAM already uses 3D capacitor technology whereas FeRAM has so far only used planar technology.

Table 9: Overview of cell parameter scaling trends

		Megabit Era		**Gigabit Era**	
		FeRAM	**DRAM**	**FeRAM**	**DRAM**
Cell size		$\sim F^2$		$\sim F^2$	
Cell voltage		$\sim F^{1/3}$		$\sim F$	
Coercive voltage		$\sim F^{1/3}$	*n/a*	$\sim F$	n/a
Dielectric thickness	t_{eff}	$\sim F^{1/3}$	$\sim F^{1/2}$	$\sim F$	*1*
Sensing signal	V_S	$\sim F$	1	1	
Bit-line capacitance	C_{BL}	$\sim F$	$\sim F^{2/3}$	$\sim F^{2/3}$	
Stored charge	Q_S	$\sim F^2$	$\sim F^{2/3}$	$\sim F^{2/3}$	
Capacitor area	A	$\sim F^2$	$\sim F^{5/6}$	$\sim F^{2/3}$	$\sim F^{-1/3}$

With a more mature FeRAM technology, these differences should decrease. In the Gigabit era, scaling trends for both memory technologies should be similar. However, there could be a major difference regarding required capacitor area. Because in DRAM, stored charge Q_S is also dependent on cell voltage, so the substantially decreasing V_{DD} ($\sim F$) in the Gigabit era forces the capacitor area A for DRAM to increase as shown next:

$$Q_S \propto V_S \cdot C_{BL} \qquad \text{and} \qquad Q_S \propto \frac{A \cdot V_{DD}}{t_{eff}}$$

$$\Rightarrow \frac{A \cdot V_{DD}}{t_{eff}} \propto V_S \cdot C_{BL}$$

$$\Leftrightarrow A \propto \frac{V_S \cdot C_{BL} \cdot t_{eff}}{V_{DD}}$$

$$\rightarrow A \propto \frac{1 \cdot F^{2/3} \cdot 1}{F} = F^{-1/3}$$

In contrast, the stored (=switched) charge in a FeRAM is in general independent from cell voltage, assuming the coercive voltage is sufficiently lower than the cell voltage so that the material can be brought into saturation. This fact might be advantageous for future FeRAMs because the capacitor area A can then decrease in proportion to C_{BL} for FeRAM:

$$Q_{SW} \propto V_S \cdot C_{BL} \qquad \text{and} \qquad Q_{SW} \propto A$$

$$\Rightarrow A \propto V_S \cdot C_{BL}$$

$$\rightarrow A \propto F^{2/3}$$

This implies that continued successful scaling of the coercive voltage is very important for ferroelectric memories. Experimental results for 60 nm and 85 nm PZT films are presented in Figure 96. Both films exhibit very low coercive voltages and saturate at less than 1.2V.

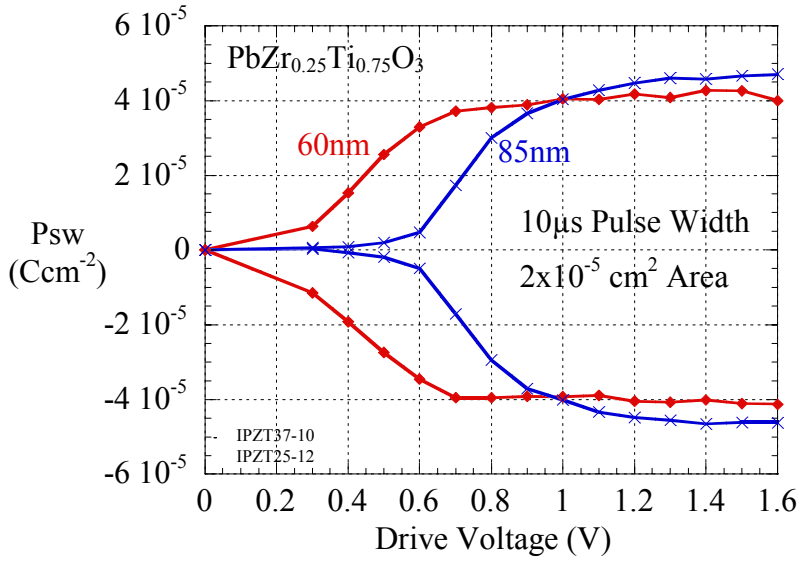


Figure 96: Scaling of coercive voltage [37]

E. Impact of Scaling on Performance

As previously described, the minimum access/cycle time of present ferroelectric memories is strongly limited by the rise/fall time of the plate-line signal, which is usually large. To determine the impact of CMOS scaling on access/cycle time, its effect on plate-line speed will be evaluated first. The characteristic time constant of the plate-line was approximated to be:

$$\tau_{PL} = R_{PLD} \cdot C_{PL} \qquad \text{with} \qquad R_{PLD} \approx \frac{V_{DD}}{I_{on} \cdot W}$$

As a first simplification, it is assumed that I_{on} does not scale down with the design rule[3]. If we further assume that the W/L ratio for each transistor in the plate-line driver will also remain constant, then I_{PLD} should be roughly proportional to F, the minimum feature size, as Figure 97 shows.

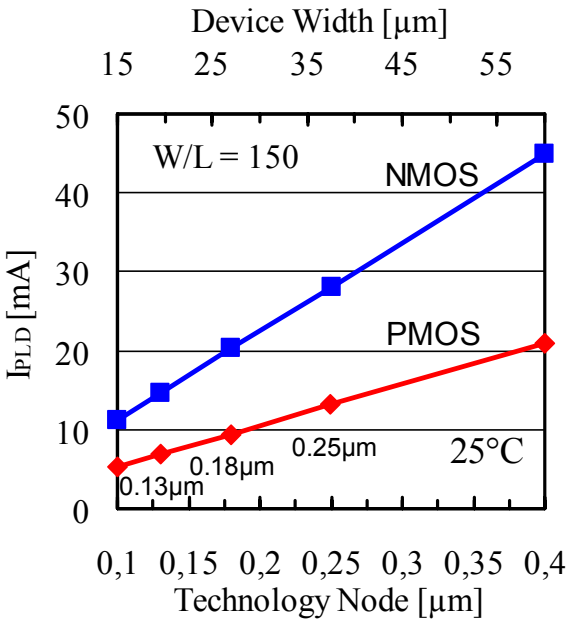


Figure 97: Plate-line driver current versus technology node and width

Scaling might also allow a more area-efficient layout of the plate-line drivers and, consequently, an increase in the W/L ratio.

[3] Int. Technology Roadmap for Semiconductors 2000 Update

It is very difficult to accurately predict the impact of scaling on performance for future FeRAMs as there are many dependencies and unknowns. However, based on what has been reported in the literature and derived in this chapter, it is possible to draw the following conclusions. Access and cycle times have already improved remarkably with design rules in the past (see Table 2 on page 4). In addition, the scaling study revealed that the time constant of the plate-line should decrease with F:

$$\tau_{PL} \propto \frac{Q_{SW} + Q_{NS}}{I_{PLD}} \propto \frac{F^2 + F^2}{F} \propto F$$

Consequently, plate-line speed should significantly improve for future FeRAM generations due to CMOS scaling alone. As the plate-line speed has a great impact on access and cycle times, these should improve as well. Thus, a substantial improvement in performance for next FeRAM generations can be expected. The large disadvantage of the slow plate-line signal (compared to DRAM) will then greatly diminish.

XI. CONCLUSION

Relevant circuits and techniques for design of present and future ferroelectric memories have been presented. The results of this work can be summarized as follows:

1. A ferroelectric capacitor model has been developed and implemented in a standard circuit simulator. The model is based on the Preisach theory for magnetic systems and accounts for the history dependency of the ferroelectric capacitor by memorizing *turning points*. A comparison between experimental and simulated hysteresis loop showed that the model achieves a very good match for both saturated and partial hysteresis loops. Furthermore, the model has been employed in an exemplary investigation of the impact of imprint on memory operation.

2. A first 0.35-µm 512-kbit FeRAM test chip has been presented. It features both a FeRAM and CFRAM memory array, whereas the FeRAM array can be switched between 2T2C or 1T1C operation. The FeRAM array also features 7 different ferroelectric capacitor sizes and a hierarchical bit-line architecture. To allow for maximum testing flexibility, all relevant timing signals and the reference voltage are externally adjustable.

3. A second 0.13-µm 4-Mbit embedded FeRAM test chip has been presented. The planning, design and testing phase for this chip was described. New circuits and techniques have been introduced that could become key to the characterization and design of future FeRAMs, including:

   (a) A novel sense amplifier that has truly independent sense- and write-back capability enabling a multitude of applications. For example, the multiple comparisons operation allows conducting a high-speed charge distribution measurement or identification of weak bits before they start causing errors. Since it is mainly comprised of the comparator well known from SRAM, there's little risk involved by employing the new sense amplifier.

(b) The high-speed measurement of cell charge distribution is capable of obtaining the full 4-Mbit charge distribution in about 45 ms. This opens new possibilities for on-chip characterization or quality control of ferroelectric memory cells.

(c) The presented grouped plate-line architecture aims to further improve plate-line speed in order to make FeRAM more competitive to DRAM. It is suitable for low voltage operation since, unlike the segmented plate-line architecture, it requires no over-voltage. In addition, it is area efficient and easy to implement.

(d) A special test mode has been implemented to reduce the time required for the memory cell fatigue test. Conventional tests require about 6 months to exercise only a few memory cells per chip 10^{14} times. The accelerated fatigue mode allows exercising many more cells (for this chip: 8,192) in only 23 days for 10^{14} times or in only 6 hours for 10^{12} times.

(e) The chip allows the characterization of key circuits such as: sense amplifier (offset voltage measurement), bit-line (capacitance measurement), and word-line boosting (measurement of write-"1" voltage).

4. Important issues for the CFRAM architecture have been presented and analyzed in detail. New circuits and techniques that try to overcome some of the issues have been proposed:

(a) The impact of chain length on plate-line capacitance is substantial and could be disadvantageous for the performance of future CFRAMs. To mitigate this effect and improve performance, the early plate-line scheme was proposed.

(b) The proposed replica chain makes the bit-line capacitance independent from cell position and successfully eliminates the readout voltage shift.

A comparison between CFRAM and FeRAM revealed that:

(a) The area advantage of CFRAM over FeRAM seems to be technology dependent.

(b) The impact of the body effect on circuit performance is much stronger for CFRAM than for FeRAM and, in fact, introduces a lower bound for the boosted voltage V_{PP}, making simultaneous low-voltage/high-speed operation difficult.

(c) The permanent gate over-voltage that unselected cell transistors experience could degrade the lifetime of their gate-oxide.

APPENDIX

A. Source Code for the Ferroelectric Capacitor Model

```
/*
** Implementation of Ferroelectric Capacitor Model
**
** Author: Juergen Rickes
**
*/

// CONSTANTS
#define E0 8.8541e-12
#define PI 3.14159265359
#define MAX_TP 50

////////////////////////////////
// SATURATION CURVE FUNCTION //
////////////////////////////////

real F(real V, real Vco, real a)
{
   return 2*atan(a*(V-Vco))/PI;
}


/////////////////////////////
// INTERFACE DECLARATION //
/////////////////////////////

module fecap (MINUS, PLUS)
   (area, d, Er, Ps_, Pr_, Vcp, Vcn, Rleak)

node [V,I] MINUS,PLUS;

parameter real area = 1.00 from (0:inf); // cap area [µm²]
parameter real d = 170 from (0:inf);     // thickness [nm]
parameter real Er = 350 from [0:inf);    // dielectric constant
parameter real Ps_ = 30 from (0:inf);    // saturation pol.[µC/cm²]
parameter real Pr_ = 25 from (0:inf);    // remanent pol.[µC/cm²]
parameter real Vcp = 1.5 from (0:50);    // coercive voltage [V]
parameter real Vcn = -1.5 from (-50:0);  // neg. coercive voltage
parameter real Rleak = 5k from (0:inf];  // leakage [R*cm²]

{

////////////////////////////////////
// VARIABLES AND INITIALIZATION //
////////////////////////////////////
```

```
 // leakage resistor and linear capacitor parallel to FeCAP
 capacitor C_linear (PLUS,MINUS) (c=E0*Er*area*1e-3/d);
 resistor R_leakage (PLUS,MINUS) (r=1e8*Rleak/area);

 real Vfe, P, Ps;       // voltage, polarization, sat. polarization
 real Vch;              // actual coercive voltage
 real Vtp[MAX_TP],Qtp[MAX_TP];  // voltage/charge of turning point
 integer ix, ix_new, s;         // index of turning point array
 real a,mi;

 initial {
  ix = 0;
  Ps = 1e-14 * area * Ps_;
  a = -tan(PI/2*Pr_/Ps_)/Vcn;
    Vfe = 0; P = Ps*F(Vfe,Vcp,a);
  Vtp[ix] = inf; Qtp[ix] = Ps;
  Vtp[++ix] = -inf; Qtp[ix] = -Ps;
  Vtp[++ix] = Vfe; Qtp[ix] = P;
}


////////////////////////////
// MAIN PART OF THE MODEL //
////////////////////////////

analog {

if ($analysis("dc")) {I(PLUS,MINUS) <- 0;} // dc current is zero
else
{
  Vfe = V(PLUS,MINUS);

 // determine last direction
     s = (Vtp[ix%2]>0)   ?   1 :   // ascending voltages
                            -1 ;   // descending voltages

 // check if direction changed
 if (Vfe*s<Vtp[ix]*s) {
  ix_new = ix + 1;
  s = -s;
 } else  ix_new = ix;

 // search for correct position in table
 while (Vfe*s>=Vtp[ix_new-2]*s) ix_new -= 2;

 // set coercive voltage for actual direction
 Vch  = (s==1) ? Vcp : Vcn;

 // Calculate scaling factor mi
 mi = (Qtp[ix_new-1]-Qtp[ix_new-2]);

 // the following if-clause prevents a division by zero
 if (mi!=0) {
     mi /= F(Vtp[ix_new-1],Vch,a)-F(Vtp[ix_new-2],Vch,a);
```

```
 }

 // Calculate P(V)
 P = Qtp[ix_new-1] - mi*(F(Vtp[ix_new-1],Vch,a)-F(Vfe,Vch,a));

 // accept new ix and store last (V,Q) pair
 ix = ix_new;
 Vtp[ix] = Vfe;
 Qtp[ix] = P;

 I(PLUS,MINUS) <- dot(P);}  // current is derivative of charge
 }
}
```

BIBLIOGRAPHY

[1] Choi, M.-K. et al.: A 0.25µm 3.0V 1T1C 32Mb Nonvolatile Ferroelectric RAM with Address Transition Detector (ATD) and Current Forcing Latch Sense Amplifier (CFLSA) Scheme. ISSCC Dig.Tech.Papers (2002), pp. 162-163.

[2] Takashima, D. et al.: A 76mm² 8Mb Chain Ferroelectric Memory. ISSCC Dig.Tech.Papers (2001), pp. 40-41.

[3] Itoh, K.: VLSI Memory Chip Design. Springer 2001. ISBN 3-540-67820-4.

[4] Itoh, K.: VLSI Memory Design. Tokyo (in Japanese): Baifukan 1994.

[5] Dennard, R. H.: Field-Effect Transistor Memory. U.S. Patent 3387286 (1968).

[6] Takahashi, T. et al.: A Multi-Gigabit DRAM Technology with 6F² Open-Bit-line Cell Distributed Over-Driven Sensing and Stacked-Flash Fuse. ISSCC Dig.Tech.Papers (2001), pp. 380-381.

[7] Kirihata, T.: A 390mm² 16 Bank 1 Gb DDR SDRAM with Hybrid Bitline Architecture. ISSCC Dig.Tech.Papers (1999), pp. 422-423.

[8] Takahashi, O. et al.: 1GHz Fully Pipelined 3.7ns Address Access Time 8kx1024 Embedded DRAM Macro. ISSCC Dig.Tech.Papers (2000), pp. 396-397.

[9] Murotani, T.: A 4-Level Storage 4 Gb DRAM. ISSCC Dig.Tech.Papers (1997), pp. 74-75.

[10] Nambu, H.: A 550-ps Access, 900-MHz, 1-Mb ECL-CMOS SRAM. Symp.VLSI Circuits Dig.Tech.Papers (1999)

[11] Various reported SRAM chips. ISSCC Dig.Tech.Papers (1999)

[12] Nozoe, A.: A 256 Mb Multilevel Flash Memory with 2 MB/s Program Rate for Mass Storage Applications. ISSCC Dig.Tech.Papers (1999), pp. 110-111.

[13] Imamiya, K.: A 130mm² 256 Mb NAND Flash with Shallow Trench Isolation Technology. ISSCC Dig.Tech.Papers (1999), pp. 112-113.

[14] Smolenskii, G. A.: Ferroelectrics and Related Materials. Gordon and Breach 1984.

[15] Warren, W. et al.: Imprint in Ferroelectric Capacitors. Jpn.J.Appl.Phys. vol. 35 (1996) no. 2B, pp. 1521-1524.

[16] Takashima, D. et al.: High-Density Chain Ferroelectric Random-Access Memory (CFRAM). Symp.VLSI Circuits Dig.Tech.Papers (1997), pp. 83-84.

[17] Koike, H. et al.: A 60-ns 1-Mb Nonvolatile Ferroelectric Memory with a Nondriven Cell Plate Line Write/Read Scheme. IEEE J.of Solid-State Circuits vol. 31 (1996) no. 11, pp. 1625-1634.

[18] Chung, Y., Jeon, B.-G., and Suh, K.-D.: A 3.3-V, 4-Mb Nonvolatile Ferroelectric RAM with Selectively Driven Double-Pulsed Plate Read/Write-Back Scheme. IEEE J.of Solid-State Circuits vol. 35 (2000) no. 5, pp. 697-704.

[19] Yamada, J. et al.: A 128kb FeRAM Macro for a Contact/Contactless Smart Card Microcontroller. ISSCC Dig.Tech.Papers (2000), pp. 270-271.

[20] Jeon, B.-G. et al.: A 0.4-μm 3.3-V 1T1C 4-Mb Nonvolatile Ferroelectric RAM with Fixed Bitline Reference Voltage Scheme and Data Protection Circuit. IEEE J.of Solid-State Circuits vol. 35 (2000) no. 11, pp. 1690-1694.

[21] Sheikholeslami, A. and Gulak, P. G.: A Survey of Circuit Innovations in Ferroelectric Random-Access Memories. Proceedings of the IEEE vol. 88 (2000) no. 5, pp. 667-689.

[22] Jeon, B.-G. et al.: A 0.4μm 3.3 V 1T1C 4 Mb Nonvolatile Ferroelectric RAM with Fixed Bitline Reference Voltage Scheme and Data Protection Circuit. ISSCC Dig.Tech.Papers (2000), pp. 272-273.

[23] Mayergoyz, I. D.: Mathematical Models of Hysteresis. New York: Springer Verlag 1991.

[24] Bartic, A. T.: Physical Modeling, Characterization and Design of Ferroelectric Non-volatile Memories. PhD Dissertation. Katholieke Universiteit Leuven 2001.

[25] Bartic, A. T. et al.: Preisach Model for the Simulation of Ferroelectric Capacitors. Journal of Applied Physics vol. 89 (2001) no. 6, pp. 3420-3425.

[26] Bartic, A. T. et al.: Implementation of a Ferroelectric Capacitor Model Using the Preisach Hysteresis Theory. Int.Ferroelectrics vol. 26 (2001) no. 1-4, pp. 987-994.

[27] Jiang, B. et al.: Computationally Efficient Ferroelectric Capacitor Model for Circuit Simulation. Symp.VLSI Technology Dig.Tech.Papers (1997), pp. 141-142.

[28] Goebel, H. et al.: Distribution function integral method for modeling ferroelectric devices. Extended Abstracts of the International Conference on Solid State Devices and Materials (1999), pp. 386-387.

[29] Geib, H. et al.: Experimental Investigation of the Minimum Signal for Reliable Operation of DRAM Sense Amplifiers. IEEE J.of Solid-State Circuits vol. 27 (1992) no. 7, pp. 1028-1035.

[30] Takashima, D. and Kunishima, I.: High-Density Chain Ferroelectric Random Access Memory (Chain FRAM). IEEE J.of Solid-State Circuits vol. 33 (1998) no. 5, pp. 787-792.

[31] Jeon, B.-G. et al.: A novel cell charge evaluation scheme and test method for 4Mb nonvolatile ferroelectric RAM. 6th International Conference on VLSI and CAD (1999), pp. 281-284.

[32] Takashima, D. et al.: A Sub-40-ns Chain FRAM Architecture with 7-ns Cell-Plate-Line Drive. IEEE J.of Solid-State Circuits vol. 34 (1999) no. 11, pp. 1557-1563.

[33] Takashima, D. et al.: A Sub40-ns Random-Access Chain FRAM Architecture with a 7ns Cell-Plate-Line Drive. ISSCC Dig.Tech.Papers (1999), pp. 102-103.

[34] Takashima, D., Oowake, Y., and Kunishima, I.: Gain Cell Block Architecture for Gigabit-Scale Chain Ferroelectric RAM. Symp.VLSI Circuits Dig.Tech.Papers (1999), pp. 102-103.

[35] Takashima, D. et al.: A 76-mm^2 8-Mb Chain Ferroelectric Memory. IEEE J.of Solid-State Circuits vol. 36 (2001) no. 11, pp. 1713-1720.

[36] Nitayama, A., Kohyama, Y., and Hieda, K.: Future Directions for DRAM Memory Cell Technology. Tech.Dig.IEEE Int.Electron Devices Meeting (1998)

[37] Moise, T.: Int.Ferroelectrics (2001)

LEBENSLAUF

Persönliche Daten

Name:	Jürgen Thomas Rickes
Geburtsdatum:	17. Juli 1973
Geburtsort:	Neuwied
Familienstand:	verheiratet, keine Kinder
Nationalität:	deutsch
Religion:	katholisch

Ausbildung

April 1980 – März 1984	Grundschule Zweifall
April 1984 – Mai 1993	Goethe-Gymnasium in Stolberg Rhld.
Mai 1993	Abitur
Okt. 1993 – Aug. 1998	Elektrotechnik-Studium an der RWTH Aachen
Oktober 1995	Vordiplom
August 1998	Diplom, „summa cum laude“
Jan. 1999 – Dez. 2002	Elektrotechnik-Promotion an der RWTH Aachen

Berufstätigkeit

Mai 1995 – Juli 1997	Studentische Hilfskraft am Institut für Technische Mechanik der RWTH Aachen (Prof. Peters)
Okt. 1995 – Feb. 1996	Studentische Hilfskraft am Institut für Angewandte Mathematik der RWTH Aachen (Prof. Bemelmans)
Feb. 1997 – Apr. 1997	Praktikum, Applied Materials, Danbury, CT, USA
Apr. 1997 – Jun. 1997	Praktikum, Philips Forschungszentrum, Aachen
März 1996 – Dez. 1998	Studentische Hilfskraft am Institut für Werkstoffe der Elektrotechnik der RWTH Aachen (Prof. Waser)
Jan. 1999 – Dez. 2001	Wissenschaftlicher Angestellter am Forschungszentrum Jülich
April 2000 – heute	Leitender Ingenieur bei Agilent Technologies, Santa Clara, CA, USA

CURRICULUM VITAE

Personals

Name:	Jürgen Thomas Rickes
Date of Birth:	07/17/1973
Place of Birth:	Neuwied, Germany.
Marital Status:	married, no children
Nationality:	German
Religion:	Roman Catholic

Education

April 1980 – March 1984	Primary school Zweifall
April 1984 – May 1993	Goethe-Gymnasium in Stolberg Rhld.
May 1993	Abitur (equivalent to A levels).
Oct. 1993 – Aug. 1998	Electrical Engineering, RWTH Aachen
October 1995	Intermediate Diploma
August 1998	Diploma, "summa cum laude"
Jan. 1999 – Dec. 2002	Ph.D. (to be awarded) in Electrical Engineering

Work Experience

May 1995 – July 1997	Student, Chair for Mechanical Engineering, Prof. Peters
Oct. 1995 – Feb. 1996	Student, Chair for Applied Mathematics, Prof. Bemelmans
Feb. 1997 – Apr. 1997	Internship, Applied Materials, Danbury, CT, USA
Apr. 1997 – Jun. 1997	Internship, Philips Research Laboratories, Aachen, Germany
March 1996 – Dec. 1998	Student, Chair for Materials Science in Electrical Engineering, Prof. Waser
Jan. 1999 – Nov. 2001	Scientific Staff Member, Research Center Juelich, Germany
April 2000 – today	Lead design engineer, FeRAM R&D, Agilent Technologies Inc., Santa Clara, CA