INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9311505

Digital design of a one megabit magneto-resistive memory incorporating efficient hardware fault tolerance

Kohl, Clinton Edward, Ph.D.

Iowa State University, 1992

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

Digital design of a one megabit magneto-resistive memory incorporating efficient hardware fault tolerance

by

Clinton Edward Kohl

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Electrical Engineering and Computer Engineering
Major: Computer Engineering

Approved: Members of Committee:

Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1992

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I thank my major professors, Dr. Art Pohm and Dr. Marwan Hassoun. I enjoyed working with them and consider it a privilege to have been a part of the MRAM design team. Both men taught me a great deal and provided needed guidance and support throughout this project. Thanks also to the other members of our design team, Indumini and Manel Ranmuthu and Chip Comstock, without whom this work would not have been completed. It was a joy to work with these colleagues. I also appreciate the technical support and the financial assistance of NonVolatile Electronics and Honeywell.

I thank my wife, Susan, and my son, Caleb, who encouraged me and brought joy to my life. Most of all, I acknowledge Jesus Christ, my Lord, who has given me every gift and ability I have, as well as the strength to complete this dissertation.

ABSTRACT

The research documented in this thesis was undertaken to improve and advance magneto-resistive (MR) memory design. This new memory technology shows great promise in many areas of modern computer systems.

Our research team completed a design and partial layout of the first MR memory to operate in the voltage mode. This thesis focuses on the related areas of architecture and hardware fault tolerance. In the area of architecture, an overall chip organization was developed. Compact and space efficient layouts of MR cells and supporting circuitry also were designed. In the area of hardware fault tolerance, various techniques for improving chip yield and reliability in the presence of hardware failures, by means of, spares and error-correcting and detecting logic, were investigated and reported.

Chapter 1 introduces computer memory systems and emphasizes the key advantages of MR memories over existing memory technologies. Chapter 2 provides information about the historical and theoretical background of MR memories. Chapter 3 provides an overview of the 1-megabit chip, in a top-down format, and provides a foundation on which the remaining more detailed chapters will be built. Chapter 4 contains details of the sense line design, and Chapter 5 discusses the details of support logic. Chapter 6 discusses yield-enhancing techniques, and Chapter 7 contains concluding remarks.

1. INTRODUCTION

1.1 Objective

The research documented in this thesis was undertaken to improve and advance magneto-resistive (MR) memory design. This new memory technology shows great promise in many areas of modern computer systems.

Our research team completed a design and partial layout of the first MR memory to operate in the voltage mode. This thesis focuses on the related areas of architecture and hardware fault tolerance. In the area of architecture, an overall chip organization was developed. Compact and space efficient layouts of MR cells and supporting circuitry also were designed. In the area of hardware fault tolerance, various techniques for improving chip yield and reliability in the presence of hardware failures, by means of, spares and error-correcting and detecting logic, were investigated and reported.

1.2 Introduction to Computer Memory Systems

Computer system performance is related closely to the memory technologies used by these systems. Many storage devices already exist, and many more are being developed and enhanced. Memory storage devices vary in speed and capacity from small but high speed cache memories to large but slow tape drives. A significant amount of memory dollars are spent on semiconductor memories such as DRAM (Dynamic Random Access Memory) and SRAM (Static Random Access Memory), and an ever greater amount is spent on hard-disk storage units. Figure 1.1 and 1.2 illustrate the basic cell designs for DRAM and SRAM.

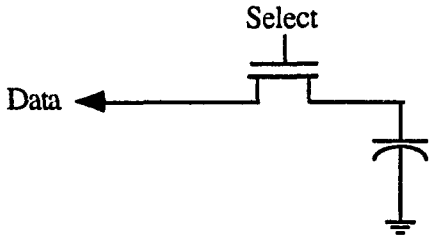

Figure 1.1 Simplified DRAM cell

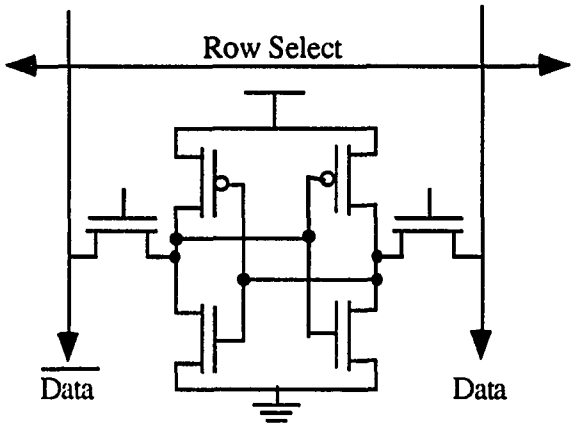

Figure 1.2 Six transistor static RAM cell

1.3 Existing Memory Technologies

Most computers benefit from the use of a memory hierarchy, which typically includes a small high-speed cache near the CPU, a much larger but slower main memory, and a hard-disk drive, which is very large but several orders of magnitude slower than main memory. Figure 1.3 illustrates the typical memory hierarchy.

Attributes of several popular storage systems are listed in Table 1.1 [1]. Figure 1.4 illustrates a plot of cost per Mbyte versus access time of the various memory technologies listed in Table 1.1 [1]. In the region between the dashed lines, which is known as a *gap*, no cost-effective memory technology exists. If a cost-effective memory were to become

available in the present memory gap, most systems could be better optimized for price and performance.

As computers increase in speed, larger and faster memory systems will be required to service them. But performance of storage systems is improving less rapidly than that of CPUs [2:427]. Figure 1.5 illustrates the increasing performance gap between CPUs and DRAMs. Although in recent years, disks have improved in terms of higher density and lower power consumption and price, only modest improvements in speed have occurred [2:518]. Such performance trends will create an even greater need for efficient and cost-effective memory systems and memory hierarchies.

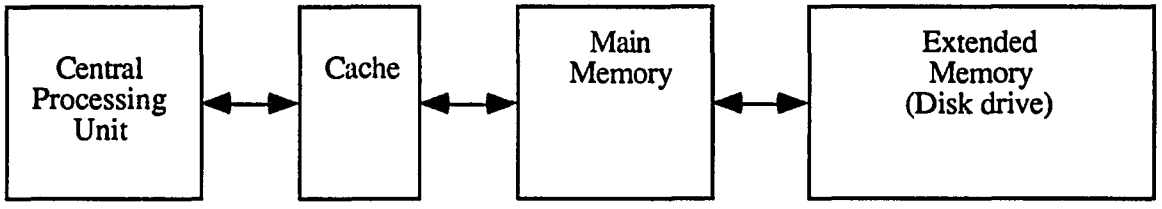

Figure 1.3 Typical memory hierarchy

Table 1.1 Attributes of popular storage systems

Technology	Cost ($/MByte)	Access Time	Unit Capacity (Mbytes)
Static RAM	760-12,200	3.5-200 ns	.03-1.0
Dynamic RAM	120-500	65-200 ns	.25-4.0
Bubble Memory	1000	10 msec	1.0
Winchester Disk	4-10	12-20 msec	20-1600
Optical Disk	12-24	40-100 msec	256-450
Half-Inch Tape	4-20	20-50 sec	95-1280
Quarter-Inch Tape	4-12	10-45 sec	10-525

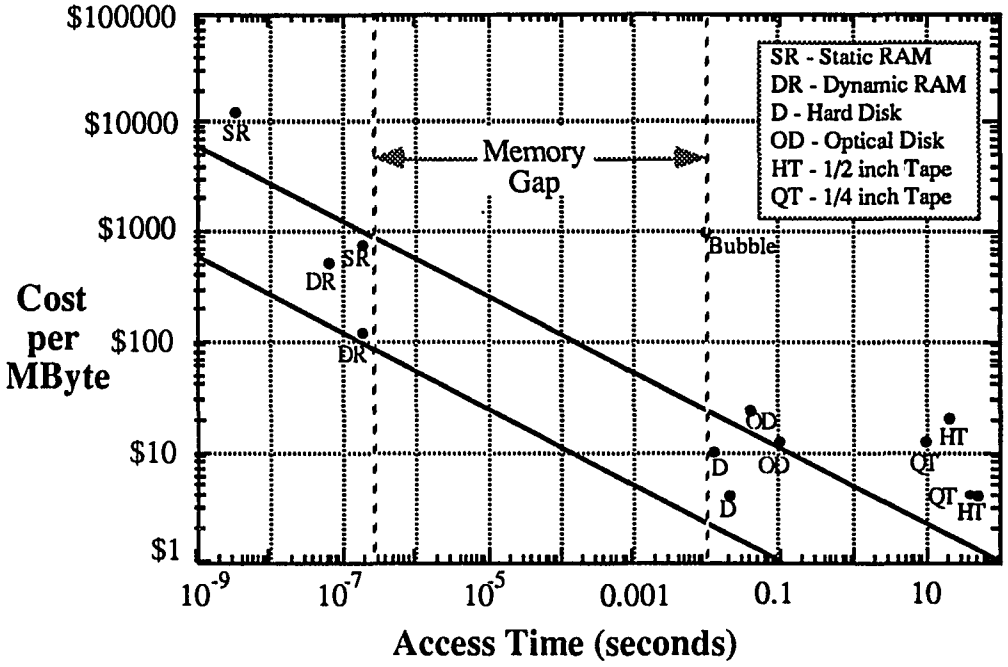

Figure 1.4 Performance attributes of popular storage devices

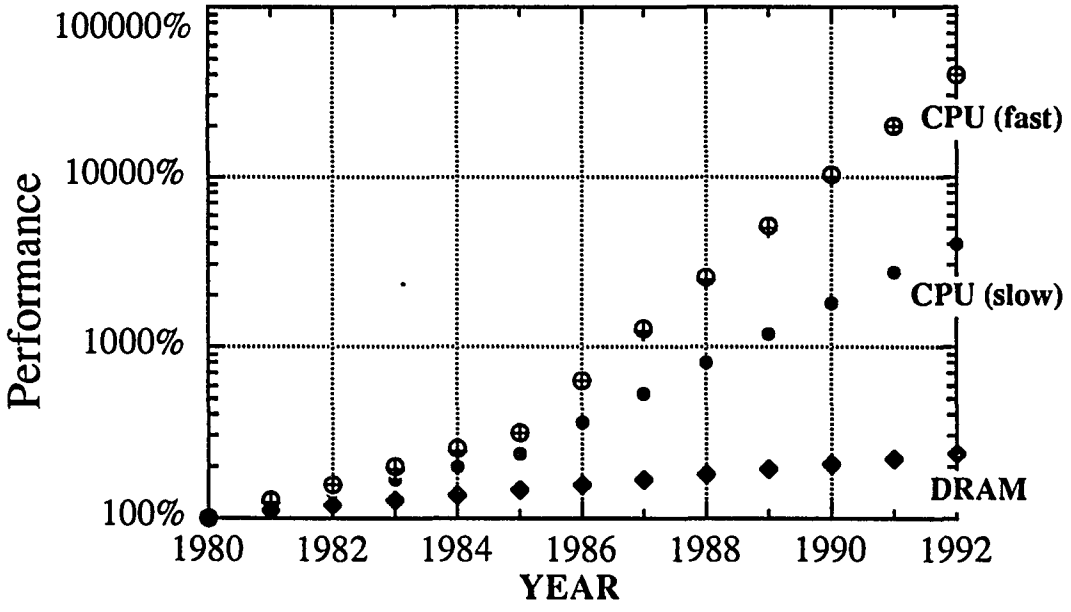

Figure 1.5 Performance of CPUs and DRAMs [2:519]

1.4 Factors Affecting Cost of Integrated Circuits

Many factors contribute to the cost of high-volume integrated circuits. The most prevalent factors are associated with processing the die, testing, packaging, and yield. Equation 1.1 describes the relations among these various cost elements [2:55]:

$$\text{Cost of Integrated Circuit} = \frac{\text{Cost of Die} + \text{Cost of Testing Die} + \text{Cost of Packaging}}{\text{Final Test Yield}}. \quad (1.1)$$

All factors must be given close attention to ensure the lowest possible cost. The cost of a die can be broken down further into its main cost contributors, as represented in equation 1.2 [2:59]:

$$\text{Cost of Die} = \frac{\text{Cost of Wafer}}{\text{Die per Wafer} * \text{Die Yield}}. \quad (1.2)$$

The wafer cost is tied closely to the number of masking steps required in processing. In short, as the number of masking levels increase, so does the cost. Yield (the ratio of good chips to the total number of chips on a wafer) also correlates closely with the number of masking steps. In that defects can be created at each step in the manufacturing process, the more steps, the more opportunity for failed parts. Lithography, wafer size, and chip area all contribute to the number of die per wafer. Many design factors, including chip area, internal redundancy, and sparing affect die yield. Minimizing chip area is essential to minimizing overall cost. Table 1.2 shows DRAM technology trends.

To be competitive an MRAM (Magnetic Random Access Memory) product will need to outperform its DRAM counterpart in one or more of the technological areas mentioned in Table 1.2. MRAM is most likely to excel in terms of having fewer masking steps and smaller cell sizes, but will likely be slower than DRAMs.

Table 1.2 DRAM technology trends

Year Introduced 1M/yr.	1983	1985	1988	1991	1994	1997	2001
DRAM Size (Mbytes)	0.256	1	4	16	64	256	1024
Vdd	5	5	5	5	3.3	3.3	3.3
Minimum Feature (μ)	1.5	1.0	0.8	0.5	0.3	0.25	0.15
Metal Pitch (μ)	5.0	3.5	2.5	1.5	1.0	0.8	0.78
Cell Size(μ^2)	38	20	9	4.5	2	0.9	0.4
Die Size (mm^2)	41	59	78	98	190	400	415
Wafer Diameter (mm)	150	150	150	250	250	250	300
Die/Wafer	360	263	191	271	162	103	100
Mask Levels (Ave..)	**10**	**16**	**18**	**22**	**25**	**28**	**32**

Source: 2001 Strategy by Microstructure Sciences TAB.

1.5 The Competitiveness of MR Memory Technology

Magneto-resistive memories have a number of qualities making them both viable and competitive with other memory products. Table 1.3 lists several of the key advantages of MRAM. These advantages, which are found individually in a variety of existing products, are all found in MRAM, making it an attractive solution for numerous memory problems. A subsection of this chapter discusses each advantage.

Table 1.3 Key advantages of magneto-resistive memories

1 Simple Manufacturing Process
2 Very High Densities Possible
3 Nonvolatile
4 Radiation Hard
5 No Wear-out
6 Fast write cycle

1.5.1 Simple manufacturing process

The present MRAM manufacturing process adds one masking step beyond the 11 masking steps of the standard CMOS process. This level of complexity stands in contrast to the 16-22 masking steps required to make today's DRAMs. Future DRAMs are projected to

require 25-32 masking steps, whereas even highly dense, advanced MRAM chips will likely require only 13-16 such steps [1]. Due to this reduction in complexity MRAM should have a significant processing cost advantage.

1.5.2 Highly dense cell configurations

Many fundamental construction differences exist between DRAM and MRAM cells. One difference is that a DRAM cell consists of two components: a capacitor to store charge and a selection transistor. This results in a ratio of one transistor per cell. Figure 1.6 illustrates a basic DRAM cell schematic.

Consisting of two key elements a sense line and a word line, an MRAM cell is quite different from a DRAM cell. The sense line used in this design and shown in Figure 1.7 contains 16 MR cells that share two transistors.

The gate transistor, when selected, allows current to flow through the sense line, and the mux transistor gates out the signal to be sensed. A detailed description of sense line operation can be found in Chapter 3. The ratio of transistors per MR cell is 1:8, much smaller than the 1:1 ratio of DRAM. Although gate and mux transistors must be larger than the selection transistor used in a DRAM, these transistors are placed under the bits themselves and add little area to the MR array. Details regarding placement and size of gate and mux transistors can be found in Chapter 3. Future MRAM designs will likely have 32 to 64 elements per sense line, which should make MRAM even more competitive than it is now. Research in the area of ultra dense M-R bits suggests that cells could be made with an area of $6\lambda^2$. This is two to three times the density achieved in current DRAMs, with the same lithography limits. Furthermore, M-R element arrays could be made with nearly the density established by first and second level metal pitch [3].

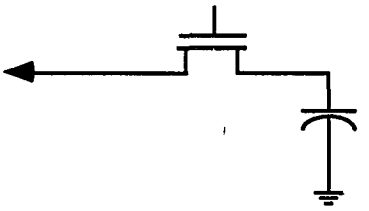

Figure 1.6 Basic DRAM cell construction

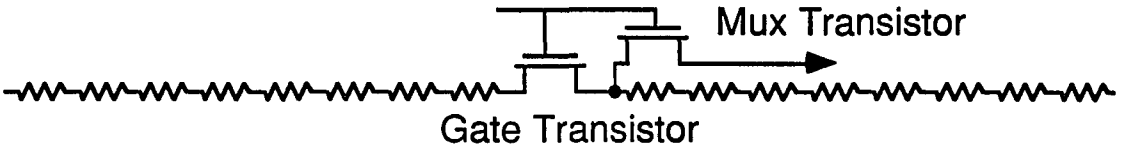

Figure 1.7 Basic MR cell construction

1.5.3 Nonvolatile

MRAM is nonvolatile, a characteristic that it shares with EPROMs and EEPROMs. In sharp contrast are DRAM and SRAM. Because the data is retained even in the absence of power, MRAM is especially useful for low-energy operations, for example, a battery-powered remote data-acquisition system. In such an environment, the memory can be powered down at all times except for brief periods during which data are actually being recorded. The nonvolatile nature of MRAM places it in a position to replace hard-disk drives. Through the use of wafer scale technology, large nonvolatile MR memories (≈ 250Megabytes) could be built to replace hard-disk drives, which are approximately 1,000 times slower [4].

1.5.4 Radiation hard

The data stored in MR memories are retained in the presence of radiation. Military and Space industries are interested in this technology because of its rad-hard properties, and NASA and the Department of Defense have funded modern MRAM development heavily.

1.5.5 No wear out

Unlike other nonvolatile IC memories such as EPROM and EEPROM, MRAM has an unlimited number of read/write cycles. Endurance tests have been performed on MR cells in which cells were written at 60 MHz over a 15-month test interval. During this time, the tested bits underwent $2x10^{15}$ write cycles, with no indication of wear [5].

1.5.6 Fast-write cycle

Another distinctive property of MRAM is its fast-write cycle. Its cells have been written in 2 ns, a time seemingly limited by the speed of the drive circuitry in bringing up the appropriate current levels. Certain high-speed data acquisition systems could benefit from the application of MRAM, particularly those, which record high speed events rather infrequently such as Digital Camera Memory. Another specialized application of MRAM is shadow RAM, which tracks the contents of a high-speed cache memory, and can be read out in the event of main cache failure [4].

2. MAGNETO-RESISTIVE MEMORY

2.1 History and State of the Art

Magneto-resistive memory elements are based on the concepts of William Thomson who, in 1857 identified the magneto-resistance effect in ferromagnetic materials [6]. His discovery was largely forgotten because of the large currents required for the materials and the sizes available at that time. Until recently, inductive sensing was the most practical and popular method of sensing magnetic memory elements. With the development of integrated circuit technology, magneto-resistance has proved a valuable means of sensing and storing information. As the size and the thickness of magnetic thin films have decreased, huge currents are no longer necessary to achieve useful results.

The predecessor to the modern MR memory cell is the crosstie wall memory, invented by L.J. Shwee in the early 1970s [7]. Although it took advantage of the magneto-resistive effect, this memory cell produced a very small signal. In 1983, the modern MRAM cell was conceived. Dr. Daughton from Honeywell & Dr. Pohm from ISU filed and received a patent from the US patent office for their longitudinal and transverse MR bits [8]. These new designs constitute the basis of state-of-the-art MR memory technology.

Following the 1983 patent, the Honeywell Corporation, having received major defense funding, developed MR memory technology further and is now producing in small quantities a 64K MR memory.

In 1989, NonVolatile Electronics (NVE) was formed as a friendly spin-off from Honeywell. The purpose of NVE is to exploit MR technology in the commercial memory market. Currently, NVE shares a free interchange of information with Honeywell.

2.2 Present Research Effort

Researchers at Iowa State University have made significant contributions to MR memory development over the past seven years and NVE recently has funded an ISU research team to further magneto-resistive memory technology. Research funding has been provided specifically to develop techniques for improving the density of MR memories. To prove the effectiveness of these new techniques, a one megabit (128k by 8) MR memory is being designed.

Recently, drastically increased interest in materials exhibiting large magneto-resistive effects has arisen. IBM and Seagate are two disk-drive manufacturers researching and developing MR head sensors to be used in high-performance hard-disk drives.

2.3 Magneto-Resistive Basics

Magneto-resistance is simply a change in the electrical resistance of a magnetic substance in the presence of a magnetic field. This effect is quantum mechanical in nature and involves electron spin-orbital interaction. Magneto-resistance is defined formally as:

$$R=R_O+\Delta R\cos^2\theta \qquad (2.1)$$

where: R_O = the resistance when $\theta = 90°$,
θ = the angle the current makes with the magnetization,
ΔR = Magneto-resistance term, and
$\Delta R/R$ =MR coefficient.

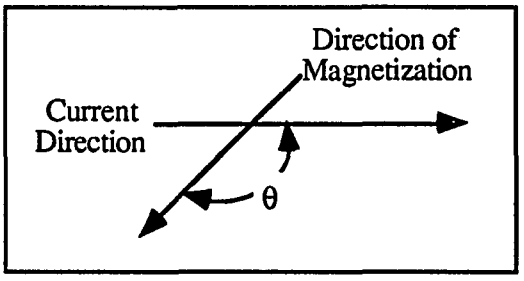

Figure 2.1 Graphic definition of theta

The signal level output from an MR element is closely related to the MR coefficient. As this coefficient increases, so does signal level. This correlation has a significant impact on both minimum element size, and speed at which the element can operate. Until recently, this coefficient has been near 2%, but recent research by NVE has produced substances yielding MR coefficients as large as 6%. A 60Ω element with a 6% MR coefficient would exhibit a maximum change in resistance from 60Ω to 60 + 60 x .06, or 63.6Ω. Figure 2.2 illustrates a plot of equation 2.1 for a 60Ω bit with MR coefficient of 6%. A simple model of a MR cell would be a resistor in series with a small variable resistor. The ratio of these two resistance's would be a function of the MR coefficient. Figure 2.3 depicts this model for a 60Ω cell with an 6% MR coefficient.

Another key characteristic of the materials used to make magneto-resistive memory cells is known as uniaxial anisotropy. Uniaxial anisotropy is the preference of magnetization to lie along a single axis in one of two antiparallel states. These two states are found along what is termed the *easy* axis. Perpendicular to the easy axis is the hard axis. With no external magnetic fields applied, the magnetization of the bit will always lie in the easy axis. Figure 2.4 illustrates these properties.

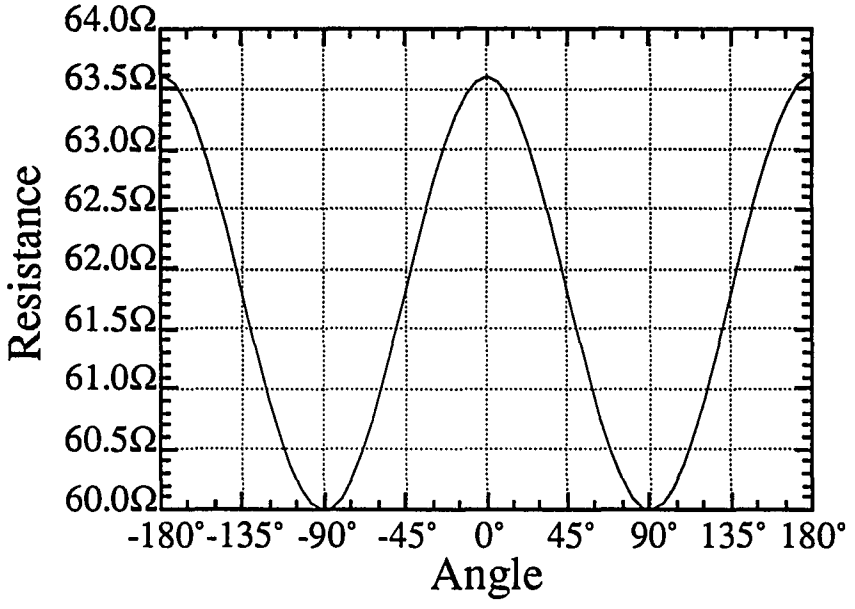

Figure 2.2 Resistance vs. angle between sense current and magnetization

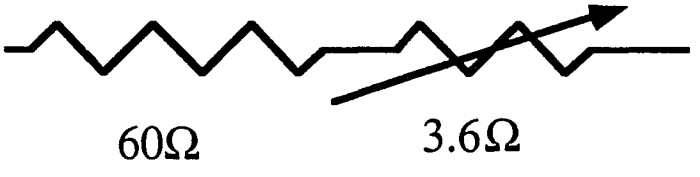

Figure 2.3 Simple MR cell model

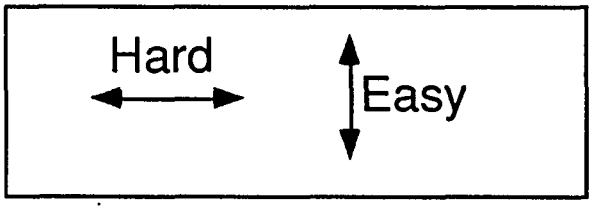

Figure 2.4 Properties of uniaxial anisotropy

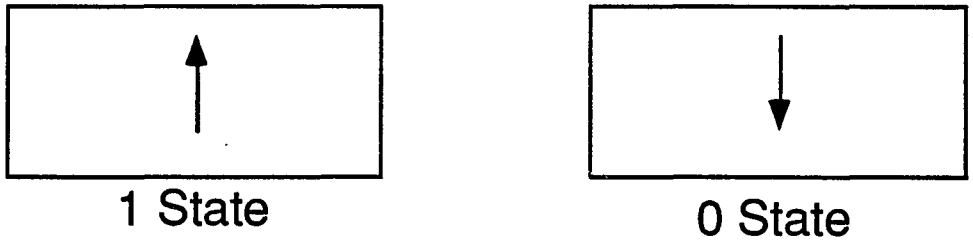

Figure 2.5 Stable states of MR cell and associated binary assignments

Figure 2.6 Gravitational example

The property of uniaxial anisotropy gives rise to the bi-stable nature of an MR cell. With no net fields applied, the cell will have its magnetization either pointing up along the easy axis or pointing down along the easy axis. We can assign the first condition to indicate a one state and the later to indicate a zero state, as shown in Figure 2.5. Figure 2.6 shows a gravitational example of a ball residing in one of two valleys with a mountain in between. One can associate these two stable states with the condition of magnetization being pointed up along the easy axis in one case (1 state) and being pointed down along the easy axis in the other (0 state). Magnetization along the hard axis would correspond to the unstable point on top of the mountain. To change the direction of magnetization, external magnetic fields must be applied.

The modern MR memory element is being manufactured using two 150 Å magnetic layers (Ni 65%, Fe 15%, Co 20%) separated by a 50 Å layer of non-magnetic tantalum oxide.

Much research is being conducted to improve the MR coefficient, and changes in the construction of these magnetic layers are likely. A word line is placed over the top and perpendicular to the length of the MR element. Figure 2.7 illustrates this basic memory element.

Both the sense and the word current produce magnetic fields that interact with the magnetic properties of the bit itself to rotate the direction of magnetization. This rotation causes a change in the resistance due to the magneto-resistive effect. A bit written to the one state exhibits a much greater change in resistance than does a bit written to a zero state. Figure 2.8 illustrates a typical output signal for a 60Ω bit with an MR coefficient of 2%, 3 mA of sense current, and the word field at the bit from -18 to +7 Oersteds. At the present distance, the word line is from the magnetic thin films, 1 mA of word current generates approximately .4 Oersteds of magnetic field. Although these signal levels are small, the elements can be read with a great degree of accuracy through the use of careful sensing techniques. In the reading of a memory element, only reverse word currents are used. Forward word currents are applied to write the bit . Details of the operation of the bit are presented later in this document.

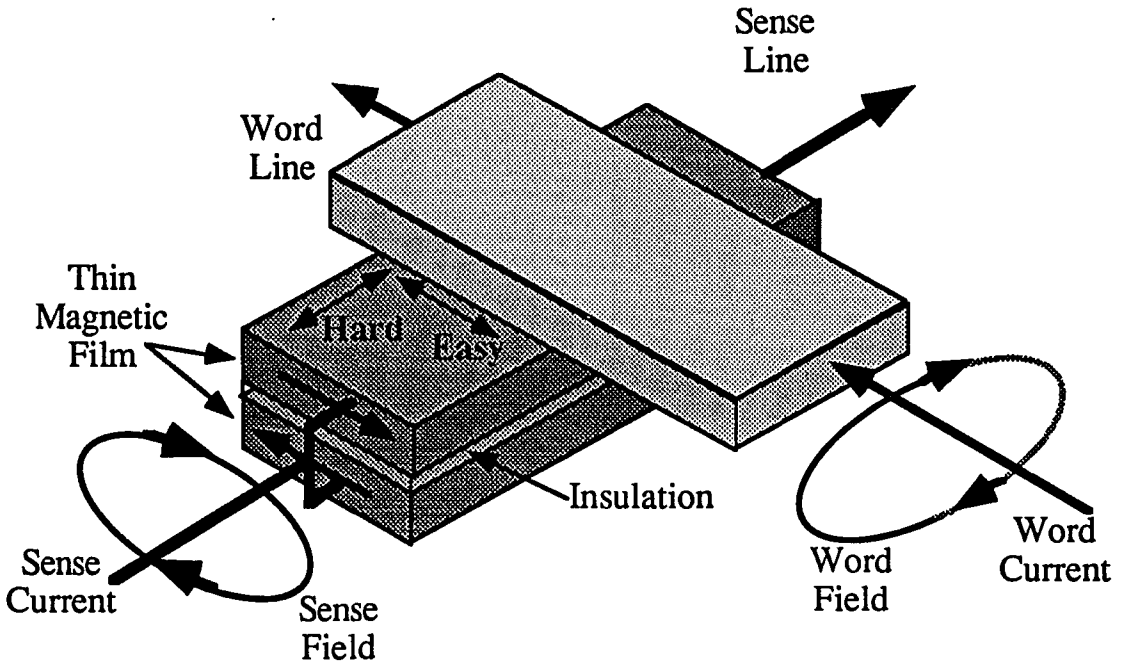

Figure 2.7 Basic MR element

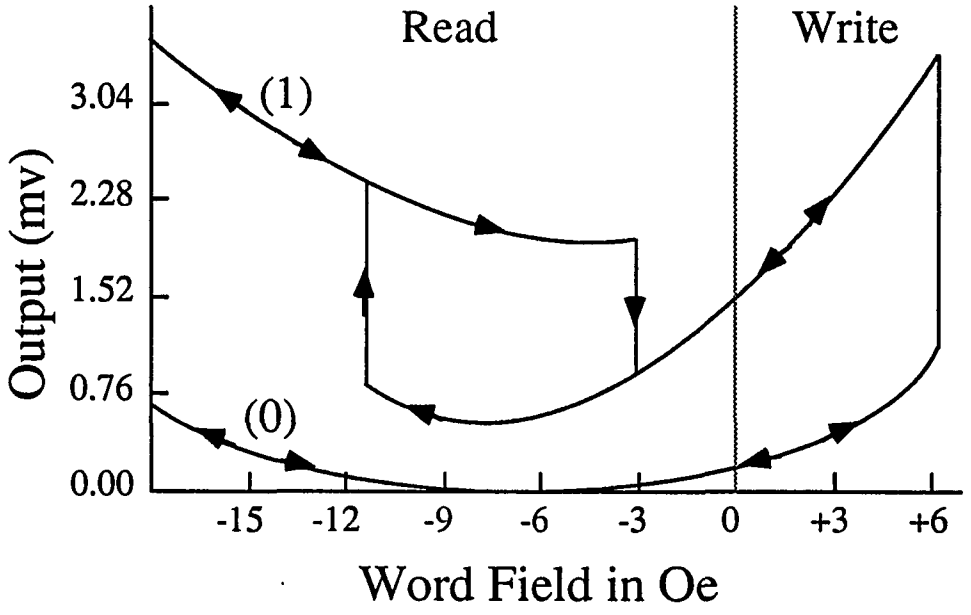

Figure 2.8 Typical MR cell output

3. ONE-MEGABIT DESIGN

3.1 Introduction

This chapter summarizes the one megabit design and provides a framework in which to consider the remaining chapters . Table 3.1 summarizes key design goals for this chip.

3.2 Overview

Figure 3.1 illustrates the overall block diagram of this one-megabit design. The chip is comprised of 33 double chunks, 32 normal chunks and one spare. Each double chunk consists of two chunks, each of which is made up of four data segments. One bit from each of eight active data segments within a double chunk is operated on simultaneously to yield a byte-wide word. Each data segment consists of 256 regular 16-element sense lines. Total number of bits therefore is 16 x 256 x 4 x 2 x 32= 1,048,576, or 1 Megabit.

Address decoding is placed in a center cross shape. Four sets of word drivers are placed on the top and the bottom of the memory, with centrally located word gates that select one of 128 word lines. To reduce switching noise due to the large word currents, a dummy word line is placed on each of the four word driver pairs. This dummy word line is turned on before the word current is needed, and is allowed to stabilize. When the word current is required, the dummy word line is turned off as the selected word line is being turned on, thus limiting unwanted transient effects. Sense line decoders and drivers are placed in the vertical column separating the left and the right half of the chip. Details of these decoders and drivers are discussed in Chapter 5. The final-stage amplifiers contain large capacitors constructed as illustrated in Figure 3.2. Metal 2 and poly will be joined to form one plate of the capacitor, and metal 1 will form the inner plate. Because all three materials are needed to

Table 3.1 Summary of design goals

Memory Size	1 Megabit (128K by 8 bit)
Chip Area	$\approx .5\ cm^2$ @ min. feature of .8μm
% of MR cell area	> 50% of chip area in MR cells
Byte Read Time	< 1 μsec
Yield	90% with bit failure rate of 1/4000

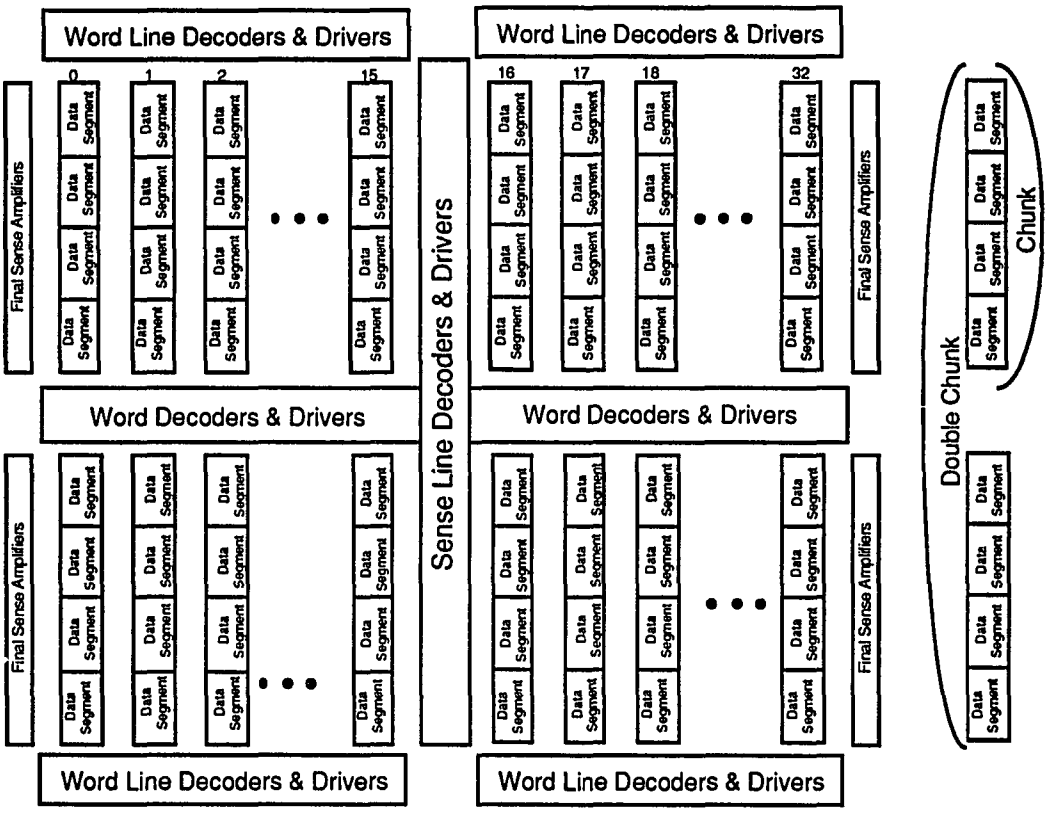

Figure 3.1 Block diagram of one-megabit chip

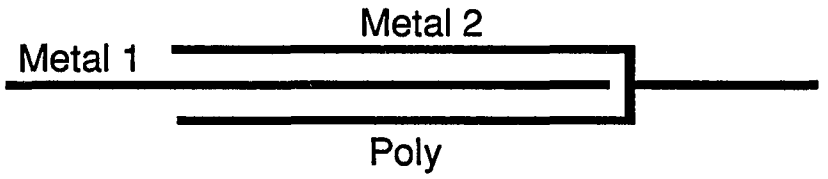

Figure 3.2 Construction of large final sense amplifier capacitors

make this type of capacitor, final sense amps are placed on the outside edges of the memory array, where all three layers are readily available. Originally, the final sense amplifiers were to be placed in the vertical center section, along with the sense line decoders and drivers. But it was necessary to use metal 1 for address lines in this channel. As a result, these capacitors could not be placed in the middle of the chip as originally designed. By placing these amplifiers on the outer edge instead of in the middle, two sets rather than one were needed, and chip area required for these final sense amplifiers doubled. One benefit of the final amplifier placement was its less restrictive shape, whereby these final amplifier circuits could be modified easily to accommodate future changes.

3.3 Data Segment Details

Figure 3.3 illustrates the contents of a data segment, which contains 256 normal sense lines and 8 dummy sense lines, each divided equally into an upper-half and, lower-half segment. In the center of each data segment resides the sense line drivers and the first-stage differential amplifier (pre-amp), as well as the necessary control logic. Figure 3.4 details the contents of a data segment. Because two sense lines share the same select line, two separate sense line drivers and rails are required. While one sense rail is on, the other rail must be grounded, so that no current is allowed to flow through it. Eight dummy lines are required

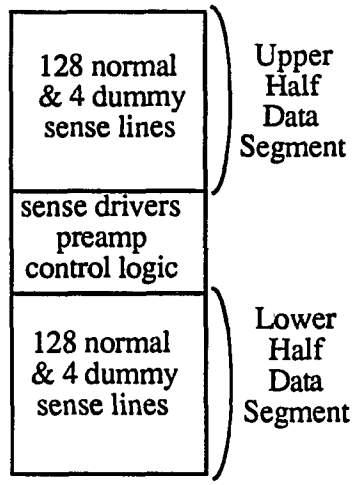

Figure 3.3 Block diagram of data segment contents

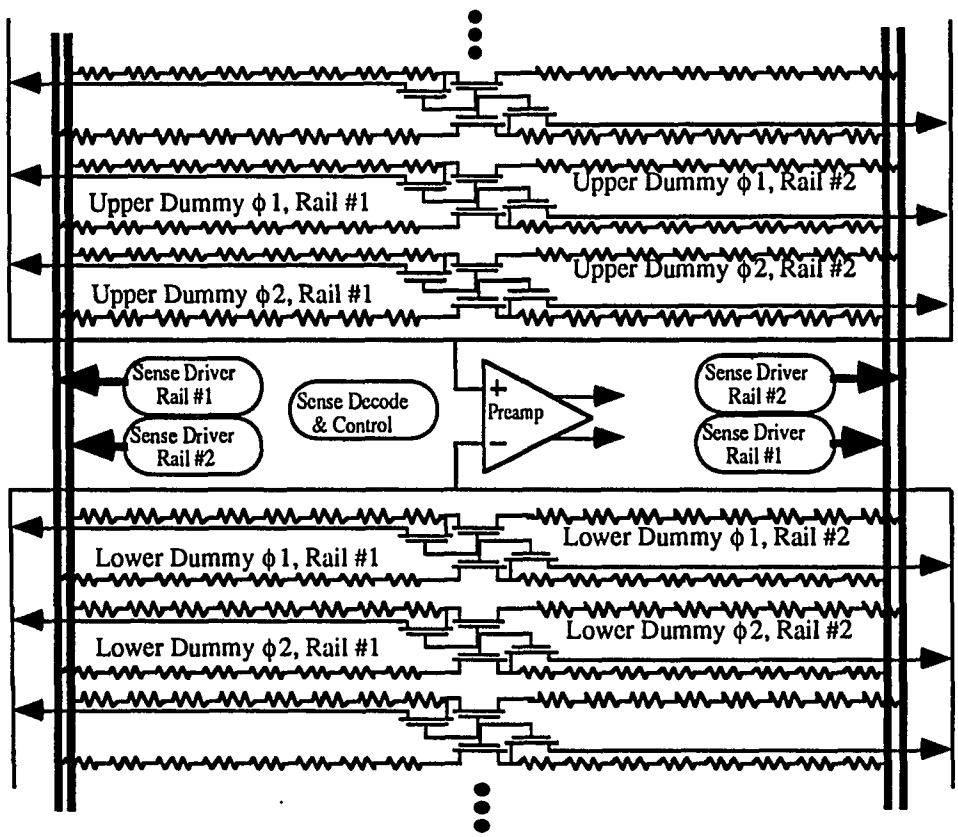

Figure 3.4 Data segment schematic diagram

for proper sensing. Chapter 4 contains detailed information about normal sense and dummy sense line design and operation. Not shown in Figure 3.4 are the word lines running perpendicular to and over the top of the sense lines. The function of a word line is to provide a magnetic field that rotates the magnetization of the MR element. Word lines are made of metal 2, and a single word line services 2K elements along a double chunk, because each word line crosses every sense line once. During a single read cycle, exactly 16 sense lines 8 normal and 8 dummy and 1 word line have current flowing through them.

3.4 MR Element Details

A single MR element is shown in Figure 3.5 This basic magnetic element is 7μm long by 1.4μm wide with a necked shorting bar of .9μm on each end. When packed as tightly as design rules allow, cell area is 24.7μm^2. This bit design is rather conservative. As processing and materials developments occur, it is likely to decrease in size dramatically. The bit is necked to ensure that when an element is written to a different state, the energy in the switching front will be annihilated when entering the tapered end. Thus, elements are prevented from switching back to the original state. The magnetic layers are laid down in the presence of a magnetic field such that the easy axis is perpendicular to the length of the element. At the edge of the bit, the magnetization must bend into the hard axis. Because this could bend in either direction, a strong magnetic field is applied to set all bits magnetization to the same edge direction following manufacturing. Figure 3.6 is a cross -section of an MR element along its long axis. The illustration is roughly to scale. Metal 2 is separated from the bit by 10,000Å of silicon dioxide insulation. Because the field strength is inversely proportional to the distance, large word currents are necessary to provide the desired field at the magnetic layers. When this distance is decreased, future bits will require smaller word

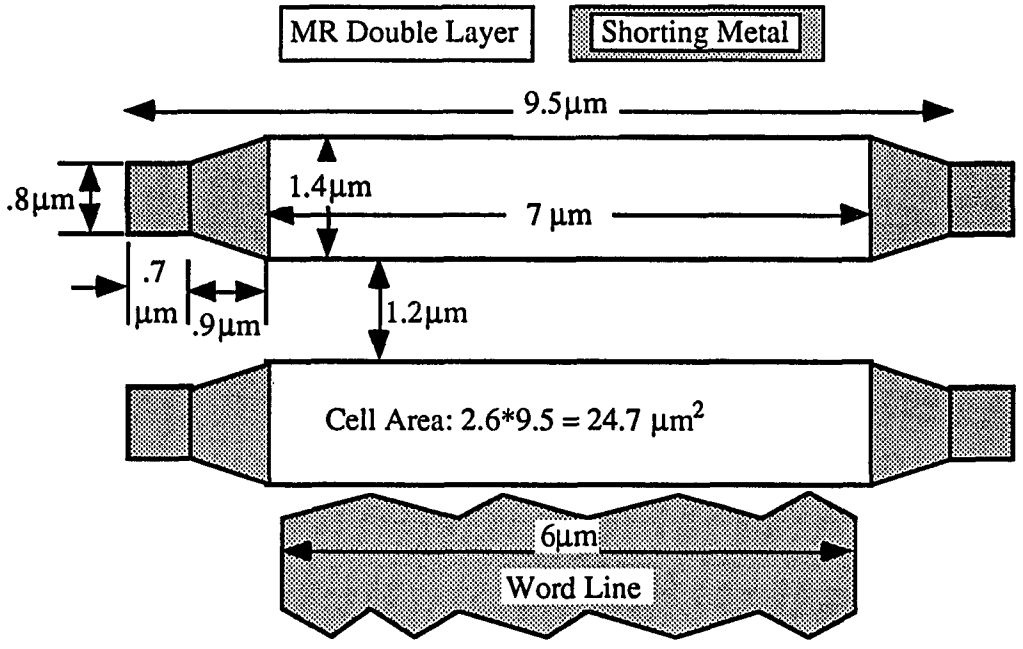

Figure 3.5 MR element details

currents. Figure 3.7 illustrates a close-up of the first-layer metal and gives greater detail of the composition of the magnetic layers.

One of most significant enhancements of this design involves the new two-phase sensing scheme with sample and hold. MR cells have a unique property whereby if the sense current is reversed, the element will act as its own complement. Thus, if a bit was previously written to the one state, during phase 1 of a read, the sense current flows one direction through the bit, and the output signal is that of a one. If during phase 2, however, the sense current through the bit is reversed, the output signal will be that of a zero. The previous 16K MR design used what is known as a ping-pong array. In this type of design two separate arrays are used to store a single bit. One array stores the bit and the other stores the bits complement. When a bit is read, the signals from the bit and its complement are compared to

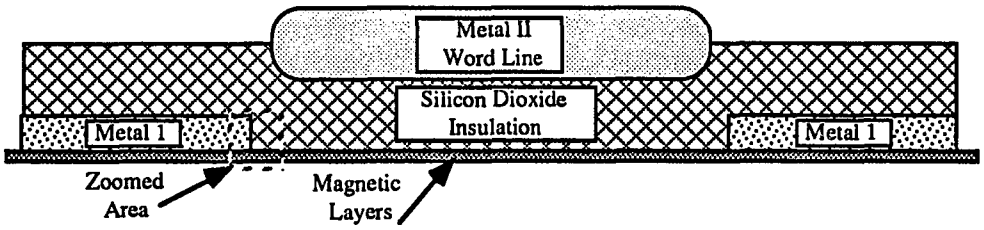

Figure 3.6 Cross section of MR element

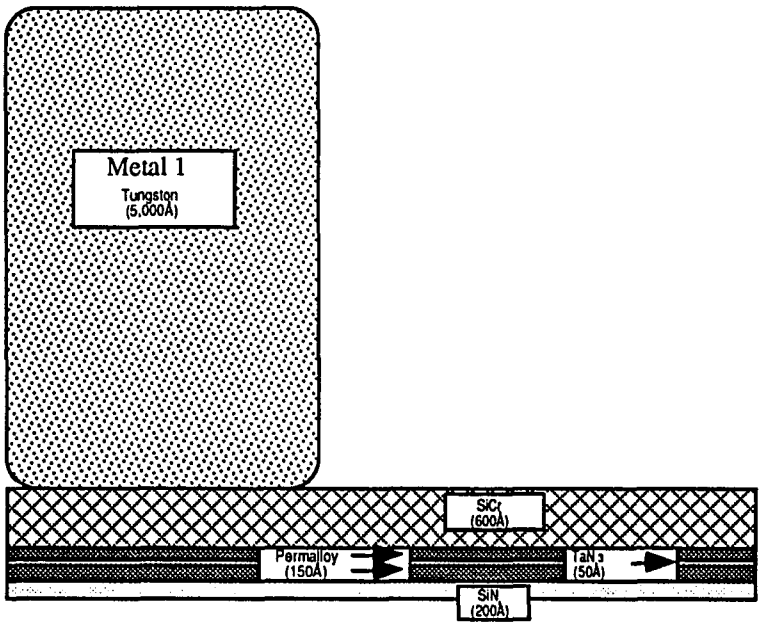

Figure 3.7 Cross-section detail of magnetic layers

yield the correct digital output. This is an effective method for optimizing signal levels as it conceals many common mode disturbances, but is costly in area.

The new design eliminates the need for a second memory array to store the complement of the bit. To read a bit using this new technique, the bit is read for one direction of sense current during phase 1 and stored. During phase 2, the sense current is reversed, and the bit acts like as its complement. This second signal is compared with the first, and the proper digital output results. The signal levels are similar to those of the previous scheme, without twice the memory area. The drawback to this design is a slower read time because two separate accesses (phase 1 & phase 2) are required. Figure 3.8 illustrates the difference between these two sensing schemes. The details of this sensing scheme, including area considerations can be found in the doctoral dissertations of Manel Ranmuthu and Indumini Ranmuthu [9, 10].

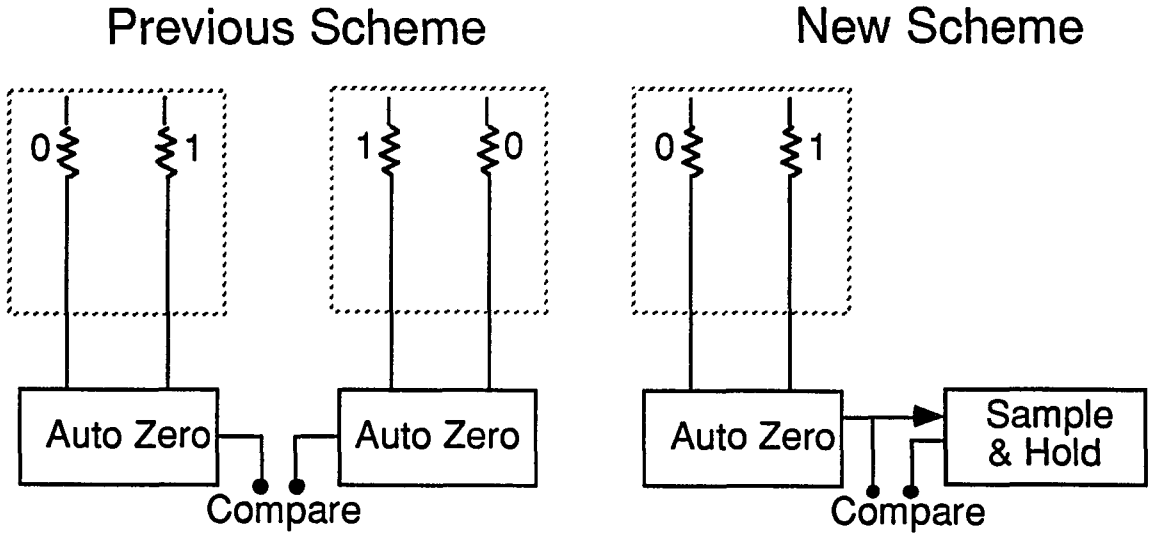

Figure 3.8 Previous and new sensing scheme

4. SENSE LINE DESIGN

4.1 Introduction

Much effort has gone into the design of the sense line which is the fundamental building block of this type of memory. Figure 4.1 is a schematic representation of a sense line, which consists of three main components: gate transistor, mux transistor, and a number of MR memory cells shown as resistors. In this design 16 elements are placed on each sense line with 8 elements on either side of the gate transistor.

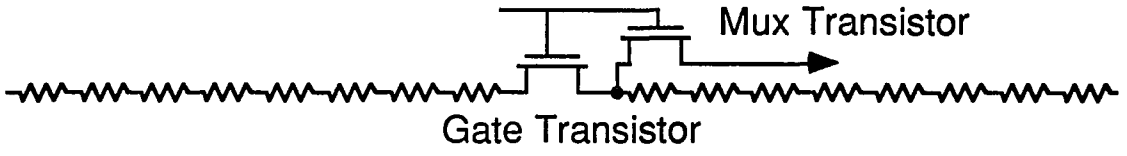

Figure 4.1 Schematic of sense line

4.2 Design Considerations and Constraints

There are Many constraints and limitations on sense line design. Because the sense line occurs thousands of times, it must be implemented in the smallest possible area. The layout architecture must be able to abut neighboring cells without requiring special wiring or adjustment. Access delay and settling time must be reasonably small and must not delay the system unnecessarily. A small voltage drop is desired across the gate transistor to allow the most voltage possible to exist across the MR elements, where it is most useful. The mux transistor should have a low output impedance to improve the signal to noise ratio. The sense line must be designed to accommodate a bi-directional current, which the new two-phase sensing scheme requires. The high current densities in the sense line also must be carefully considered. Reliability and redundancy needs also constrain sense line design.

4.3 Buried Gate and Mux Transistor

One of the most successful features incorporated in this design is the placing of the gate and mux transistors underneath the MR bits themselves. Because the MR cells consist of layers at and above metal 1, the layers underneath them, which comprise all the materials necessary to make transistors, are available for other structures. Thus gate and mux transistors can be placed underneath the MR cells. If these transistors are buried, many benefits are realized, including a dramatic area savings. Only a modest area is needed beyond that required for the bits themselves. Figure 4.2 is a schematic representation of two sense lines which share a common polysilicon gate. Figure 4.3 illustrates a simplified compact layout of the two central gate transistors.

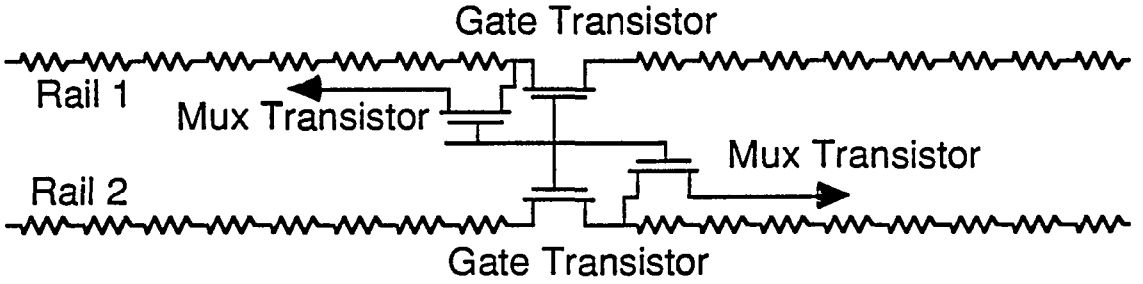

Figure 4.2 Two sense lines with common polysilicon gate

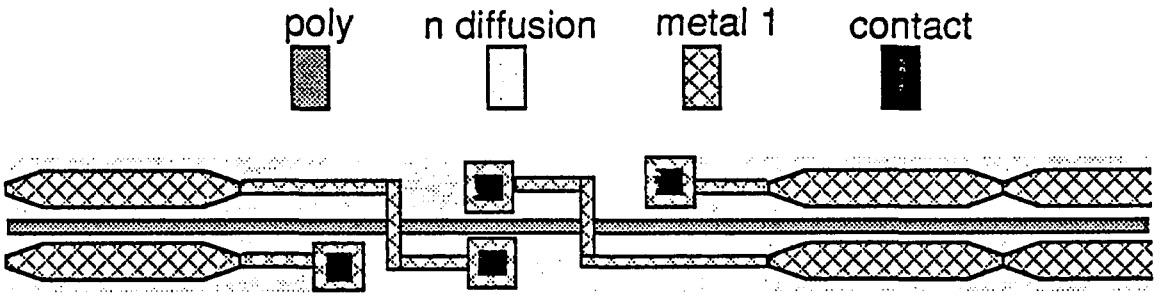

Figure 4.3 Layout view of buried central gate transistors

In addition to savings in area, these transistors can be made relatively large at no added area cost. This change results lowered voltage drop across gates and improved noise performance.

4.4 Spreading Resistance Considerations

When these two transistors are buried, their layouts become unusual and require special attention. Because contacts are unavailable along either side of these transistors, significant spreading resistance results. *Spreading resistance* refers to the inherent sheet resistance of materials used to manufacture integrated circuits. In this instance, the current flowing through the far end of the transistor must first pass through a long resistive drain across the gate and back along a resistive source region before entering the contact and continuing on its path to ground. As a consequence, the transistor shown in Figure 4.4 needs to be modeled accurately. An accurate value of the voltage at the far end of this transistor is needed because the mux transistor attaches there, and the voltage is next gated to the first-stage pre-amp for sensing. The current-carrying capabilities of the gate transistor are reduced as a result of the spreading resistance, and the consequences of these effects need to be understood.

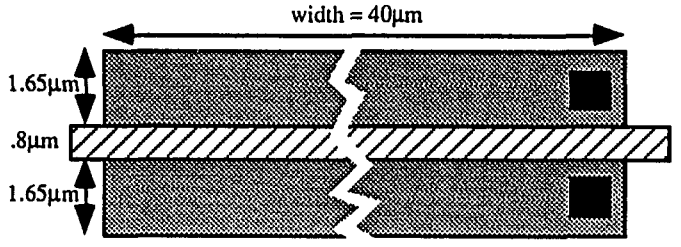

Figure 4.4 Transistor to model

4.4.1 SPICE simulations

A common approach to solving this problem is to use a SPICE simulation of the transistor as several segments, with series resistance due to diffusion sheet resistance taken into account. Figure 4.5 is a schematic using this approach with a five-transistor model in which the sum of the transistor widths is equal to the total transistor width. Two SPICE simulations were performed, one using a five-transistor model, and the other using a ten-transistor model. With an N diffusion sheet resistance of 3.03 Ω/square, the results of Table 4.1 were obtained and graphed in Figure 4.6.

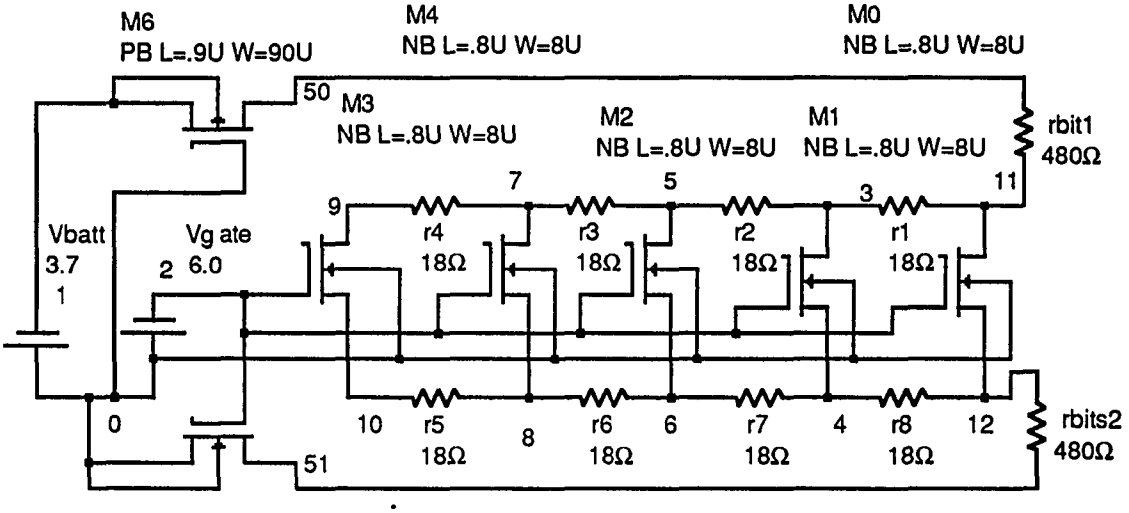

Figure 4.5 SPICE modeling by segmentation

Table 4.1 SPICE modeling table of results

Transistor #	Five Transistor Model Current in mA	Ten Transistor Model Current in mA
1	1.0625	0.585386
2	0.75595	0.498008
3	0.55780	0.426087
4	0.439632	0.367389
5	0.3845	0.320094
6		0.282733
7		0.254148
8		0.233451
9		0.22
10		0.213377
Total	3.200382	3.400673
V(mux)	1.8435 Volts	1.9718 Volts

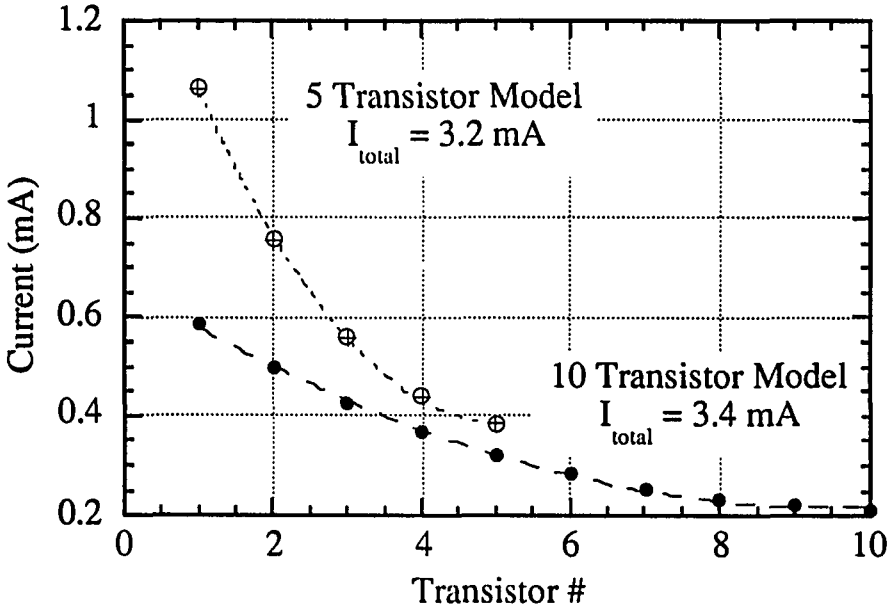

Figure 4.6 SPICE segmentation modeling results

The important aspect of these simulation results is the marked decrease in the current capacity of those transistors farthest from the contacts located on the end of the original transistor.

4.4.2 Resistive ladder approach

A second, more flexible, method of solving the spreading resistance problem is use of a resistive ladder network and differential equations. The gate transistor resides midway between the driving rails, with an equal number of resistive elements on either side. As a result, the gate transistor will always operate in the ohmic region. The second-order equation governing its current is given in equation 4.1. Although this simplified equation does not take into account many complex phenomena known from device physics, it is suitable for improving understanding of the effects of spreading resistance.

$$I_{DS} = K_N(\frac{W}{L})\left[(V_{GS} - V_t)V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{4.1}$$

The resistance per unit length can be found by dividing the voltage by the current to obtain the resistance of the transistor gate. Equation 4.2 gives this result.

$$r_p = \frac{V_{DS}}{I_{DS}} = \frac{1}{K_N(\frac{W}{L})\left[(V_{GS} - V_t) - \frac{V_{DS}}{2}\right]} \tag{4.2}$$

If a differential equation is used to solve the resistive ladder network shown in Figure 4.7, the solution is of the form:

$$\frac{\partial V}{\partial x} = -r_s I \qquad \frac{\partial I}{\partial x} = -\frac{V}{r_p} \tag{4.3}$$

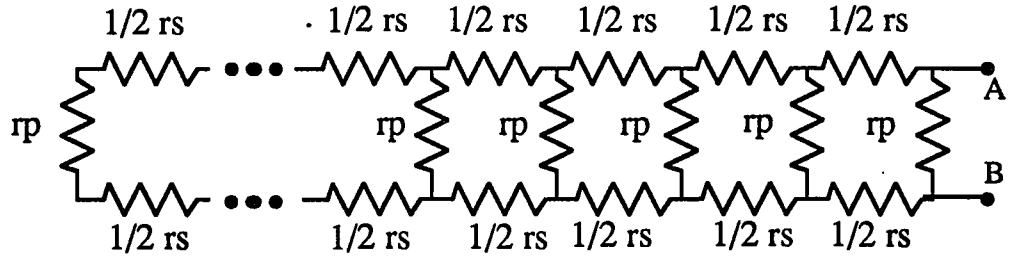

Figure 4.7 Resistive ladder network

Written in operator form, equations 4.3 become:

$$DV + r_s I = 0 \qquad DI + \frac{V}{r_p} = 0 \tag{4.4}$$

By multiplying the first equation of 4.4 by $1/r_p$ and the second by D and subtracting the two, equation 4.5 results.

$$-D^2 I + \frac{r_s}{r_p} I = 0 \tag{4.5}$$

The characteristic equation is therefore:

$$I(-D^2 + \frac{r_s}{r_p}) = 0 \text{ , where } D = \pm\sqrt{\frac{r_s}{r_p}} \quad \text{and} \quad \alpha = \sqrt{\frac{r_s}{r_p}} \tag{4.6}$$

The form of the solution is:

$$I(x) = C_1 e^{\alpha x} + C_2 e^{-\alpha x} \qquad V(x) = C_3 e^{\alpha x} + C_4 e^{-\alpha x} \tag{4.7}$$

By substituting into equation 4.3a, he form of the solution is given in (4.8).

$$\alpha C_3 e^{\alpha x} - \alpha C_4 e^{-\alpha x} = -r_s C_1 e^{\alpha x} - r_s C_2 e^{-\alpha x} \tag{4.8}$$

Because equation 4.8 must hold for all values of x, C_3 and C_4 can be solved for in terms of C_1 and C_2, as given in equations (4.9)

$$C_3 = -\frac{r_s C_1}{\alpha} \qquad C_4 = \frac{r_s C_2}{\alpha} \tag{4.9}$$

The boundary condition for the current at width W is zero because no current is present at the very end of the transistor. This leads to equation 4.10.

$$I(W) = 0 = C_1 e^{\alpha W} + C_2 e^{-\alpha W} \Rightarrow C_1 = -C_2 e^{-2\alpha W} \tag{4.10}$$

Substituting the results of equation 4.10 into 4.7 results in equation 4.11.

$$I(x) = C_2(e^{-\alpha x} - e^{\alpha(-2W+x)}) \quad V(x) = \frac{C_2 r_s}{\alpha}(e^{-\alpha x} + e^{\alpha(-2W+x)}) \tag{4.11}$$

Solving for r_{eff} yields equation 4.12.

$$\frac{V(x)}{I(x)} = r_{eff} = \frac{\sqrt{r_s r_p}(e^{-\alpha x} + e^{\alpha(-2W+x)})}{(e^{-\alpha x} - e^{\alpha(-2W+x)})} \quad \text{where } \alpha = \sqrt{\frac{r_s}{r_p}} \tag{4.12}$$

Solving for x=0, (at the terminal A & B) equation (4.13) results.

$$\frac{V}{I} = r_{eff} = \frac{\sqrt{r_s r_p}(1 + e^{\alpha(-2W)})}{(1 - e^{\alpha(-2W)})} \quad \text{where } \alpha = \sqrt{\frac{r_s}{r_p}} \tag{4.13}$$

This equation has several interesting consequences. As expected, there comes a point after which the effective resistance is not significantly reduced by continuing to add transistor width. By allowing the width (W) to approach ∞, the minimum resistance approaches $\sqrt{r_s r_p}$. Figure 4.8 illustrates a plot of equation 4.12, where $\sqrt{r_s r_p}$ is set equal to 1, and the effective resistance is computed versus the quantity αW. From this plot, it is obvious that the law of diminishing returns is at work. When α W is greater than 1.5, negligible benefit is realized from adding transistor width.

By means of SPICE simulations, the estimates V_{GS}=4.26V and V_{DS}=.174V were made. From these values, parallel resistance is found to be 3055Ω. This can be seen in equation 4.14.

$$r_p = \frac{V_{DS}}{I_{DS}} = \frac{1}{81.49x10^{-6}(\frac{1}{.8})\left[(4.26-.96) - \frac{.174}{2}\right]} = 3055\Omega \tag{4.14}$$

The series resistance is found using 3.03Ω/square (Silisided source and drain lower sheet resistance significantly), and the height of the source and drain diffusion is 1.65μm. Therefore source and drain series resistance per micrometer is:

$$2\left(\frac{3.03}{1.65}\right) = 3.673\ \Omega/\mu m \tag{4.15}$$

Alpha would therefore be $\alpha = \sqrt{\frac{3.68}{3055}} = 0.03471$, and the length of the device is L = 40μm.

The final effective resistance shown in equation 4.16 is 120Ω.

$$\frac{V}{I} = r_{eff} = \frac{r_s(1+e^{\alpha(-2L)})}{\alpha(1-e^{\alpha(-2L)})} = \frac{3.68(1+.06225)}{.03471(1-.06225)} = 120\Omega \qquad (4.16)$$

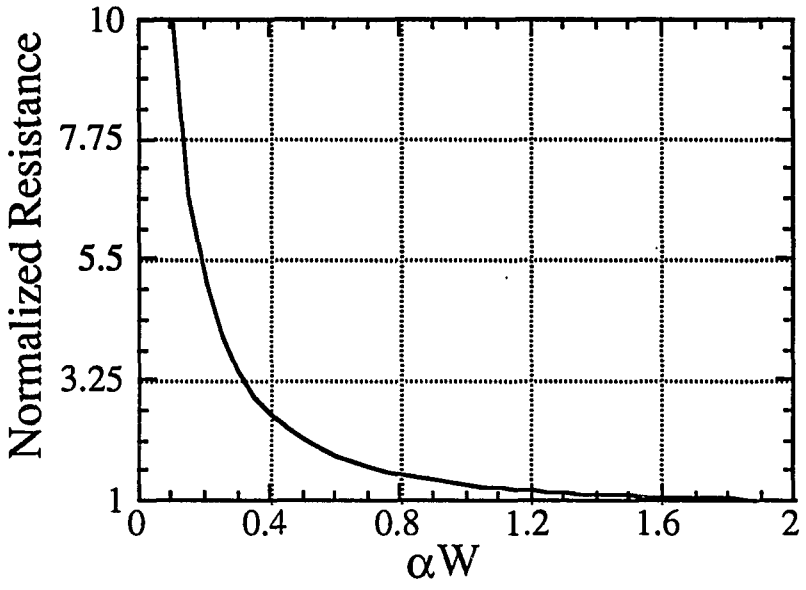

Figure 4.8 Normalized resistance vs. αW

4.5 Current Density Restrictions

This design requires 3mA of sense current. The current density in a single 1μm x 1μm contact is $3x10^9$ Amps/m^2. Process parameters differ from one vender to the next, but it is likely that a second contact will need to be added to handle the high current requirements. The existing 16K design also requires approximately 3mA of sense current and is implemented with only one contact, it is however 1.4μm x 1.4μm in size. The current densities in the magnetic layers are even higher. The 3mA sense current must flow through the two 150Å thick by 1.4 μm wide permalloy layers (see Figures 3.5 & 3.7), resulting in current densities as high as $7.15x10^{10}$ Amp/m^2. Fortunately, neither tungsten (metal 1) nor

the permalloy layers experience metal migration. The only well-known failure mechanism that these materials evidence is melting.

4.6 Heating and Cooling Effects

The large currents that flow through the sense line causes heating in MR cells. Because these elements have a large temperature coefficient of resistance, the desired change in resistance due to magneto-resistance can be masked out easily by the change of resistance due to change in temperature. Thus, careful consideration must be paid to ensure that the temperature of the bit being sensed is similar to that of the dummy with which its signal is being compared. These elements heat up and cool down exponentially and behave similarly to an electrical RC circuit. In much the same way as an RC circuit *remembers* its past inputs, these bits *remember* their past heating conditions. This temperature *memory* is referred to as thermal history. A paper entitled "Temperature Transient Effects in M-R Memory Elements" by A. V. Pohm elaborates these heating and cooling effects [11].

4.7 Number of Elements on Each Sense Line

To increase density and to reduce support electronics overhead, it is desirable to have as many elements as possible on a single sense line. The two main factors affecting this number are the MR coefficient and the magnetic materials resistance. The present design has 16, 60Ω elements on each sense line. At 25° C, these elements have a total combined resistance of 960Ω for all 16 elements. If, for instance, the MR coefficient improved so that the signal level doubled, the number of elements could be increased to 32 while their resistance could be decreased to 30Ω, while maintaining the total 960Ω. If this were done, the same signals would result and similar performances would be expected, but higher densities would be achieved.

4.8 Dummy Sense Line Design

Because the signal levels are small, and various operating conditions affect the absolute signal level, dummy sense lines are needed to compare the bit being sensed. Dummy sense lines Should be as much like ordinary sense lines as possible, so as to maintain close matching. Temperature change is also key, and timing must be considered to ensure similar thermal histories. To meet these demands, timing circuits may need to be lengthened to allow for sufficient cool-down periods [11].

Several designs were considered, and the most promising is described here. Figure 3.4 shows the schematic details of these dummy sense lines. For ease in layout and matching purposes, the dummy lines are identical to ordinary sense lines, with the exception of decoding and drive circuits. Four dummy lines are required in the single-redundancy design. To match impedance and capacitive loading of the pre-amps, a dummy in the opposite upper or lower half segment is activated to be compared with the desired cell. Because it is always desirable to compare with a stored zero, the dummies must be in the zero state. With the dual-phase sensing scheme, two separate dummy lines are needed to sense one selected bit, one which is zero for one polarity of sense current, and one which is zero for the opposite polarity of sense current [9,10].

This implementation yields an inherent thermal history problem. During the first phase of sensing, both the selected bit and the dummy bit are likely to be cool, but during the second phase, the selected bit is hotter than the second phase dummy. As a result of this inequity, sufficient cool-down time of the selected bit and/or preheat of the second phase dummy is necessary to match thermal histories.

5. SUPPORT LOGIC DESIGN

5.1 Introduction

This chapter deals with digital design and layout of necessary support logic. These circuits include decoders for sense lines and word lines, as well as decoders and control logic for the dummy sense lines.

5.2 Primary Sense Line Decoders and Drivers

Several sense line decoding designs were considered. Originally planned fully static decoding was to be used to ease testing and to eliminate dynamic decoding timing considerations. Several narrow fully static layouts were designed, but none was sufficiently narrow to abut the sense lines. Stacking of several decoders was attempted, but area increased dramatically. With these considerations in mind, a dynamic decoder was designed and laid out which was narrow enough to abut to the sense lines directly. This dynamic decoder required much less than half the area of the fully static design with little reduction in performance. A narrow drive inverter also was designed to abut the sense line. The next section discusses proper sizing and spacing of sense line gate drivers and re-powering circuits. Table 5.1 list key process and design parameters.

Table 5.1 Process and design parameters

Parameter	Value
Sheet resistance of Polysilicon gate	3.52Ω/square
Thickness of gate oxide (t_{ox})	170Å
NMOS gate length	.8μm
Gate width per data segment	200μm

The gate capacitance (C_{OX}) is calculated as shown in equation 5.1

$$C_{ox} = \frac{\varepsilon_o \varepsilon_{SIO_2}}{t_{ox}} = \frac{8.854\text{x}10^{-12}\frac{F}{m} * 3.9}{170\text{x}10^{-10}\,m} = 2.0312\text{x}10^{-3}\frac{F}{m^2} = 2.0312\ \frac{fF}{\mu m^2} \qquad (5.1)$$

The gate capacitance of one sense line is therefore:

$$C_g = \text{Area} * C_{ox} = (0.8\ \mu m)(200\mu m)(2.0312\ \frac{fF}{\mu m^2}) = 325\ fF \qquad (5.2)$$

The resistance of one sense line gate is :

$$R_g = \text{Squares} * \frac{\Omega}{\text{Square}} = \frac{200\mu m}{.8\mu m} * \frac{3.53\Omega}{\text{Square}} = 882.5\Omega \qquad (5.3)$$

The central driver needs to drive 16 sense line gates in series, if lumped, the total would become:

$$16 * 882.5 = 14.12K\Omega \qquad (5.4)$$

$$16 * 325 = 5200\ fF \qquad (5.5)$$

The resulting RC time constant would be 73.42 ns, which is unacceptably slow. The lumped model in this instance is overly pessimistic and yields inaccurate results. This is especially true in that the output signal cannot be sensed until the sense line has reached 99.99% of its final value. This long settling time is required to satisfy the signal-to-noise requirements of the sensing circuitry.

A SPICE simulation of a lumped RC circuit and the same RC distributed into 16 segments were compared. The distributed RC reached the 99.99% level 2.28 times sooner than the lumped circuit did. To achieve 99.99% of the final voltage, -ln(1-.9999) = 9.21 time constants are needed. Taking the 2.28 result from the distributed RC line, the net effect would be the same as waiting 9.21/2.28 = 4.0 time constants using the lumped value. Figure 5.2 illustrates four different re-powering strategies, and Table 5.2 lists the results of using each strategy.

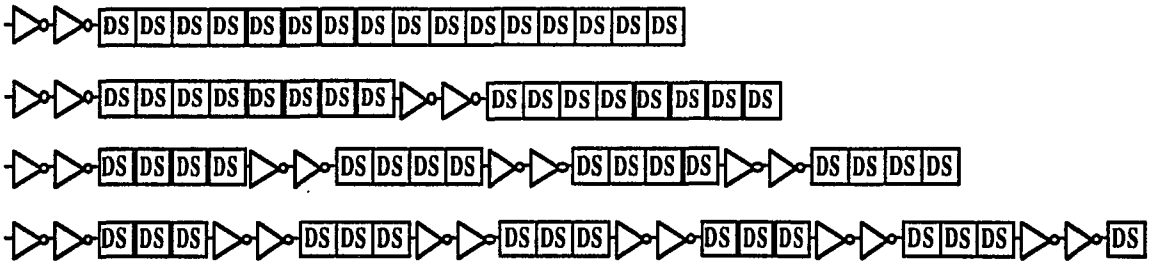

Figure 5.1 Several re-powering alternatives

Table 5.2 Results of various re-powering strategies

Number of Sense lines before re-powering	Total Lumped Resistance per Driver	Total Lumped Capacitance per Driver	4 Time Constants of one Driver	Number of Drivers Required	Peak Current Requirement of Driver
16	14120 Ω	5200 fF	293.7 ns	1	0.4 mA
8	7060Ω	2600 fF	73.4 ns	2	0.8 mA
4	3530 Ω	1300 fF	18.4 ns	4	1.7 mA
3	2648 Ω	975 fF	10.3 ns	6	2.3 mA
2	1765 Ω	650 fF	4.6 ns	8	3.4 mA
1	883 Ω	325 fF	1.15 ns	16	6.8 mA

Peak current requirement is found by:

$$I(t) = \frac{V}{R} e^{\left(\frac{-t}{RC}\right)} \Rightarrow \text{ for } t = 0 \Rightarrow I(\text{Max}) = \frac{V}{R} \tag{5.6}$$

For this design, re-powering every four data segments is sufficient to meet the timing requirements. Furthermore, this design adds the less area than does any faster solution.

5.3 Decoding for Dummies

Special decoding circuits are required with which to select the appropriate pair of dummy sense lines. The key is to determine if the bit being sensed is in the upper or lower half data segment. Once this is determined, the dummy lines located in the opposite half data

segment on the same rail will be accessed in sequence. Four sets of dummies are required to improve the impedance matching on both inputs of the first stage pre-amp [9,10]. With four sets of dummies, nearly identical matching is obtained, and little difference in signal levels will be seen between a data element and a dummy element. Because the selected bit is always to be compared with a zero, four dummy sense lines are required in each half data-segment. One pair is needed for each drive rail, and within a pair one is needed for use during phase one and the other for use during phase two.

By carefully assigning the normal decoding, a single address line should be sufficient to determine whether a bit lies in the upper-half or in the lower-half data segment. An enable line is unnecessary because only one sense driver is on at any given time along a given dummy select line. Two digital signals are necessary to implement the dummy decoding circuit. A signal indicating whether the bit being selected is in the upper-half or in the lower-half of the selected data segment and a signal indicating which phase of the sensing cycle is being performed are also necessary. The phase signal is timed carefully to ensure proper operation and to implement all necessary preheat and/or cool-down periods required to ensure matching thermal history. This decoding and driving circuitry is placed in the middle of the chip and is similar to normal sense line decoders and drivers. Figure 5.2 illustrates this dummy decoding logic.

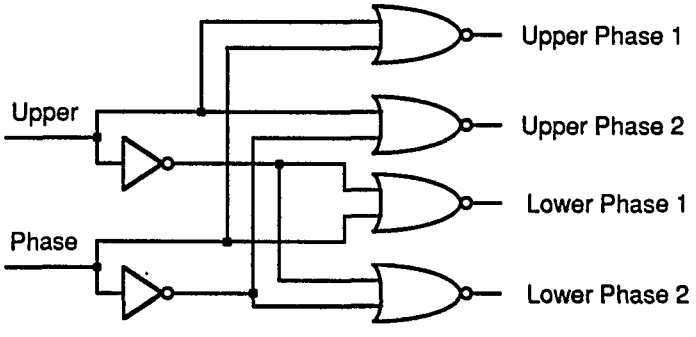

Figure 5.2 Dummy decoding logic

5.4 Word Gate Decoders and Drivers

Word gate decoders and drivers are similar to sense line decoders and drivers, with several minor exceptions. Seven bits of full decoding are necessary to select one of 128 on a given word driver pair. A controlling enable signal also is needed to ensure that proper timing considerations are met. These word decoder and driving circuits lie in a horizontal line dividing the top from the bottom half of the chip. Because of the circuits location, a wide metal 2 word line was placed over their top. This precluded any use of metal 2 in the word decoder and driver layouts. The drive capabilities are not as demanding as those for the sense line drivers. The word gates themselves have a W/L of 90/.8 (113 squares), resulting in a resistance of 400Ω and a gate capacitance of 147fF. The RC time constant (.006 ns) in this instance is not a limiting factor, nor are the current capabilities of the driver. A minimum size inverter drives this load in less than 1 ns.

6. YIELD ENHANCING TECHNIQUES

6.1 Introduction

The manufacturing process by which these circuits are being built is undergoing major modifications and likely will continue to change for some time. Yields are below product goals. Thus, it is desirable to build components into the architecture and the circuitry so as to reduce or to eliminate the ill effects of failed parts. These approaches are of critical importance and certainly will be used in the future, for no competitive manufacturing process is ever defect free. This chapter deals with several techniques to enhance yield.

6.2 Dual Redundant Elements

One of the most significant methods used to improve the yield of MR memories is the use of dual redundant elements. This method uses two elements rather than one to store a single data bit. A sense line still consists of 16 MR elements but only stores 8 data bits. For this technique to work, sensing circuits must be designed so that signal levels from a single functioning element are large enough to yield the correct result. Because the original sensing circuitry was designed for signals from a single MR element, no modification of these circuits is necessary to accommodate the dual redundant method. If dual redundant elements are used, one of the two selected bits could fail to function, and the memory would continue to work properly. For a bit to fail, both MR elements storing that bit must fail simultaneously, which occurs rarely. If the error rate is one bad bit in 4,000, and a random distribution is assumed, then the chance that one element fails is given in equation 6.1.

$$\text{Probability of a single element failure} = \left\{ \frac{1}{4000} \right\} \tag{6.1}$$

The chance that the two elements which store the information for a single bit both fail is given in equation 6.2.

$$\text{Probability of a dual redundant bit failure} = \left\{ \left(\frac{1}{4000}\right)^2 \right\} = 6.25 \times 10^{-8} \qquad (6.2)$$

For a data segment of 4096 elements (2K bits) the probability that a dual redundant data segment would be fully functional is given in equation 6.3.

$$P\{\text{failure free data segment}\} = \left[1-\left(\frac{1}{4000}\right)^2\right]^{2048} \approx \left[1-2048*\left(\frac{1}{4000}\right)^2\right] = 0.999872 \qquad (6.3)$$

This technique will be used in conjunction with other sparing methods to increase yield dramatically. If one spare sense line per data segment is used, then if one bit (two elements) fails, a spare sense line will be substituted. For a data segment of 4096 elements (2K bits), the probability that a dual redundant data segment with one spare sense line would be fully functional is given in equation 6.4.

$$P\{\text{failure free data segment}\} = 1-\left[(2048)\left(\frac{2047}{2}\right)\left(\frac{1}{4000}\right)^4\right] = 0.999999991 \qquad (6.4)$$

Because there are 512 data segments in a one-megabit chip, overall chip yield is given in equation 6.5.

$$P\{\text{failure free 1 Megabit chip}\} = \left[1-(2048)\left(\frac{2047}{2}\right)\left(\frac{1}{4000}\right)^4\right]^{512} = 0.9999958 \qquad (6.5)$$

This technique is effectively eliminates the negative impact of failures ,even when the failure rate is higher than one bad bit in 4,000. If these same calculations are made for a failure rate of one bad bit in 512, the yields is 0.9845037136. When the bit failure rate is 1 bad bit in 256, a dual redundant system will one spare sense line per data segment would yield of 0.7788483497, still acceptable by today's standards.

An obvious drawback to dual redundant elements is the reduction in half of the memories' size. One positive consequence of employing dual redundant elements manifests itself as defect density diminishes. If the failure rate decreases to the point at which sparing techniques are sufficient to compensate for single element failures, then memory speed could

be increased substantially because the signal levels would be twice as large. As a result, it would be advantageous to be able to adjust timing of the sensing circuits easily such that they could accommodate this change in signal level and allow the memory to operate much faster.

Figure 6.1 shows a dual redundant schematic, and Figure 6.2 the layout for both a single and a dual redundant sense line. Only a minor change in the Metal I layer distinguishes the two. This small change allows for a single metal mask change to convert from the dual redundant layout to a single redundant layout. Drive rail and selection circuits also would need modification to convert between these two schemes.

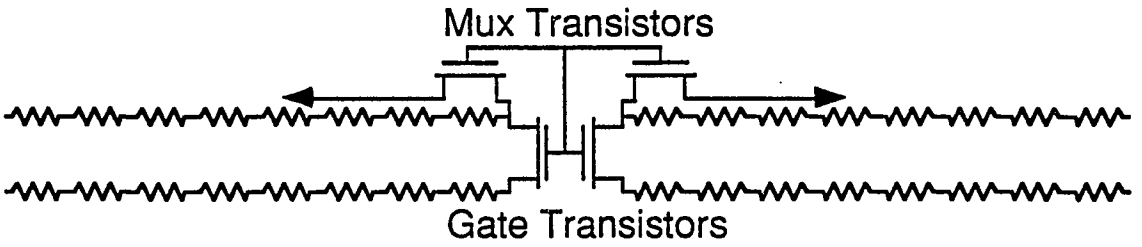

Figure 6.1 Dual redundant sense line schematic

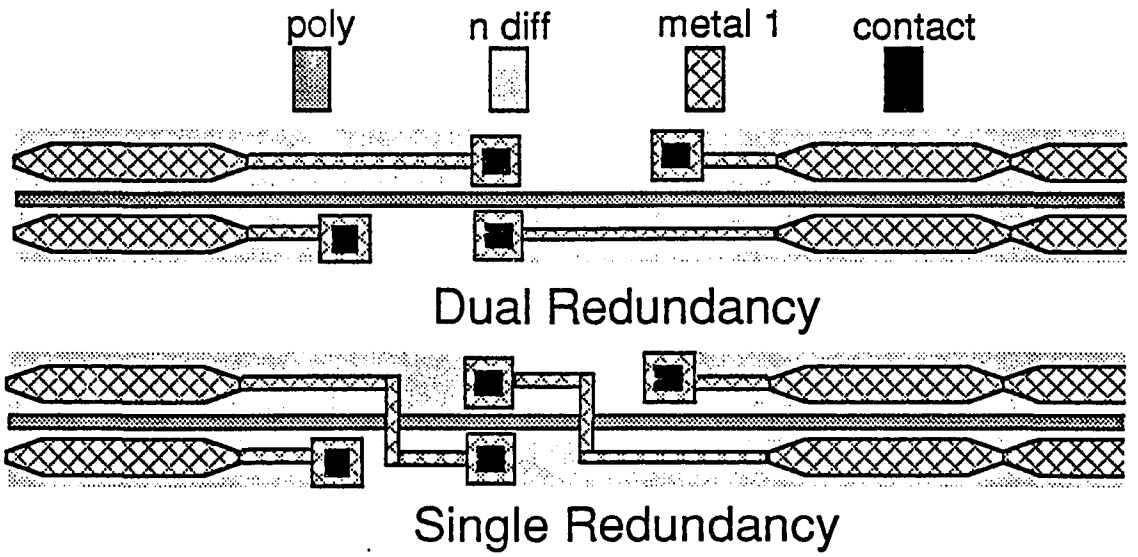

Figure 6.2 Approximate layouts of single and dual redundant sense lines

6.3 Spare Replacements and Fusing Considerations

Most modern mass memories on the market today benefit from the use of redundant row and column elements. These redundant parts are available for sparing or replacing defective areas of the chip. The result is a much higher yield than could otherwise be achieved. A key issue involves the method by which the bad parts are disconnected or ignored and how the good replacement parts are activated. The next sections discuss traditional laser fusing and a new method taking advantage of the nonvolatile nature of MRAM.

6.3.1 Laser fuses

Laser fusing is a common method used to break electrical connections on VLSI chips. Special areas on the chip are laid out in such a way that if a high-intensity laser is directed at them, the polysilicon link is broken. Laser fusing has the advantages of requiring little chip area, and of being reliable. It is costly, however because the laser fusing system is mechanical in nature, the speed by which fuses can be blown is less than that of electronic methods [4].

6.3.2 All-electronic solution

Although laser fusing is widely used, it can add a significant cost to the overall chip if many fuses must be blown. For large memories, this is likely, especially since the manufacturing process is still undergoing development and error rates for MR memory parts are great. Consequently a completely electronic fuse is desirable. Because MRAM is nonvolatile, information containing the address of a bad sense line could be stored in certain special high-speed MR elements. These elements would be written only once during device

testing. Because they do not wear out, they would remain for the life of the chip. A block diagram of such a system is presented in Figure 6.3.

In this system, either at power-up or when the input address points to a different region in memory, a high-speed bad address would be read and stored in a bad address register. This address would then be compared with all incoming address requests, and if a match occurs, indicating that the original sense line was defective, a disable signal would be sent to the standard decoder, and a spare sense line would be activated. This technique also allows for a flexible implementation. If a high-speed memory is desired, more comparator and bad address register circuits could be placed on the chip so that all or most of the bad addresses could be ready and available after power-up, with no or little time penalty during normal use. If, on the other hand, the speed requirements of the memory were not as great, area could be saved by implementing one comparator and bad address register circuit per output bit. Although this would result in a denser chip, it would increase read access time.

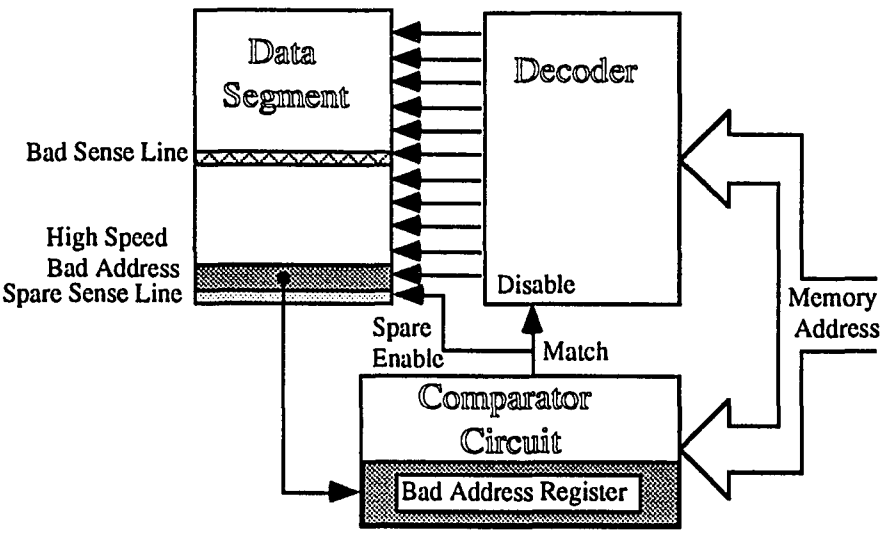

Figure 6.3 Block diagram of an all electronic sparing technique

6.4 Error correction and detection techniques

To achieve a high degree of fault tolerance, error correction circuitry can be used in conjunction with the previously mentioned redundancy plan [12]. Because of the nature of error correcting systems, they are inefficient with small data sets (<8 bytes) [13,14]. As the size of data increases, however, these techniques become more practical to implement. For this reason, a block organization has been proposed to support error correcting techniques. Recent research published by Hitachi has demonstrated a block-oriented DRAM. This paper adds validity to the usefulness of block memory organizations [15]. One also can argue that in many contemporary main memory applications, the smallest granularity of data moved in main memory is a cache-size block.

This proposed design would be randomly accessible at the block level (8 or 16 bytes plus error correcting bytes), but within a block, all bytes must be written to or read from, in order. At the end of every block of data, a Single Bit Error Correction/Double Bit Error Detection (SBEC/DBED) byte would be stored [13]. Table 6.1 presents appropriate error correcting techniques under differing conditions.

Table 6.1 Various yield enhancing schemes

Bit Error Rate	Large Memory Application	Smaller Memory Application
1 bad bit in 1000	•16 Byte Block ->DBEC/TBED •Parallel Error Correcting Circuitry Off Chip organized 18 Chips Wide •1 Block per Memory Cycle •1 Electronic Fuse Spare per 8 Data Segments (32Kbits)	•16 Byte Block ->DBEC/TBED •Serial Error Correcting Circuitry On Chip •Organized 1 Chip Wide •1 Block per 18 Memory Cycles •1 Electronic Fuse Spare per Data Segment (4Kbits)
1 bad bit in 4000	•8 Byte Block ->SBEC/DBED •Parallel Error Correcting Circuitry Off Chip •Organized 9 Chips Wide •1 Block per Memory Cycle •1 Electronic Fuse Spare per Data Segment (4Kbits)	•8 Byte Block ->SBEC/DBED •Serial Error Correcting Circuitry On Chip •Organized 1 Chip Wide •1 Block per 9 Memory Cycles •1 Electronic Fuse Spare per Data Segment (4Kbits)
1 bad bit in 1600	•Fully Random Access •No Error Correcting Circuitry •Organized 1 Chip Wide •1 Byte (or Bit or Nibble) per Memory Cycle •4 Fused Spares per Data Segment (4Kbits)	•Fully Random Access •No Error Correcting Circuitry •Organized 1 Chip Wide •1 Byte (or Bit or Nibble) per Memory Cycle •4 Fused Spares per Data Segment (4Kbits)

7. CONCLUSIONS

Many critical design issues for a one-megabit voltage mode magneto-resistive memory were addressed in this thesis. An overall chip organization was developed, as were compact, space efficient layouts of MR cells and supporting circuitry. Various techniques to improve chip yield in the presence of hardware failures also were investigated and reported.

This one-megabit design demonstrated practical solutions to the problems of designing a large voltage mode MRAM. Many of these design features it is hoped will find their way into successful commercial products.

7.1 Future Research

Magneto-resistive memories and the technologies that they use are undergoing tremendous changes. As a result, many modifications to this design will be necessary to take full advantage of these improvements. The most notable improvement in recent days is a dramatic increase in the MR coefficient of new magneto-resistive materials. These new materials are said to possess Giant Magneto-Resistance (GMR) and have usable MR coefficients of approximately 6.8% in contrast to the 2.2% used for the design described in this dissertation. Benefits of using GMR materials include increased signal levels, smaller cell size, increased density, faster operation, larger noise margins, and simpler sensing schemes.

REFERENCES

[1] K. E. Spears, "Design and simulation of a wafer scale integration magneto-resistive memory architecture," Doctoral dissertation, Iowa State University, 1990.

[2] J. L. Hennesey and D.A. Patterson, *Computer Architecture A Quantitative Approach*, San Mateo, CA: Morgan Kaufmann Publishers, 1990.

[3] G.B. Granley, J.M. Daughton, A.V. Pohm, C.S. Comstock; "Properties of 1.4 x 2.8 μm2 active area M-R elements," *IEEE Transactions on Magnetics* (Nov. 1991):5517-5519.

[4] K. E. Spears, "Ultra-Dense Magneto-resistive Mass Memory.", Second Quarter Report for NASA contract # NAS7-1152, January 30, 1992.

[5] A.V. Pohm, C.S. Comstock, G.B. Granley, J.M. Daughton; "Write Stability of 2.0 X 20μm^2 M-R Memory Cells", *IEEE Transactions on Magnetics,* (Nov. 1991):5515-5517.

[6] W. Thomson, "On the electrodynamic qualities of metals : Effects of magnetization on the electric conductivity of nickel and iron." *Proceedings of the Royal Society* 8 (1857): 546-550.

[7] L.J. Schwee, "Proposal on Cross-Tie Wall and Bloch Line Propagation in Thin Magnetic Films", IEEE *Transactions on Magnetics.*, 1972 INTERMEG Conference Proceedings, (1972):405-407.

[8] A.V. Pohm and J. Daughton, US patent #4,780,848, 1983.

[9] K. T. M. Ranmuthu, "A low noise small signal sensing scheme in voltage mode for high density MRAMS," Doctoral dissertation, Iowa State University, 1993.

[10] I.W. Ranmuthu, "A switched capacitor, self referencing sensing scheme for high density magneto-resistive memories," Doctoral dissertation, Iowa State University, 1992.

[11] A.V. Pohm, "Temperature Transient Effects in M-R Memory Elements," personal communication.

[12] A.V. Pohm; O.P. Agrawal. *High Speed Memory Systems*, Englewood Cliffs, NJ: Prentice-Hall, (1983):119-143

[13] T.R.N. Rao; E. Fujiwarq. *Error-Control Coding for Computer Systems..* Englewood Cliffs, NJ: Prentice-Hall, 1989.

[14] W. W. Peterson; E. J. Welder, Jr., *Error Correcting Codes..* 2nd edition, Cambridge Mass, MIT Press, 1972.

[15] K. Kimura; T. Sakata; K. Itoh; T. Kaga; T Nishida and Y. Kawamoto, " A Block-Oriented RAM with Half-Sized DRAM cell and Quasi-folded Data-Line Architecture,"*IEEE Journal of Solid State Circuits* (Nov. 1991) :1511-1518.

[16] A.V. Pohm; J. Daughton; C.S. Comstock; H. Y. Yoo and J. Hur, "Threshold Properties of 1, 2 and 4 μm Multilayer Magneto-Resistive Memory Cells," *IEEE Transactions on Magnetics* (Sept 1987) :2575-2577.

[17] V. Mehra, "Implementation of a Sensing Technique for Non-volatile MR memories," Master's thesis, Iowa State University, 1988.

[18] A.V. Pohm; J.S. T. Haung; J. Daughton; D.R. Krahn; V. Mehra, "The Design of a One Megabit Non-volatile MR Memory Chip using 1.5x5 um^2 Cells," *IEEE Transactions on Magnetics* (Nov. 1988) :3117-3119.

[19] A.V. Pohm; C.S. Comstock; J.M. Daughton and D.R. Krahn, "Ultra High Density Non-destructive Readout MR memory Cells," Comp Euro '89, Hamburg, Germany, May 1989.

[20] A.V. Pohm; C.S. Comstock; and A.T. Hurst, "Quadrupled Non-destructive Outputs from MR Memory Cells using Reversed Word Fields," Paper CA-13 in 34th Conference on Magnetism and Magnetic Materials, Boston, Nov. 1989.

[21] K. T. M. Ranmuthu; I.W. Ranmuthu; A.V. Pohm; C.S. Comstock; and M. Hassoun, "10-35 Nanosecond Magneto-Resistive Memories," International Magnetics Conference, Brighton, UK, 1990.

[22] I.W. Ranmuthu, "Reprogrammable Logic Array Using MR Elements," Master's thesis, Iowa State University, 1990.

[23] K.T. M. Ranmuthu, "10-35 Nanosecond Magneto-Resistive Memories," Master's thesis, Iowa State University, 1990.

APPENDIX. KEY DESIGN PARAMETERS

Key Layout Parameters

minimum Poly width	.8μm
minimum Poly to Poly space	1.2μm
all contacts	1μm x 1μm
minimum Metal 1 width	.8μm
minimum Metal 1 to Metal 1 space	1.2μm
all vias	1.2μm x 1.2μm
minimum Metal 2 width	1.2 μm
minimum Metal 2 to Metal 2 space	1.6μm
minimum PMOS transistor	width 2μm, length .9μm
minimum NMOS transistor	width 2μm, length .8μm

Sheet Resistance

N+ Source/Drain	3.03 Ω/square
P+ Source/Drain	2.64 Ω/square
Polysilicon	3.52 Ω/square
Metal 1	0.088 Ω/square
Metal 2	0.049 Ω/square

PSPICE Models

```
.MODEL PB PMOS
(LEVEL=2 VTO=-.817 TOX=.02U NSUB=6E16 UO=267
+ THETA=.175 GAMMA=.64 LD=.07U  VMAX=3.9E5  XJ=.22U CJ=3.4E-4
+ MJ=.455 CJSW=5.2E-10 MJSW=.26 CGSO=3.7E-10 CGDO=3.7E-10 CGBO=5.6E-10
+ KAPPA=.01 ETA=.01 DELTA=.933 PB=.71 RSH=4.2 )

.MODEL NB NMOS
 (LEVEL=2 VTO=.730 TOX=.02U NSUB=1E16 UO=655
+ THETA=.13 GAMMA=.5 LD=.025U  VMAX=1.6E5  XJ=.18U CJ=2.1E-4
+ MJ=.458 CJSW=5.2E-10 MJSW=.282 CGSO=1.9E-10 CGDO=1.9E-10 CGBO=6.1E-10
+ KAPPA=.076 ETA=.003  DELTA=1.0 PB=.76 RSH=4.0 )
```