Iowa State University
Digital Repository

Retrospective Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

1990

Design and simulation of a wafer scale integration magnetoresistive memory architecture

Kurt Eugene Spears
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Electrical and Electronics Commons

Recommended Citation
Spears, Kurt Eugene, "Design and simulation of a wafer scale integration magnetoresistive memory architecture " (1990). *Retrospective Theses and Dissertations*. 9894.
https://lib.dr.iastate.edu/rtd/9894

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

المنارة للاستشارات

www.manaraa.com

1

91

10567

U·M·I

MICROFILMED 1991

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript 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 9110567

Design and simulation of a wafer scale integration magnetoresistive memory architecture

Spears, Kurt Eugene, Ph.D.

Iowa State University, 1990

Copyright ©1991 by Spears, Kurt Eugene. All rights reserved.

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

Design and simulation of a

wafer scale integration

magnetoresistive memory architecture

by

Kurt Eugene Spears

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 the Committee:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1990

Copyright © Kurt Eugene Spears, 1990. All rights reserved.

CONTENTS

1. Introduction 1
1.1 Purpose of Thesis 1
1.2 Methodology 1

2. MEMORY TECHNOLOGIES 3
2.1 Introduction 3
2.2 Existing Memory Technologies 3
2.2.1 Electro-Mechanical Memory 4
2.2.2 Semiconductor Memory 14
2.3 Memory Hierarchy 18
2.3.1 Performance Impact of Magnetoresistive Memory 19
2.4 Magnetoresistive Memory 21
2.4.1 Introduction 21
2.4.2 Wafer Scale Application 25

3. WAFER SCALE INTEGRATION 27
3.1 Introduction 27
3.2 History of Wafer Scale Integration 27
3.3 Wafer Scale Integration Technologies 27
3.4 Current Wafer Scale Integration Products and Projects 29
3.4.1 Anamartic Wafer Stack 30
3.4.2 Inova 1 Mbit SRAM 30
3.4.3 Inova 8-Mbit SRAM 31
3.5 Wafer Scale Integration Technologies For Memory 31

4. WAFER SCALE DESIGN 33
4.1 Yield versus Capacity Trade-off 33
4.2 Testability 33
4.3 Reconfigurability 35
4.4 Power Requirements 36
4.5 Performance 38
4.6 Architecture and Redundancy 38
4.7 Address Decoding 41

5. DESIGN SIMULATION 43
5.1 Introduction 43
5.2 Overview of Model 43
5.3 16-Kb Model 44
5.4 1-Mb Model 46
5.5 Wafer Yield Model 47

6. SIMULATION RESULTS 51
6.1 Introduction 51
6.2 Limitations and Approximations 51
6.3 Interpreting the Results 52
6.4 Numerical Results 55
6.5 Optimizing Redundancy 59
6.6 Characteristics of the Optimized Design 60

7. CONCLUSIONS 66
7.1 Future Research 66

8. REFERENCES 67
9. ACKNOWLEDGEMENTS 69
10. APPENDIX A: APPROXIMATION VERIFICATION CODE 70
11. APPENDIX B: SIMULATION SOURCE CODE 73
12. APPENDIX C: GRAPHICAL SIMULATION RESULTS 80

LIST OF FIGURES

Figure 1. Cost and Access Time for Current Memory Technologies 20
Figure 2. Magnetoresistive double layer 22
Figure 3. Magnetoresistive memory element 23
Figure 4. Stoner-Wolhfarth threshold for a 1.5x5 um element 24
Figure 5. Output curves for a magnetoresistive element 25
Figure 6. Redundancy locations for lowest levels 40
Figure 7. Address decoding of host address 42
Figure 8. Block diagram of simulation 43
Figure 9. Effectiveness of Redundancy for M=64 47
Figure 10. Manufacturing Process Range and Distributions 53
Figure 11. Normal Distribution Used in Simulation 55
Figure 12. Distribution of Top Ten Mean Distribution Combinations 62
Figure 13. Capacity Probability 62
Figure 14. Number of Functional Modules Probability 62
Figure 15. Yield of Capacities (Number of Functional Modules) 63
Figure 16. Yield of Capacities over 200 MB 63
Figure 17. Cost and Access Times of Memory Technologies with 1993 Additions 65
Figure 18. Mean Yield Capacities (grouped by level 1 redundancy) 81
Figure 19. Centered Yield Capacities (grouped by level 1 redundancy) 82
Figure 20. LL_Centered Yield Capacities (grouped by level 1 redundancy) 83
Figure 21. LR_Centered Yield Capacities (grouped by level 1 redundancy) 84
Figure 22. UL_Centered Yield Capacities (grouped by level 1 redundancy) 85
Figure 23. UR_Centered Yield Capacities (grouped by level 1 redundancy) 86
Figure 24. Mean Yield Capacities (grouped by level 2 redundancy) 87
Figure 25. Centered Yield Capacities (grouped by level 2 redundancy) 88
Figure 26. LL_Centered Yield Capacities (grouped by level 2 redundancy) 89
Figure 27. LR_Centered Yield Capacities (grouped by level 2 redundancy) 90
Figure 28. UL_Centered Yield Capacities (grouped by level 2 redundancy) 91
Figure 29. UR_Centered Yield Capacities (grouped by level 2 redundancy) 92

LIST OF TABLES

TABLE 1. 1990 Winchester Disk Attributes 6

TABLE 2. Projected 1993 Winchester Disk Attributes 6

TABLE 3. 1990 Optical Disk Attributes 7

TABLE 4. Projected 1993 Optical Disk Attributes 7

TABLE 5. 1990 Reel-to-Reel Tape Attributes 9

TABLE 6. Projected 1993 Reel-to-Reel Tape Attributes 9

TABLE 7. 1990 Half-Inch Cartridge Longitudinal Tape Attributes 10

TABLE 8. Projected 1993 Half-Inch Cartridge Longitudinal Tape Attributes 10

TABLE 9. 1990 Cartridge Longitudinal Tape Attributes 11

TABLE 10. Projected 1993 Cartridge Longitudinal Tape Attributes 11

TABLE 11. 1990 Helical Scan Tape Attributes 12

TABLE 12. Projected 1993 Helical Scan Tape Attributes 12

TABLE 13. 1990 Optical Tape Attributes 13

TABLE 14. Projected 1993 Optical Tape Attributes 13

TABLE 15. 1990 Autochanger System Attributes 14

TABLE 16. Projected 1993 Autochanger System Attributes 14

TABLE 17. 1990 Static RAM Attributes 16

TABLE 18. Projected 1993 Static RAM Attributes 16

TABLE 19. 1990 Dynamic RAM Attributes 17

TABLE 20. Projected 1993 Dynamic RAM Attributes 17

TABLE 21. Online Memory Attributes 18

TABLE 22. Projected Cost/Performance For DRAM, MRAM, and Hard Disk Mass Memories 19

TABLE 23. Miss Ratios for Three Commercial Systems 19

TABLE 24. Disk Cache Performance for Three Commercial Systems 21

TABLE 25. Previous WSI Projects 28

TABLE 26. Design Requirements for Wafer Scale Magnetoresistive Memory 32

TABLE 27. Test times for memory module testing 34

TABLE 28. Test costs of memory module testing 34

TABLE 29. Redundancy costs for binary tree structure 39

TABLE 30. Areal cost of redundancy at 16-Kb level 48

TABLE 31. Areal cost of redundancy at 1-Mb level 48
TABLE 32. Percent of Usable Area at Wafer level 49
TABLE 33. Number of 1-Mb Modules on Wafer 49
TABLE 34. Yields vs. Parameters 52
TABLE 35. Values of Normal Distribution 56
TABLE 36. Numerical Results of the Simulation 57
TABLE 37. Numerical Results of the Simulation 58
TABLE 38. Top Ten Yields for Distributions 59
TABLE 39. Estimated Product Costs 64
TABLE 40. Wafer Scale Product Attributes 64

1. Introduction

1.1 Purpose of Thesis

The purpose of this thesis is to investigate the implementation of magnetoresistive memory using wafer scale integration. Wafer scale integration involves the integration of very large numbers of semiconductor devices on a single semiconductor wafer. The primary concern of the thesis is examining the types of defects and their distribution to facilitate the design of redundancy on the wafer to allow acceptable yields.

1.2 Methodology

The methodology used to arrive at the optimized wafer design parallels the format of the thesis. First, current memory technologies are examined to establish performance, capacity and yield requirements for a new memory product. This work is detailed in Chapter Two. Chapter Two also examines how magnetoresistive memory competes with existing memory technologies and why wafer scale integration is required to allow magnetoresistive memory to succeed. Given the need for wafer scale integration, Chapter Three examines the current state of wafer scale integration technology. Chapter Three examines current trends and products utilizing wafer scale integration and then focuses on wafer scale memory products and their associated problems. These problems consist mainly of various types of semiconductor defects and their effect on product yields. With knowledge of the types and densities of defects in wafer scale integration, Chapter Four describes the variety of methods of compensating for defects. Each of the methods has advantages and disadvantages in terms of yield, capacity and testability. Chapter Four examines the interdependencies of yield, capacity and testability and proposes feasible memory designs. These memory designs and defect probabilities are modelled in Chapter Five using a computer simulation. The computer simulation

determines whether the redundancy in the memory design can compensate the simulated defect density and distribution. The results of the computer simulations are detailed in Chapter Six. The results indicate the optimal redundancy for wafer scale magnetoresistive memory. The results are utilized in Chapter Six to optimize the design of the wafer scale memory. Finally, Chapter Seven summarizes the feasibility of wafer scale magnetoresistive memory and describes future research.

2. MEMORY TECHNOLOGIES

2.1 Introduction

Memory systems are a critical component of computer systems. Memory systems in computers consist of multiple components with differing size, speed and cost attributes. This chapter examines existing memory technologies and discusses the advantages and disadvantages of each. Memory hierarchies which combine memory technologies into a memory system are summarized. This chapter also examines magnetoresistive memory technology and how it can be used to enhance memory system performance.

2.2 Existing Memory Technologies

Existing memory technologies can be divided into two categories, electro-mechanical and semiconductor. Electro-mechanical memory consisting of winchester disks, optical disks, and tape systems is comprised of a mechanical moving media and a read-write subsystem for writing and reading data from the media. Semiconductor memory consists of static and dynamic random access memory (RAM). Bubble memory will not be considered because it is not widely used and is not an emerging technology. Semiconductor memory consists of electronic devices on a semiconductor substrate and has no moving media. Each of these technologies is discussed in detail below and summarized according to the following attributes.

Access time — Access time refers to the amount of delay associated with fulfilling a random request for data.

Capacity — Capacity is the amount of on-line data which can be stored on a physical unit.

Transfer rate	Transfer rate is the amount of data per second which can be written or read from the device on a sustained basis.
Cost per megabyte	Cost per megabyte is a merit which evaluates a technology based on unit price and capacity. Cost per megabyte only includes the capacity which can be accessed without operator intervention. For stand alone units this is a single side of a single piece of media. For autochangers, it is the entire amount of media which can be present in the autochanger.
Removable media cost	Removable media cost is the cost per megabyte of removable media for technologies with removable media.
Areal density	Areal density is the number of bits which can be stored in one square inch.

2.2.1 Electro-Mechanical Memory

Electro-mechanical memory consists of a read-write subsystem and some type of media. The electro-mechanical memory typically uses either magnetic or magneto-optical properties to store data. Electro-mechanical memory is characterized by slow access times, high capacity, moderate transfer rates and moderate to low cost per megabyte.

2.2.1.1 Winchester Disks Winchester disks use magnetic recording technology to record information on rigid magnetic media. The recording subsystem consists of rotating magnetic media and a read-write head which can traverse the tracks of data which are recorded in concentric circles on the media.

Winchester disk technology continues to advance despite assertions that it is a mature technology with little room for improvement. Key areas of advancement are in areal recording density and access time. Areal recording density in combination with recording area determines the capacity of a platter. Multiple platters are present in high performance high capacity winchester

disks. Areal recording density is the product of two factors, recording density per track and track density. Recording density per track has improved with the use of thin-film plated media and lower flying heights for heads.[1] Improved recording techniques have also contributed to increased recording density per track.[2] Track density has been improved through the development of dedicated servo surfaces and embedded positioning systems. Dedicated servo surfaces use one of the available platters to control the tracking system for all the remaining platters. Embedded positioning systems use servo information which is embedded in the recording tracks on all the platters. Access time has improved through the use of voice-coil actuators and higher rotational speeds. Voice-coil actuators position the read/write head over the proper track and have reduced access time due to positioning the head and have also contributed to increased track density. Higher rotational speeds have reduced access time by reducing the latency due to disk rotation. Winchester disks are expected to advance technologically in the following areas.[3]

- Thin-film and magnetoresistive heads
- Disk medium using film barium ferrite or metal particles.
- Embedded servo for increased tracks per inch
- Improved signal processing to compensate for decreased signal-to-noise ratio from smaller bits.
- Low mass actuators and air bearings to improve head to track registration.

Winchester disk attributes for the current year(1990) and projected attributes for 1993 are listed in Tables 1 and 2 respectively.

2.2.1.2 Optical Disks Rewriteable optical disks consist of an optically sensitive surface and an optical read-write head. Optical disks currently on the market utilize magneto-optical recording. A laser diode writes the data on the surface at a high laser power and reads data using a lower laser power. A biasing magnet is used to control the state of the bits heated by the high power laser. A

TABLE 1. 1990 Winchester Disk Attributes

Access Time	12 - 20	milliseconds
Capacity	20 - 1600	megabytes
Transfer Rate	2 - 15	megabits/second
Cost per Megabyte	4 - 10	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	40 x 10e6	bits per square inch
Source: IEEE Spectrum Feb 1987		

TABLE 2. Projected 1993 Winchester Disk Attributes

Access Time	10 - 20	milliseconds
Capacity	40 - 6400	megabytes
Transfer Rate	2 - 30	megabits/second
Cost per Megabyte	2 - 10	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	80 x 10e6	bits per square inch
Source: IEEE Spectrum Feb 1987		

combination of factors including the mass of the biasing magnet and the time required for the bits to cool to below the Curie temperature necessitate a two pass write operation. The first pass is used to set all bits to a known state while the second pass is used to write bits which should be in the opposite state.

Rewriteable optical disks are a recent entry in the mass storage market. Optical disks are expected to compete with winchester disks in the future but currently are slower due to larger mass heads and multiple pass write operations. Due to the size and expense of the optical head, optical drives only write on one side of the media. Using the other side of the media requires that the media be removed and flipped by a mechanical system or a human operator. Because of this fact, capacity is quoted as the single side capacity and cost per megabyte is calculated based on the single side capacity. Multiple pass write operations decrease the transfer rate for write operations and as a result two transfer rates are quoted. The first rate refers to the read transfer rate while the value in parentheses is the write transfer rate.

Optical disks have the advantage of removable media and high areal density. The removable media allows a single optical mechanism to record data on a number of surfaces and allows for the removal and archiving of data. Higher areal density should allow more data in smaller packages as optical heads are reduced in size and mass. Optical disk research is expected to advance in the following areas.[3]

- Shorter wavelength and higher power laser diodes.
- Direct-overwrite media which allows one pass write operations.
- Improved signal processing to compensate for decreased signal-to-noise ratio from smaller bits.
- Low mass actuators and air bearings to improve head to track registration.
- Integrated low mass optical heads.

The current(1990) attributes and projected 1993 attributes are listed in Tables 3 and 4 respectively.

TABLE 3. 1990 Optical Disk Attributes

Access Time	40 - 100	milliseconds
Capacity	160 - 512	megabytes (per side)
Transfer Rate	2.4 - 6.6(1.2 - 3.3)	megabits/second
Cost per Megabyte	12 - 24	$/megabyte
Removable Media Cost	0.38	$/megabyte
Areal Density	350 x 10e6	bits per square inch
Source: Rothchild Consultants - The Optical Memory Report Dec. 1989		

TABLE 4. Projected 1993 Optical Disk Attributes

Access Time	15 - 100	milliseconds
Capacity	256 - 2000	megabytes
Transfer Rate	2.4 - 24	megabits/second
Cost per Megabyte	3 - 12	$/megabyte
Removable Media Cost	0.10 - 0.38	$/megabyte
Areal Density	1.4 x 10e9	bits per square inch
Source: Rothchild Consultants - The Optical Memory Report Dec. 1989		

2.2.1.3 Reel-to-Reel Tapes Reel-to-Reel tape has existed in the computer industry for many years and is still the current archive for most data processing operations. Large libraries of archived data on half inch tape perpetuate the use of the technology which has changed very little since its introduction.

Half inch reel-to-reel tape uses a nine track read-write head to record data in parallel tracks as the tape media is moved over the stationary head. Eight of the tracks contain data while the ninth track contains parity. The data are recorded in either 800, 1600 or 6250 bits per inch density. Major drawbacks of tape media are the rewind time encountered at the end of media and reposition cycles or wasted media if data are not ready when the drive is correctly positioned and moving down the tape.

Data compression has recently increased capacity by a factor of about three and is expected to become widely used. Other areas of technological advancement are increased recording densities and higher linear tape speed. Recording density is projected to double to 12500 bits per inch and later double again to 25000 bits per inch. Technological advancement areas are listed below.

- Data compression.
- Increased recording density.
- Higher linear tape speed.

The current(1990) attributes and projected 1993 attributes are listed in Tables 5 and 6 respectively.

2.2.1.4 Half-Inch Cartridge Longitudinal Tapes Half-inch cartridge longitudinal tape is a relatively new tape technology based on IBM's 3480 tape cartridge. The cartridge measures 1 x 4 x 5 inches and is between 450 and 600 feet in length. The tape is pulled out of the cartridge and wound on a take-up reel during writing or reading. The tape is rewound back into the cartridge to be unloaded.

TABLE 5. 1990 Reel-to-Reel Tape Attributes

Access Time	45 - 120	seconds
Capacity	40 - 500	megabytes
Transfer Rate	0.8 - 10.0	megabits/second
Cost per Megabyte	40 - 80	$/megabyte
Removable Media Cost	0.10 - 0.20	$/megabyte
Areal Density	100 x 10e3	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Half-Inch Reel and Cartridge Products		

TABLE 6. Projected 1993 Reel-to-Reel Tape Attributes

Access Time	45 - 120	seconds
Capacity	40 - 700	megabytes
Transfer Rate	0.8 - 18.0	megabits/second
Cost per Megabyte	30 - 80	$/megabyte
Removable Media Cost	0.10 - 0.20	$/megabyte
Areal Density	200 x 10e3	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Half-Inch Reel and Cartridge Products		

The data are recorded in parallel longitudinal tracks on the tape. The tape contains between 18 and 48 tracks which are recorded in a single pass for high performance drives and are recorded in multiple passes for lower performance drives. Areas of technological advancement are listed below.

- Data compression.
- Track density
- Higher linear tape speed.

Current(1990) attributes and projected 1993 attributes for half-inch cartridge longitudinal tape are listed in Tables 7 and 8.

2.2.1.5 Quarter-Inch Cartridge Longitudinal Tapes Quarter-inch cartridge longitudinal tape is typically used on lower-end systems and does not have the performance of half-inch tape. The technology records data in parallel longitudinal tracks on tape contained in a cartridge. The tape

TABLE 7. 1990 Half-Inch Cartridge Longitudinal Tape Attributes

Access Time	20 - 50	seconds
Capacity	95 - 1280	megabytes
Transfer Rate	0.4 - 24.0	megabits/second
Cost per Megabyte	4 - 20	$/megabyte
Removable Media Cost	0.10 -0.20	$/megabyte
Areal Density	2 x 10e6	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Half-Inch Reel and Cartridge Products		

TABLE 8. Projected 1993 Half-Inch Cartridge Longitudinal Tape Attributes

Access Time	20 - 50	seconds
Capacity	40 - 1280	megabytes
Transfer Rate	0.4 - 24.0	megabits/second
Cost per Megabyte	3 - 20	$/megabyte
Removable Media Cost	0.08 - 0.16	$/megabyte
Areal Density	5 x 10e6	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Half-Inch Reel and Cartridge Products		

may contain between 9 and 80 tracks which are recorded during multiple passes. Higher performance drives contain heads which write four tracks in parallel while lower performance drives write one track at a time. Areas of technical advancement for quarter-inch tape are listed below.

- Data compression.
- Increased coercivity for higher recording and track densities.
- Increased length cartridges.
- Higher linear tape speed.

Current(1990) attributes and projected 1993 attributes for quarter-inch cartridge longitudinal tape are listed in Tables 9 and 10.

2.2.1.6 Helical Scan Tapes Helical scan technology records data in parallel tracks which are recorded diagonally across the tape. Helical scan technology typically uses 4 mm or 8 mm tape

TABLE 9. 1990 Cartridge Longitudinal Tape Attributes

Access Time	10 -45	seconds
Capacity	10 - 525	megabytes
Transfer Rate	0.032 - 1.0	megabits/second
Cost per Megabyte	4 - 12	$/megabyte
Removable Media Cost	0.10 - 1.00	$/megabyte
Areal Density	1.4 x 10e6	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Data Cassette and Data Cartridge Products		

TABLE 10. Projected 1993 Cartridge Longitudinal Tape Attributes

Access Time	10 - 45	seconds
Capacity	10 - 1250	megabytes
Transfer Rate	0.032 - 1.8	megabits/second
Cost per Megabyte	2 - 10	$/megabyte
Removable Media Cost	0.05 - 1.00	$/megabyte
Areal Density	3.0 x 10e6	bits per square inch
Source: Freeman Reports Computer Tape Outlook 1989 Data Cassette and Data Cartridge Products		

although a few systems use 1/4" and 1/2" tape. Helical scan technology is very similar to television video tape technology. Several systems including Honeywell's VLDS helical scan tape recorder utilize T120 VHS tapes. Units using T120 VHS tapes can store between 2.5 and 5.2 Gbytes. 8 mm units from Exabyte store 2.3 Gbytes on an 8 mm video cassette. 4 mm products utilize Digital Audio Tape (DAT) technology and store between 700 Mbytes and 1.3 Gbytes on a cassette.

Helical scan technologies use magnetic heads mounted on a rotating drum to record data at an angle across the tape. The drum rotates at 1800 to 3600 rpm while the tape moves by at a slow speed. This results in a very high track density of 800 to 2000 tracks per inch. Combined with a recording density between 15,000 and 60,000 bits per inch this track density corresponds to an areal density of 100 million bits per square inch. Helical scan technology is expected to advance in the following areas.

- Data compression.
- Track density.
- Longer tape.
- Recording density.

Current(1990) attributes and projected 1993 attributes for helical scan tape are listed in Tables 11 and 12.

TABLE 11. 1990 Helical Scan Tape Attributes

Access Time	20	seconds
Capacity	700 - 5200	megabytes
Transfer Rate	0.96 - 1.5	megabits/second
Cost per Megabyte	2 - 4	$/megabyte
Removable Media Cost	0.004 - 0.014	$/megabyte
Areal Density	100 x 10e6	bits per square inch
Source: IEEE Spectrum October 1989.		

TABLE 12. Projected 1993 Helical Scan Tape Attributes

Access Time	20 -25	seconds
Capacity	700 - 10,000	megabytes
Transfer Rate	2.5 - 5.0	megabits/second
Cost per Megabyte	1 - 4	$/megabyte
Removable Media Cost	0.004 - 0.014	$/megabyte
Areal Density	250 x 10e6	bits per square inch
Source: IEEE Spectrum October 1989.		

2.2.1.7 Optical Tape Optical tape technology utilizes a dye polymer on a polyester-based material to achieve high capacity write-once storage. The data are recorded on a 35 mm wide tape. Creo Electronics Corporation has an optical tape drive which stores one Terabyte. The drive records 28,500 bits per inch and has a track density of 15,875 tracks per inch. The unit costs $225,000 which is $0.225 per megabyte. The removable media cost is $10,000 for one terabyte which is $0.01 per megabyte. Optical tape technology is expected to advance in the following areas.

- Data compression.
- Rewriteable media.
- Lower capacities at lower price.

Current(1990) attributes and projected 1993 attributes for optical tape are listed in Tables 13 and 14.

TABLE 13. 1990 Optical Tape Attributes

Access Time	28	seconds
Capacity	1,000,000	megabytes
Transfer Rate	24	megabits/second
Cost per Megabyte	0.225	$/megabyte
Removable Media Cost	0.01	$/megabyte
Areal Density	450 x 10e6	bits per square inch
Source: Freeman Reports: Optical Data Storage Outlook		

TABLE 14. Projected 1993 Optical Tape Attributes

Access Time	28	seconds
Capacity	8,000 - 2,500,000	megabytes
Transfer Rate	2 - 60	megabits/second
Cost per Megabyte	0.09 - 0.60	$/megabyte
Removable Media Cost	0.0025 - 0.005	$/megabyte
Areal Density	1130 x 10e6	bits per square inch
Source: Freeman Reports: Optical Data Storage Outlook		

2.2.1.8 Autochanger Systems Autochanger systems utilize a robotic autochanger, removable media and multiple drive mechanisms to provide extremely large amounts of on-line storage. Autochanger systems utilize optical or cartridge tape. The large access time is due to delays while the mechanical robotic system retrieves the required media from a storage slot.

Optical autochangers utilize 14", 12", 8" or 5.25" media. Cartridge accesses involve removing the current media from the drive, storing the media, retrieving the new media and loading it into the drive. The cartridge access takes between ten and fifteen seconds on an optical drive. Tape autochangers utilize cartridge tapes. 8 mm and 4 mm tapes are used for large capacity systems

while 3480 type cartridges are used for high performance systems. Technological advancement areas for autochangers are listed below.

- Higher speed and more reliable robotic assemblies.
- Technological advancements in appropriate media.

Current(1990) attributes and projected 1993 attributes for optical tape are listed in Tables 15 and 16.

TABLE 15. 1990 Autochanger System Attributes

Access Time	10 - 90	seconds
Capacity	1000 - 552,000	megabytes
Transfer Rate	media dependent	megabits/second
Cost per Megabyte	0.08 - 2.00	$/megabyte
Removable Media Cost	media dependent	$/megabyte
Areal Density	media dependent	bits per square inch

TABLE 16. Projected 1993 Autochanger System Attributes

Access Time	10 - 90	seconds
Capacity	1000 - 1,300,000	megabytes
Transfer Rate	media dependent	megabits/second
Cost per Megabyte	0.04 - 1.00	$/megabyte
Removable Media Cost	media dependent	$/megabyte
Areal Density	media dependent	bits per square inch

2.2.2 Semiconductor Memory

Semiconductor memory consists of electronic devices on a semiconductor substrate. Semiconductor memory can utilize any of several semiconductor technologies although Complementary Metal Oxide Semiconductor (CMOS) is the dominant technology today. Bipolar CMOS (BiCMOS) and Emitter Coupled Logic technologies are expected to be the dominant technologies in the 1990's due to increased speed requirements.[4]

Semiconductor memory has diverged into three broad categories. These categories are static random access memory (SRAM), dynamic random access memory (DRAM) and bubble memory. Static RAM is the fastest of the semiconductor memories while bubble memory is the slowest. Bubble memory is non-volatile which means it retains its contents when power is removed. SRAM and DRAM lose their contents if power is removed.

Semiconductor memory is characterized by fast access times, moderate capacity, high transfer rates and high cost per megabyte. Access time refers to the amount of delay associated with fulfilling a random request for data. Capacity is the amount of data which can be stored on a physical unit. Transfer rate is the amount of data per second which can be written or read from the device on a sustained basis. Cost per megabyte is a merit which evaluates a technology based on unit price and capacity. The 1993 estimated cost per megabyte is based on semiconductor trends which indicate that chip capacities quadruple every three years with chip size increasing by a factor of 1.2. Lower yields which result from increased chip size and increasingly complex production prevent cost from dropping proportional to density gains. Each of these characteristics is discussed below for the existing semiconductor technologies.

2.2.2.1 Static RAM Static RAM is the fastest semiconductor memory. The density and capacity of static RAM is lower than dynamic RAM because static RAM requires four to six transistors per memory cell compared to dynamic RAM which requires one. The static RAM is faster because refresh operations are not required. Static RAM is used in computer systems near the high speed CPU's. Advances in CPU cycle time have required static RAM to become increasingly faster. Static RAM using BiCMOS technology can operate in the sub 10 nanosecond range.

Capacity is quoted for banks or other combinations of integrated circuits which create byte wide memory. Higher capacities can be obtained with multiple banks on a board. Cost per megabyte is based on quoted original equipment manufacturer (OEM) prices for the ICs multiplied

by a factor of three to compensate for board cost, production cost, distribution, profit and other factors. Technical advancement areas for static RAM are listed below.

- BiCMOS process.
- GaAs technological advancements.
- Integrated circuit yield improvements.

Tables 17 and 18 contain current and projected 1993 static RAM attributes. Projected attributes are based on long term semiconductor trends.

TABLE 17. 1990 Static RAM Attributes

Access Time	3.5 - 200	nanoseconds
Capacity	0.03 - 1.0	megabytes
Transfer Rate	40 - 2250	megabits/second
Cost per Megabyte	760 - 12,200	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	10 x 10e6	bits per square inch
Source: EDN Feb. 15, 1990		

TABLE 18. Projected 1993 Static RAM Attributes

Access Time	1 - 200	nanoseconds
Capacity	0.03 - 4	megabytes
Transfer Rate	40 - 8000	megabits/second
Cost per Megabyte	250 - 4000	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	33 x 10e6	bits per square inch

2.2.2.2 Dynamic RAM Dynamic RAM can achieve higher densities than static RAM because it only requires one transistor per memory cell. The value of a bit is stored as a charge on a capacitor which is accessed through the transistor. The charge on the capacitor leaks off with time. This requires that dynamic RAM memory cells be read and refreshed at periodic intervals.

Capacity is quoted for banks or other combinations of integrated circuits which create byte

wide memory. Higher capacities can be obtained with multiple banks on a board. Cost per megabyte is based on quoted end user prices from BYTE magazine January 1990 for low end prices. High end prices are based on manufacturer list prices for minicomputer systems. Dynamic RAM is expected to advance in the following technological areas.

- Lithography improvements.
- Reduced capacitor size.
- Integrated circuit yield improvements.
- Built-in or simplified refresh.

Tables 19 and 20 contain current and projected 1993 dynamic RAM attributes. Projected attributes are based on long term semiconductor trends.

TABLE 19. 1990 Dynamic RAM Attributes

Access Time	65 - 200	nanoseconds
Capacity	0.25 - 1.0	megabytes
Transfer Rate	40 - 120	megabits/second
Cost per Megabyte	120 - 500	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	50 x 10e6	bits per square inch

TABLE 20. Projected 1993 Dynamic RAM Attributes

Access Time	50 - 200	nanoseconds
Capacity	0.25 - 16.0	megabytes
Transfer Rate	40 - 160	megabits/second
Cost per Megabyte	40 - 170	$/megabyte
Removable Media Cost	N.A.	$/megabyte
Areal Density	170 x 10e6	bits per square inch
Source: Electrical Engineering Times Issue 594		

2.3 Memory Hierarchy

Computer systems use several memory technologies to form a complete memory system. A memory system which consists of at least two memories of differing speed and size is called a memory hierarchy.[5] A memory hierarchy which has properly selected components can provide the storage capacity of the largest (and typically slowest) technology at a speed only slightly slower than the fastest most expensive memory technology.

The attributes of existing technologies are summarized in Table 21 for online capacities. This means that the cost per megabyte is calculated based on online capacity.

TABLE 21. Online Memory Attributes

Technology	Cost ($/MByte)	Access Time	Unit Capacity(Megabytes)
Static RAM	760 - 12,200	3.5 - 200 nsec	0.03 - 1.0
Dynamic RAM	120 - 500	65 - 200 nsec	0.25 - 4.0
Bubble Memory	1000	10 msec	1
Winchester Disk	4 - 10	12 -20 msec	20 - 1600
Optical Disk	12 -24	40 - 100 msec	256 - 450
Helical Scan Tape	2 - 4	20 - 60 sec	700 - 5200
Half-inch Cartridge Tape	4 - 20	20 - 50 sec	95 - 1280
Quarter-inch Cartridge Tape	4 - 12	10 - 45 sec	10 - 525
Reel-to-Reel Tape	40 - 80	60 sec	40 - 500
Autochanger (optical)	0.5 - 2.0	10 - 15 sec	1000 - 250,000
Autochanger (tape)	0.08 - 1.00	20 - 50 sec	1000 - 552,000
Optical Tape	0.225	28 sec	1,000,000

The access time and cost of the various memory technologies are shown in Figure 1. The region between the dashed lines in Figure 1 is known as a gap. The gap is a region where no cost effective memory technology exists which could allow memory hierarchies to perform more effectively. Magnetoresistive memory is targeted as a gap technology. The following comparison details how magnetoresistive memory would improve memory system performance for an equivalent cost.

2.3.1 Performance Impact of Magnetoresistive Memory

This section will examine the performance gains which can be achieved by using magnetoresistive memory as a gap technology. The cost and performance attributes of the three technologies used are from Granley and Daughton[6] and are shown in Table 22.

TABLE 22. Projected Cost/Performance For DRAM, MRAM, and Hard Disk Mass Memories

	DRAM	MRAM	Hard Disk
Approximate Cost ($/MB)	100	10	1
Access Time (usec)	0.1	10	10,000

The performance impact of the magnetoresistive memory is evaluated using disk cache measurements from Smith.[7] Smith performed trace measurements on three commercial systems and evaluated the impact of disk caching. The measurements and simulated disk caching indicated the following miss ratios.

TABLE 23. Miss Ratios for Three Commercial Systems

System Name	On-line Disk Space	Miss Ratio for 500MB Cache	Miss Ratio for 50MB Cache	Miss Ratio for 250MB Cache	Miss Ratio for 25MB Cache
Crocker Bank	11.8 GB	0.09	0.14	--	--
Hughes Aircraft	9.9 GB	0.02	0.06	--	--
Stanford LA Center	2.5 GB	--	--	0.015	0.060

Using a simple model for evaluating effective access times, the performance impact of disk caches located in RAM and MRAM can be evaluated. Assuming the same amount of money is being expended to improve the system, the amount of MRAM which could be added is ten times greater than the DRAM. For the first two systems, the comparison is between 500MB of MRAM and 50 MB of DRAM. For the third system, the comparison is between 250 MB of MRAM and 25 MB of DRAM.

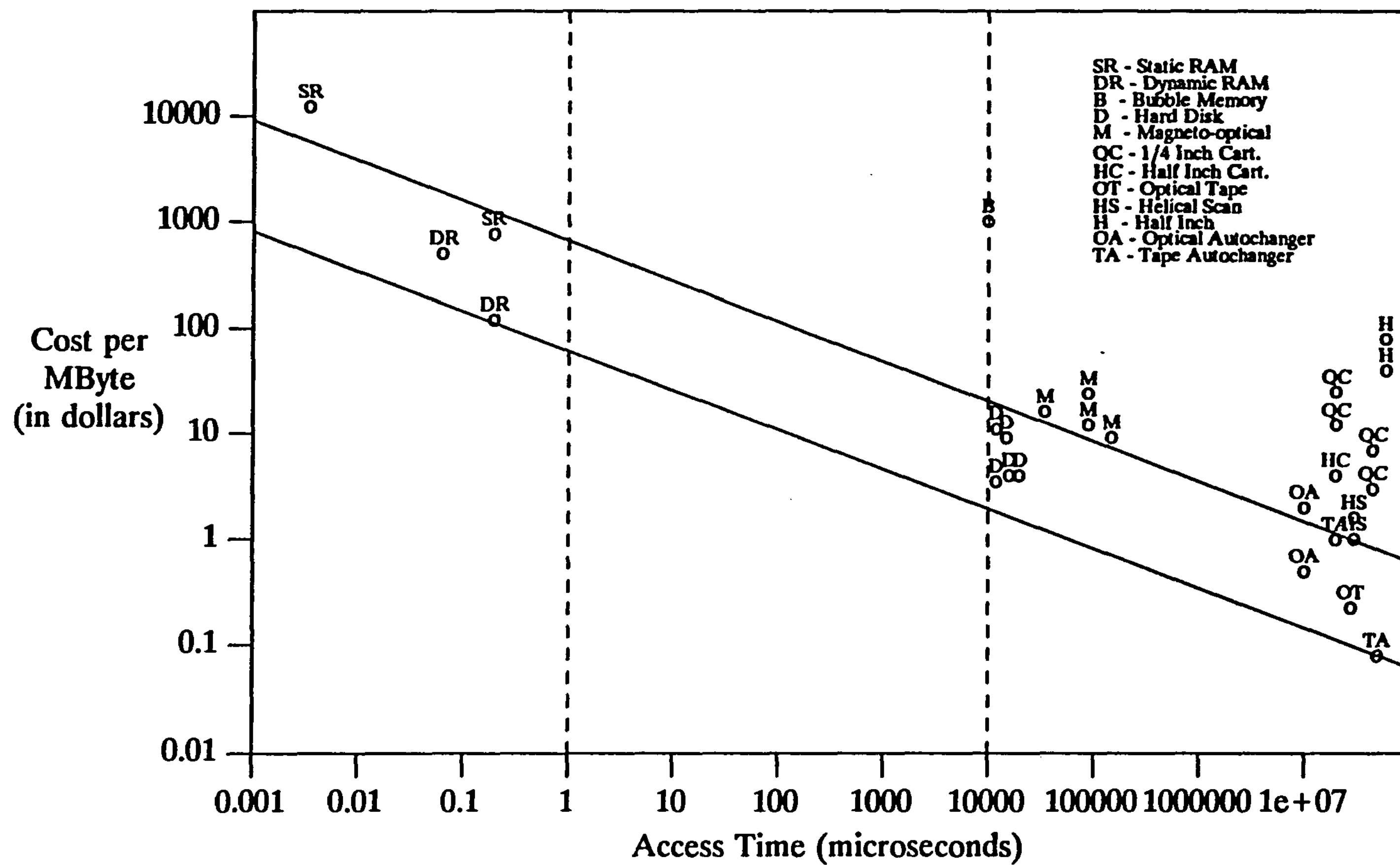

Figure 1. Cost and Access Time for Current Memory Technologies

The equation used to evaluate the effective access time is

$$T_{eff} = (1\text{-}MR)\cdot T_{cache} + MR\cdot T_{disk}\cdot(1+WR) \qquad (2.1)$$

where MR = Miss Ratio
and WR = Write Ratio (assumed to be 0.3)

The results of equation 2.1 are shown in Table 24. The memory systems using MRAM performed between 1.5 times and 3.8 times faster than the systems using DRAM disk caches costing equivalent amounts of money. These Figures show that MRAM can be used to cost effectively improve memory system performance. For Crocker Bank the performance is 1.5 times faster, while performance for Hughes Aircraft and the Stanford Linear Accelerator Center are 2.9 and 3.8 times faster respectively.

TABLE 24. Disk Cache Performance for Three Commercial Systems

System Name	Effective Access Time with MRAM	Effective Access Time with DRAM	Performance Gain using MRAM instead of DRAM
Crocker Bank	1179.1 usec	1820.1 usec	1.54
Hughes Aircraft	269.8 usec	780.1 usec	2.89
Stanford LA Center	204.9 usec	780.1 usec	3.81

2.4 Magnetoresistive Memory

2.4.1 Introduction

Magnetoresistive memory uses a physical phenomena known as magnetoresistance to store data. A magnetoresistive memory element can be magnetized into one of two possible states. This section discusses magnetoresistance and magnetoresistive memory elements. Magnetoresistance will be discussed first, followed by a detailed description of how magnetoresistive memory elements can be used to store digital data.

Magnetoresistance is based on the magnetoresistive effect. The magnetoresistive effect is the change in the resistance R of a substance when it is subjected to an external magnetic field.[8] The resistance R changes as the angle between the current and the direction of magnetization changes.

The magnetoresistive material used for the memory cells is a magnetoresistive double layer of ternary alloy. It consists of two magnetic layers separated by a thin non-magnetic conducting layer as shown in Figure 2. The magnetic layers are composed of $Ni_{0.65}Fe_{0.15}Co_{0.20}$ and are 100 to 150 Angstroms thick. The conducting layer is tantalum and is 40 to 50 Angstroms thick. The cell exhibits a 2.5% resistance change when subjected to appropriate strong fields.

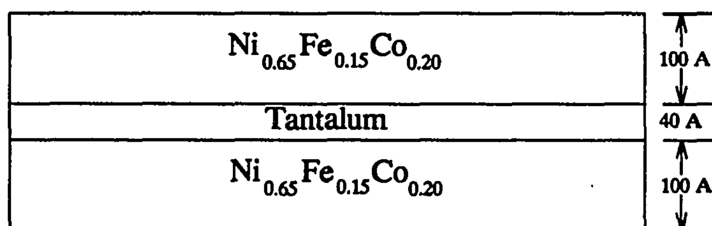

Figure 2. Magnetoresistive double layer

Current expiremental elements are as small as 1.5 um x 3.0 um; with advanced lithography smaller elements can be achieved. Very dense proposed memory cells or elements are shown in Figure 3. The line consisting of the elements is called the sense line. Multiple elements can be on a sense line with shorting metal between them to provide magnetic isolation. The memory elements are 0.6 um wide and 0.6 um long with the easy axis perpendicular to the sense line. The elements are referred to in other literature as transverse elements. The internal magnetization of the element can lie in either direction of the easy axis. In the actual elements the magnetizations of the layers are in opposite directions which allows for flux closure in the individual element. This is also the configuration the element assumes when a strong current flows through the sense line. The word line is a conducting metal and is used to bias the element for reading or writing.

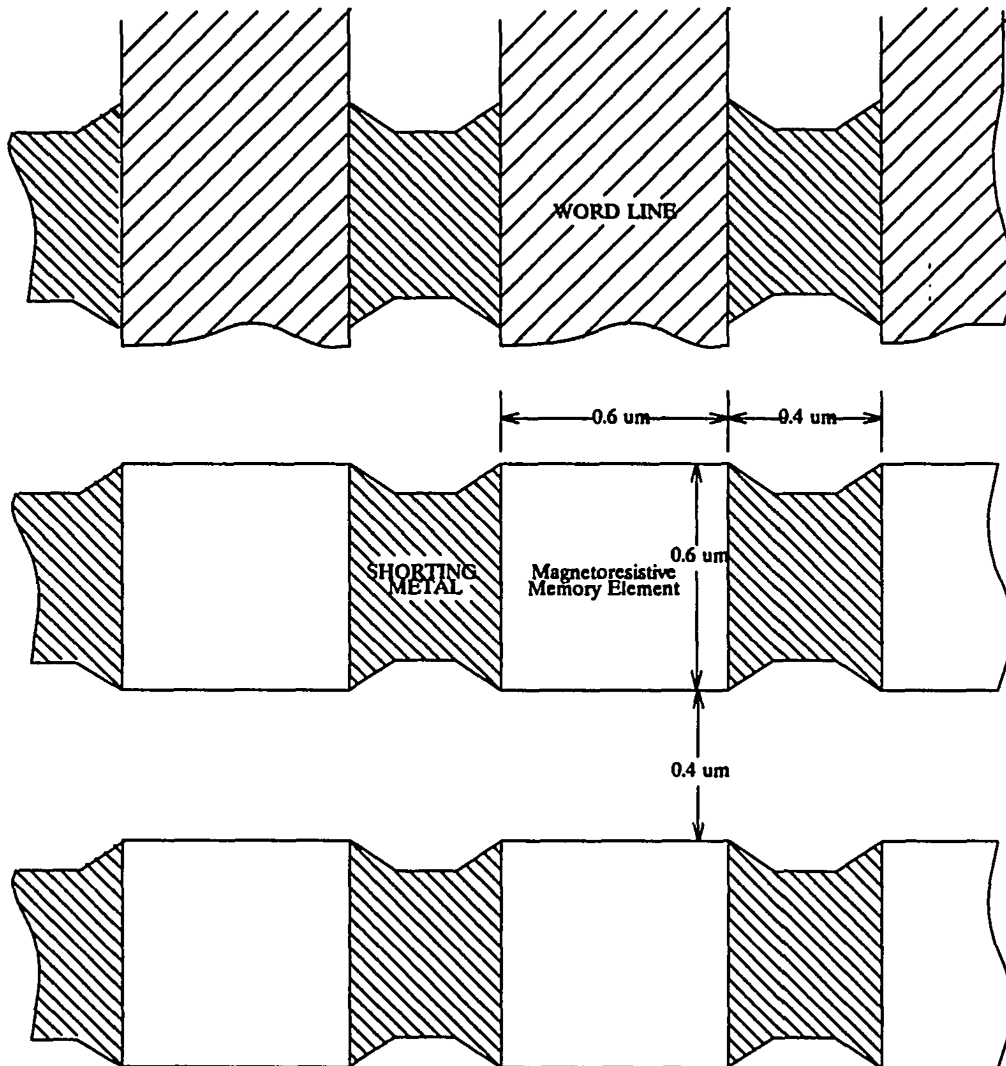

Figure 3. Magnetoresistive memory element

The sense line and word line function to select a single cell in a manner analogous to row and column select in standard random access memory(RAM). The memory cell is written and read in the following manner. The sense line current direction determines whether a "1" or a "0" is being written. The sense current is not set high enough to switch the cell since it is driving multiple cells along the sense line. The proper word line is enabled to provide the required additional field.

Figure 4 shows the Stoner-Wolhfarth threshold for a typical 1.5x5 um cell. When the sense and word currents are below the threshold, the cell is stable in it's current configuration. If the sense and word currents exceed the threshold, the cell will switch into a new state which is determined by the direction of the current in the sense line.

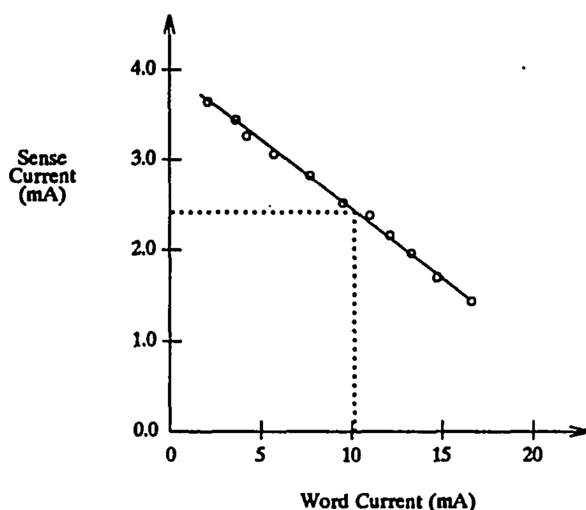

Figure 4. Stoner-Wolhfarth threshold for a 1.5x5 um element

For reading the cell, a moderate sense current is established in a predetermined direction. If the direction of the field from the sense current is the same as the magnetization of the magnetic layers, then the line will be in a low resistance state and the resistance of the sense line will only change slightly as the word field is applied. If they are in opposite directions then the resistance of the line and the output voltage across it will be higher when the word field is applied as shown in Figure 5. Due to the small output voltage, the cell is sampled multiple times and the signal is summed. This improves the signal to noise ratio since the signal is proportional to the number of samples while the noise is proportional to the square root of the number of samples. A 1.5x5 um cell can be sampled at a 10 MHz rate for 1.25 microseconds to yield a signal to noise ratio of 15 and a bit error rate of 10^{-12} or better.[9] An advanced read mode, utilizing a reverse word current, quadruples the signal output. The mode significantly enhances the speed and error rate of the

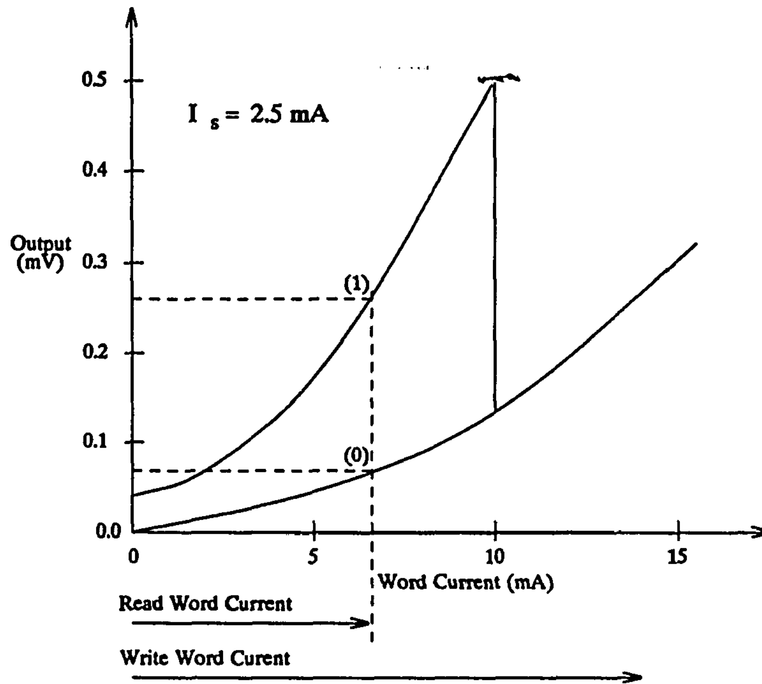

Figure 5. Output curves for a magnetoresistive element

device.

2.4.2 Wafer Scale Application

Magnetoresistive memory appears to be a good candidate for Wafer Scale Integration. The technology is compatible with standard VLSI processing and functional 8K bit ICs have been fabricated by Honeywell. Past wafer scale memory projects have typically failed because they had densities comparable to standard memory or had inefficient redundancy.[10] Magnetoresistive memory handles both problems. Magnetoresistive memory can achieve densities many times greater than standard semiconductor memory because of small bit sizes due to transistor sharing. Semiconductor memory requires at least one transistor per cell which consumes available wafer area. Magnetoresistive memory shares transistors by stringing multiple cells on a sense line. Pohm et al.

recently presented details on theoretical studies for bit densities of 10^8 bits/cm^2 with advanced lithography.[11] Magnetoresistive memory can potentially operate with much lower redundancy than current semiconductor memory due to the fact it is insensitive to alpha-particle induced soft errors.

Magnetoresistive wafer scale memory is an ideal candidate for the gap between conventional semiconductor memory and magnetic or magneto-optical recording. Recording technologies have access times in the 12-100 millisecond range while semiconductor memory typically operates in the 25-100 nanosecond range. This presents a 10^5 speed difference which is difficult for memory hierarchies to bridge effectively. Magnetoresistive wafer scale memory is targeted to operate in the 1-10 microsecond range and will significantly enhance current memory technology.

3. WAFER SCALE INTEGRATION

3.1 Introduction

Wafer Scale Integration is a key technology for development of magnetoresistive memory as a cost-effective gap technology. Wafer scale integration is the integration of millions of IC circuit components on a single wafer. The wafer is not cut into die as is done with traditional VLSI. Instead the wafer is utilized in its uncut form in the final packaging. This chapter examines wafer scale integration technology and describes the critical areas to be addressed in designing a wafer scale integration memory.

3.2 History of Wafer Scale Integration

Wafer scale integration has been attempted on a number of products dating back to a Texas Instrument design in 1966.[10] Table 25 provides a summary of these products. The primary problem associated with the earliest WSI attempts was a lack of redundancy at the end of fabrication while later designs suffer because of low density due to inefficient redundancy methods. The technologies used to utilize redundancy are described in the next section.

3.3 Wafer Scale Integration Technologies

WSI has attempted to use a range of integration technologies. The first products attempted to use discretionary wiring or pad relocation. These techniques utilize a custom metallization layer to interconnect functional cells. This requires a wafer probing before fabrication is complete to determine which cells are functional. The custom mask is then created to connect the functional

TABLE 25. Previous WSI Projects

Name	Originator	Description
Discretionary Wiring	Texas Instruments (1966)	Discretionary wiring of SSI gates.
Pad Relocation Full Wafer LSI	Hughes (1969)	Custom mask to connect SSI/MSI cells.
Wafer Scale Computer	Trilogy (1983)	WSI attempt for 30 MIPs IBM compatible mainframe.
Wafer Scale Hybrid	Mosaic Systems (1985)	CMOS ICs bonded onto aluminum interconnect grid.
Adoptive WSI	McDonnell Douglas (1979)	Operating cells were addressed by non-volatile decoder.
Reconstructable VLSI	MIT/Lincoln Lab (1983)	Orthogonal Array with laser-programmable links.
Wafer Scale RAM	NTT (1984)	256K x 6 RAM using redundancy.
Full Wafer MOS RAM	NTT (1980)	1 Mbit MOS RAM using redundancy.
Cellular VLSI	Hughes (1983)	3D wafer stack.
VHSIC-2 Superchip	TRW (future)	9 square inch chips with over 1 million gates.
WASP	UK 5G Computer Program (future)	Wafer Scale Associative String Processor.
Memory by Configuration Logic	Sinclair Research (1985)	1/2 M byte memory using logic to isolate defects.
Source: Carlson and Neugerbauer.		

cells together. These methods suffer from expensive custom mask generation and the lack of redundancy to compensate for defects which occur in the manufacturing cycle after the wafer probing. These problems were eliminated with the use of more sophisticated redundancy techniques. These techniques, which are the key approaches used today, are described below.

Programmable Reconfiguration: Programmable reconfiguration utilizes on-chip electronically controlled interconnections to allow selection of functional cells after wafer processing has been completed. The original implementation by Manning allowed cells to communicate with nearest neighbors. Subsequently, Aubusson examined more sophisticated arrays which allowed five neighbors instead of three and increased wafer yields. More generalized strategies have since been presented which utilize interconnecting bus structures running along the sides of the cells. The bus design allows a variety of mappings including linear arrays, trees, grids and pyramids.[12]

Redundancy Optimization: Redundancy optimization is concerned with maximizing wafer yield through efficient redundancy design. Redundancy can occur at several levels in the wafer scale design. However, excess redundancy at any level decreases usable wafer area and may cause the design to become economically unattractive.

Laser Restructuring: Laser Restructuring was developed in 1983 by MIT/Lincoln Laboratories. Laser restructuring involves using a laser to link functional cells into the interconnection system. All the cells are originally isolated so they can be tested individually. Then a computer program determines the optimal routing scheme and uses the laser to properly interconnect the cells.

3.4 Current Wafer Scale Integration Products and Projects

WSI has resulted in some current products and many possible products are being investigated. The products which will be discussed in this section are products which are used for computer memory. The Anamartic Wafer Stack and the Inova 1-Mbit SRAM are released products. The

Inova 8-Mbit SRAM is currently under development.

3.4.1 Anamartic Wafer Stack

The Anamartic Wafer Stack is a solid-state memory targeted at the gap between semiconductor memory and disk drives. The Wafer Stack utilizes dynamic RAM technology on six inch wafers. The Wafer Stack uses two six inch wafers to form a 40 MByte module. Each wafer contains 202 one-megabit die. The one-megabit DRAMs utilize a 1.3 um n-well CMOS process and measure 13.65 x 4.4 mm. The DRAM cells are 20 percent larger than traditional DRAM because they are designed with spare cells for redundancy. Each DRAM die also contains configuration logic which allows it to connect with any of its four neighbors. An external controller tests all the die and then programs the configuration logic for each die to form a single bi-directional path through the good die.[13] The Wafer Stack has an access time of 200 microseconds and a 20 megabit per second transfer rate. A 40 MByte module costs $11,680 and a 160 MByte four module system costs $28,760.

3.4.2 Inova 1 Mbit SRAM

The Inova 1-Mbit SRAM is not a true wafer scale product. It is an intermediate product which was developed to facilitate the WSI program at Inova. The wafer actually contains thirty of the 1-Mbit SRAM parts, each of which measures 9.1 x 34.1 mm. This is about six times the size of 1-Mbit dynamic RAM parts which are currently in production. An important attribute of the 1-Mbit SRAM is the redundancy used in the part. The part actually contains 40 blocks. 32 functional blocks are selected and the remaining blocks are isolated by blowing poly fuses. Finally, a non-discretionary metal layer is laid to interconnect the blocks. This final metal layer utilizes a relaxed 20 micron pitch to minimize possible yield loss due to the step. [14]

3.4.3 Inova 8-Mbit SRAM

The Inova 8-Mbit SRAM is a continuation of the 1-Mbit program. The 8-Mbit part is comprised of eight adjacent 1-Mbit parts. The resulting part requires 35.6 x 76.2 mm. The 8-Mbit part is mounted on a composite substrate from Westinghouse to facilitate heat removal. [15]

3.5 Wafer Scale Integration Technologies For Memory

Utilizing WSI for memory products involves the identification of properties which are unique to memory products and tuning the wafer scale design to complement memory attributes while minimizing cost. In memory products it is important to minimize access time and maximize transfer rate for a given capacity. Access time is determined by the memory technology and the memory architecture. Transfer rate is determined by cycle time and memory architecture. Capacity is determined by the wafer area, memory element size and redundancy. Redundancy can occur at a number of levels in the wafer design and it is important to not implement excess redundancy because capacity will be reduced. The following wafer scale design technologies will be the focus of the remainder of this thesis.

Block Architecture	Determines memory speed and transfer rate.
Programmable Reconfiguration	Determines levels of sparing and blocking.
Redundancy Optimization	Determines the yield, cost and capacity of the wafer scale memory.

The design methodology will consist of

1. Designing the block architecture to meet performance goals

2. Modification of architecture to allow self-test and reconfiguration capabilities.

3. Optimizing Redundancy to maximize wafer yields for the given architecture.

Design requirements are determined from Chapter 2 where competing and complementary technologies were evaluated. The requirements are based on 1993 predicted attributes since magnetoresistive memory must be competitive at the time of its release. The requirements are summarized in Table 26.

TABLE 26. Design Requirements for Wafer Scale Magnetoresistive Memory

Attribute	Semiconductor Memory Range	Winchester Disk Range	Wafer Scale MR Requirement
Access Time (seconds)	1e-9 - 200e-9	10e-3 - 90e-3	10e-6
Transfer Rate (Mb/sec)	40e6 - 8000e6	2e6 - 15e6	40e6 - 100e6
Cost per Megabyte ($/MB)	40 - 4000	2 - 10	4 - 20

4. WAFER SCALE DESIGN

The focus of Chapter 4 is examining the range of trade-offs which exist in wafer scale integration(WSI) memory. These trade-offs primarily involve capacity and yield, but other factors such as testability, reconfigurability, power requirements and performance must also be examined.

4.1 Yield versus Capacity Trade-off

Yield on a WSI memory is determined by the number of defects, the distribution of the defects and the redundancy (fault-tolerance) of the architecture. The redundancy on the wafer impacts the capacity of the memory. This is because any area used for redundancy decreases the amount of area available for actual capacity. Therefore, the amounts and locations of redundancy on a WSI memory are critical in determining the capacity and yield. The emphasis of the remaining chapters of this thesis pertains to modelling redundancy variations to optimize the WSI memory architecture. The architecture and redundancy of the WSI memory relate directly to the testability, reconfigurability and performance of the WSI memory. The remainder of this chapter will examine the basic technology involved in testability, reconfigurability, power requirements and performance and how the architecture and redundancy impact these features.

4.2 Testability

Testability refers to the ease and efficiency of testing the WSI memory. The architecture of the memory determines the amount of test time required to fully test the product. Conventional semiconductor memory products are generally tested by algorithms which utilize tests of order $O(N^2)$, $O(N^{3/2})$, or $O(N)$.[16] Testing which requires $O(\sqrt{N})$ have been proposed by Inoue et al.[16] but it requires sophisticated hardware in each memory module to test the words in parallel. The test times for various module sizes are shown in Table 27. Tests which require $O(N^2)$ or

$O(N^{3/2})$ patterns become unreasonable even for small module sizes. Test time costs were estimated by Tuszynski[17] to be four dollars per minute. Table 28 shows testing costs as a function of module size.

TABLE 27. Test times for memory module testing

Module	All times in milliseconds			
Size(K)	O(1/2)	O(1)	O(3/2)	O(2)
1	0.3	10.2	327.7	10485.8
4	0.6	41.0	2621.4	167772.2
16	1.3	163.8	20971.5	2684354.5
64	2.6	655.4	167772.2	42949672.0
256	5.1	2621.4	1342177.2	687194752.0
1024	10.2	10485.8	10737418.0	10995116032.0
4096	20.5	41943.0	85899344.0	175921856512.0
16384	41.0	167772.2	687194752.0	2814749704192.0
65536	81.9	671088.6	5497558016.0	45035995267072.0
262144	163.8	2684354.5	43980464128.0	720575924273152.0
1048576	327.7	10737418.0	351843713024.0	11529214788370432.0
4194304	655.4	42949672.0	2814749704192.0	184467436613926910.0

TABLE 28. Test costs of memory module testing

Module	All costs in dollars			
Size(K)	O(1/2)	O(1)	O(3/2)	O(2)
1	0.000	0.001	0.022	0.699
4	0.000	0.003	0.175	11.185
16	0.000	0.011	1.398	178.957
64	0.000	0.044	11.185	2863.312
256	0.000	0.175	89.478	45812.984
1024	0.001	0.699	715.828	733007.750
4096	0.001	2.796	5726.623	11728124.000
16384	0.003	11.185	45812.984	187649984.000
65536	0.005	44.739	366503.875	3002399744.000
262144	0.011	178.957	2932031.000	48038395904.000
1048576	0.022	715.828	23456248.000	768614334464.000
4194304	0.044	2863.312	187649984.000	12297829351424.000

MR memory testing can be conducted with tests of length $O(N)$. This is because MR memory elements do not exhibit the coupling faults, also called pattern sensitive faults(PSFs) which occur in semiconductor RAM. By testing MR memory cells in parallel, test time and cost can be

further reduced. The following calculations estimate test time for a two gigabit wafer with a data bus of 64 bits.

$$\begin{aligned}
Capacity &= N = 2\,Gb\,(= 2 \times 10^9) \\
Bus\ Width &= 64\,bits \\
Read\ Cycles &= Write\ Cycles = 2N \\
Read\ Cycle\ Time &= 1\,usec \\
Write\ Cycle\ Time &= 0.2\,usec
\end{aligned}$$

$$Test\ Time = \frac{Write\ Cycles}{Bus\ Width} \times Write\ Cycle\ Time + \frac{Read\ Cycles}{Bus\ Width} \times Read\ Cycle\ Time$$

$$TestTime = \frac{4 \times 10^9}{64} \times (0.2 \times 10^9) + \frac{4 \times 10^9}{64} \times (1.0 \times 10^9)$$

$$Test\ Time = 75\,seconds$$

$$Test\ Cost = \frac{75\,seconds}{60\,seconds\,/minute} \times 4.00\,dollars\,/minute = \$5.00$$

These calculations demonstrate that testing can occur in a reasonable amount of time and at a reasonable cost. Some incremental time would be required to reconfigure the memory to spare out defective bits. The number of bits to be spared is approximately 2×10^5 which implies that 0.25 seconds is required for writing and verifying the spare addresses. An additional second would be required to verify that all spares are defect free.

4.3 Reconfigurability

Reconfigurability is the capacity for reconfiguring the memory to spare out defective sectors. The hardware for allowing reconfiguration can be either programmable cells, laser fuses or polysilicon fuses. All three hardware implementations store the address of the defective bit, row or column. A compare circuit continuously compares the defect address against the address on the bus. When a match occurs, the defective element is disabled and the spare element is enabled for the

memory operation. Programmable cells are either loaded from the system after power-up or read internally stored addresses on power-up. The stored addresses were written during the original configuration. Laser and polysilicon fusing utilize fuses to pull-down the input line to a predetermined state(0). Blowing the fuse disconnects the line from the original pull-down and the input line is pulled-up to assume the alternate state(1). Laser fusing utilizes a laser to blow fuses while polysilicon fuses are blown electronically by high currents.

The wafer scale MR memory will utilize programmable cells for the following reasons,

1. Programmable cells require very small areas compared to laser fuses.
2. Programmable cells utilizing triple redundancy bits have a low failure rate of 10^{-8} or less. Laser and polysilicon fuses have substantially higher failure rates.
3. Programmable cells do not cause physical damage on the silicon when they are programmed. Laser and polysilicon fuses require significant heat and may cause surface damage.
4. Programmable cells can be written and verified quickly. Laser fuses require long cycle times for locating and blowing the fuse.
5. Programmable cells can be reconfigured at a later time. Laser fuses can only be programmed in a factory environment. Fuses are not reconfigurable after they have been blown.
6. MR memory technology easily supports programmable cells. Laser and polysilicon fuses require additional mask steps for constructing the fuse structures.

4.4 Power Requirements

Power requirements and the implied heat dissipation are important considerations in wafer scale projects. MR memory requires less power than dynamic RAM or static RAM because MR memory cells are non-volatile. MR memory cells only require current when they are actively being

written or read. The address flip-flops and address comparators require small amounts of current constantly. The power requirements for the MR wafer scale memory are estimated below.

Current requirements for each active 1 Mb module:

15 mA Word Current

8 mA Sense Current (8 lines @ 1 mA)

10 mA Amplifiers and Support Circuitry

----

33 mA Total

x 8 1 Mb Modules Active simultaneously

----

264 mA Current for Active Modules

0.5 mA Flip-flops and Comparators for all 1 Mb Modules

x 2048 Total number of 1 Mb Modules

1024 mA Current demand for address comparing

Total Wafer demand = Active Modules + Address Compare Current

Total Wafer Current = 1.288 Amperes

Total Wafer Power = 6.44 Watts

The MR memory wafer's current requirement of 1.3 Amperes and power requirement of 6.5 Watts are very low for a wafer scale product. In comparison, Anamartic's 20 megabyte wafer requires 8 Amperes or 42 Watts.[18]

4.5 Performance

The performance of the MR memory can be expressed in terms of access time and transfer rate. Chapter 3 established the 1993 requirements of the MR memory as 1-10 microsecond access time and 40-100 megabit per second transfer rate. Access time depends on the cycle time of the MR memory cell. Transfer time depends on the cycle time of the MR memory cell and the number of cells being accessed in parallel.

The performance goals of the MR memory can be achieved by designing cells which have a one microsecond cycle time and by accessing 64 bits in parallel. This provides a one microsecond access time and a 64 megabit per second transfer rate. The one microsecond access time can be achieved using new MR material which demonstrates a four percent MR effect and by utilizing the reverse read process which yields a non-destructive read signal which is six times greater than the non-destructive read signal from the forward read process. The method for accessing 64 bits in parallel is discussed in the next section.

4.6 Architecture and Redundancy

Determining the optimal redundancy for a WSI memory is a complex problem. There are several types of redundancy and many different architectural levels where the redundancy could be implemented. Defects in semiconductors exhibit random distributions with a degree of clustering. Clustering is the tendency for multiple defects to be found in groups at random locations. For wafer scale integration, multiple levels of redundancy are required to compensate for both the random distribution and the clustering phenomena.

Redundancy at any level reduces the capacity of the product. When redundancy occurs at any level, there is an areal cost associated with it. When redundancy occurs at low levels, the actual

redundancy may require minimal area. However, when this area is multiplied by the number of units on the entire wafer, the areal cost may become significant. The impact of redundancy on two architectures is examined below. The two architectures are the binary tree structure and the bus structure.

The binary tree architecture implies that 2 units at level n-1 combine to form a unit at level n. The usable area at any level with redundancy is on 67 percent of the area available without redundancy. The amount of usable area out of the total wafer area then is $(2/3)^m$ where m is the number of binary levels with redundancy. Table 29 examines the percent of the wafer usable for actual capacity when the architecture is a binary tree. Clearly, a binary tree with redundancy is very inefficient.

TABLE 29. Redundancy costs for binary tree structure

Number of Levels with Redundancy	Percentage of Total Area for Memory Capacity
0	100.00
1	66.67
2	44.44
3	29.63
4	19.75
5	13.17
6	8.78
7	5.85
8	3.90

The architecture at the opposite end of the spectrum from the binary tree would be the bus structure. In the bus structure all of the modules on the wafer are hooked to a common bus. Through some type of programmable structures, the wafer can select which of the modules correspond to which address. When redundancy is performed at a single level, there must be sufficient spares to compensate for all possible defect patterns. For MR memory, the defect level is expected to be relatively high for the MR elements. Therefore the number of defects per module will be quite high and many spares will be required for each module. With the high failure rate of the MR elements, single level redundancy is not practical.

For wafer scale integration, multiple levels of sparing are required to provide reasonable yield. Redundancy at a low level is required to compensate for isolated defects. Redundancy at a higher level is required to compensate for clusters of defects. The sparing on the MR wafer scale memory will occur at three levels. The low level sparing occurs at the 16Kb level. The goal of the low level sparing is to compensate for scattered defects. The next level of sparing occurs at the 1MB level with the goal of compensating for clustered defects. The third level of sparing is at the system bus level. The goal of the system level sparing is to compensate for large clusters of defects which cannot be handled at the 1 Mb level. Figure 6 shows the sparing for the two lowest levels.

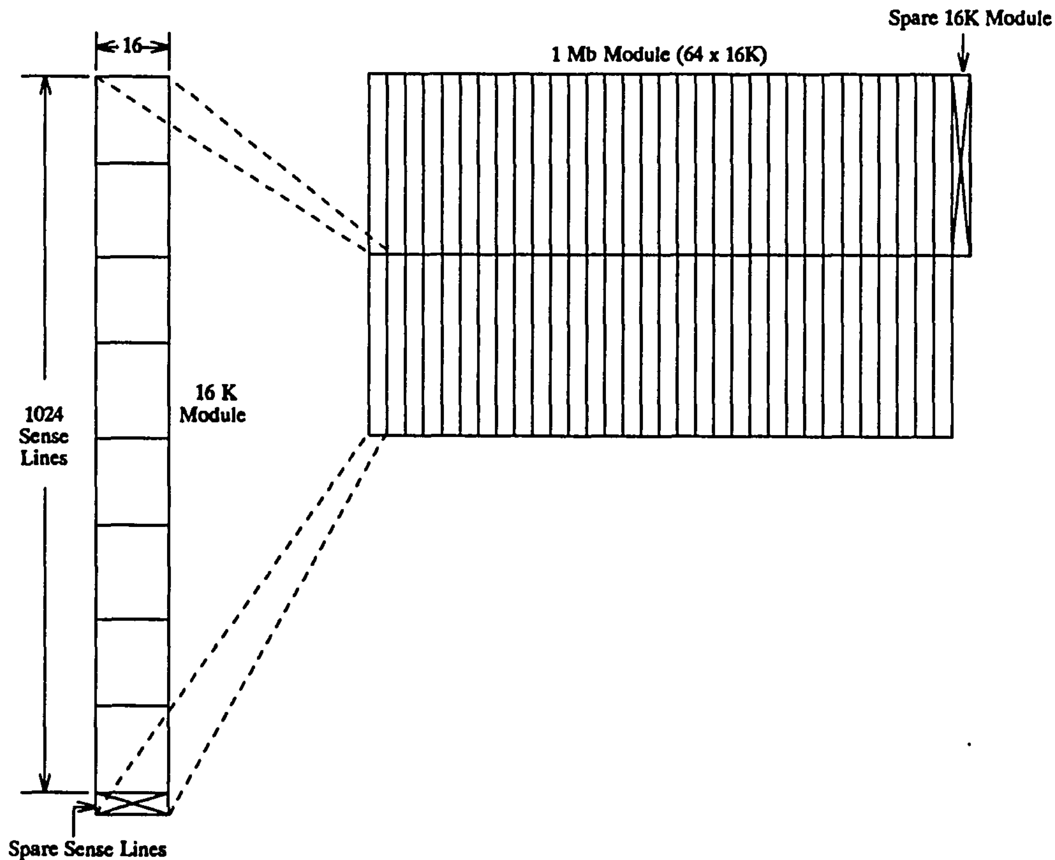

Figure 6. Redundancy locations for lowest levels

Sparing at the system bus level will occur in the following manner. Every 1Mb module will have a default test address and a programmable address. In test mode, each module is accessed via its unique test address. After the wafer has been tested and all functional modules have been identified, the programmable addresses of the functional modules will be written to form a contiguous memory space. The programmable addressing of the 1 Mb modules has several advantages,

1. Programmable addressing allows the memory capacity to increase when very few modules are defective. Traditional spares are not utilized in memory space.
2. Programmable addressing allows defective modules to occur in any pattern. Traditional sparing assigns certain spares to certain regions of the media. Defect patterns which require more than the available regional spares cause the sparing to fail even though the number of total spares is greater than the total number of defects.

4.7 Address Decoding

The programmable address of the 1 Mb module is thirteen bits long. The top ten bits define groups of eight 1 Mb modules. The eight 1 Mb modules in each group access data in parallel. Each 1 Mb module internally accesses eight bits of data in parallel from a single 16 Kb module. The eight bits are placed on the wafer's 64 bit data bus at the byte location determined by the three least significant bits of the 1 Mb module's programmable address. The address decoding of the host address is shown in Figure 7.

One consequence of the addressing scheme is that the number of 1 Mb modules which can be utilized on the wafer must be a multiple of eight. Accordingly, up to seven of the wafer's functional

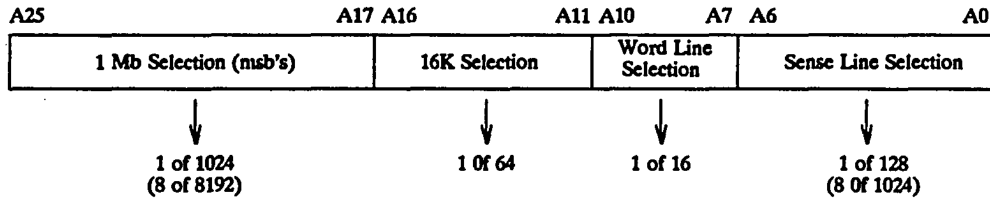

Figure 7. Address decoding of host address

1 Mb modules may be unutilized. If higher transfer rates are required, the addressing can be modified to access sixteen or 32 modules in parallel and then cycle the data out during the next read cycle. An alternative solution for higher transfer rates is to use wafers in parallel. This will require that the wafers all have the same utilized capacity.

5. DESIGN SIMULATION

5.1 Introduction

The design simulation uses a computer-based model to estimate the yield and capacity of various redundancy combinations for the wafer scale architecture. The computer-based model calculates the effects of the redundancy for a range of defect densities and clustering parameters.

5.2 Overview of Model

The model will calculate the wafer capacity and yield using three yield models, one for each level of redundancy. The three yield models are called the 16-Kb model, the 1-Mb model and the wafer yield model. The 16-Kb model calculates the yield for the 16-Kb modules using structure, redundancy, defect level and clustering level as inputs. The output of the 16-Kb model is used as the input of the 1-Mb model along with the amount of redundancy at the 1-Mb level. The 1-Mb model calculates the yield of the 1-Mb modules. The wafer level model calculates the capacity and yield of the wafer given the 1-Mb module yield as an input. A block diagram of the simulation is shown in Figure 8.

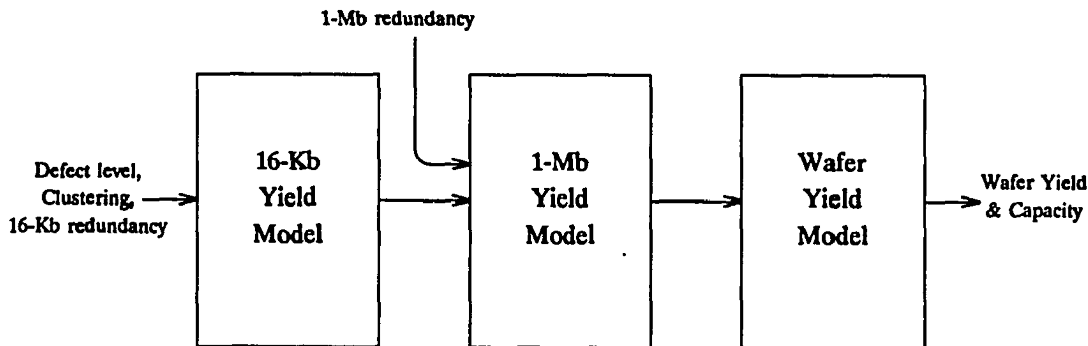

Figure 8. Block diagram of simulation

5.3 16-Kb Model

The 16-Kb model utilizes the negative binomial model which was proposed by Stapper et al.[19] and is widely used in VLSI memory yield calculations. The negative binomial model calculates yield based on defect frequency and the degree of clustering. The equation for yield in the negative binomial model is

$$Y = \left(1 + \frac{\lambda}{\alpha}\right)^{-\alpha} \tag{5.1}$$

where λ = *defect frequency*
α = *clustering parameter*

The negative binomial yield equation can be extended to combine multiple types of defects. When the equation is expanded to accommodate multiple types of defects it becomes

$$Y = \prod_{i=1}^{n} \left(1 + \frac{\lambda_i}{\alpha_i}\right)^{-\alpha_i} \tag{5.2}$$

where n = *number of types of defects*
λ = *defect frequency* for *defect type n*
α = *clustering parameter* for *defect type n*

The 16-Kb model uses two types of defects. The first type of defect is called the Sense Line Kill (SLK) defect and consists of defects which can render a sense line or a portion of a sense line unusable. SLK defects can be spared if a spare sense line is available in the 16-Kb module. The second type of defect is the Chip Kill (CK) defect. A CK defect cannot be spared so the 16-Kb module is killed or rendered unusable. Chip kill defects are defects which render non-spareable portions of the 16-Kb module unusable. Chip kill defects include amplifier circuit defects, word line defects, address decoding defects and defects in other circuitry which is critical for proper functioning and which does not have redundant circuits for replacement.

For modeling, MR elements are assumed to occupy 51 percent of the area while electronics

occupy the remaining 49 percent. It is also assumed that 50 percent of the defects in the electronics result in CK defects while the remainder result in SLK defects.

For the 16Kb model equation 5-2 can be rewritten as

$$Y_{16K} = Y_{CK}\, Y_{SLK} \tag{5.3a}$$

where

$$Y_{CK} = \left(1 + \frac{\lambda_{CK}}{\alpha_{CK}}\right)^{-\alpha_{CK}} \tag{5.3b}$$

$$Y_{SLK} = \left(1 + \frac{\lambda_{SLK}}{\alpha_{SLK}}\right)^{-\alpha_{SLK}} \tag{5.3c}$$

If redundancy is added for compensating for SLK defects, equation 5.3c can be replaced by Harden and Strader's equation for circuits with redundancy

$$Y_{SLK} = \sum_{i=0}^{N-M} (-1)^{N-M+i} \binom{N}{N-i} \binom{N-1-i}{M-1} \left(1 + (N-i)\frac{\lambda_{SLK}}{N\alpha_{SLK}}\right)^{-\alpha_{SLK}} \tag{5.4}$$

where

$N\text{-}M$ = *number of redundant sense lines*
M = *number of required sense lines.*

which reduces to equation 5.3c if N=M. For the computer model, $\alpha_{SLK} = \alpha_{CK}$, which assumes that the level of clustering in the CK defects is the same as the level of clustering in the SLK defects.

The computer model uses the following areal calculations which can be achieved with 0.8 μm lithography.

16-Kb area = $0.108\, mm^2$

electronics area = $0.0530\, mm^2$
CK area = $0.0265\, mm^2$
SLK area = $0.0265\, mm^2$

number of MR elements = 16,384

$$\lambda_{SLK} = 16{,}384 \cdot D_{MR} + (SLK\ area) \cdot D_{EL} \tag{5.5}$$

$$\lambda_{CK} = (CK\ area) \cdot D_{EL} \tag{5.6}$$

The range of MR defects which will be simulated is from 10^{-4} to 10^{-6} defects per MR element. The defect densities for electronic circuits which will be simulated are from 10 defects per cm^2 to 0.1 defects per cm^2. The range of α to be simulated is from 0.1 to 10. Stapper reports ranges for α of 0.3039 to 2.382 in dynamic RAM integrated circuits.[20] Moore recommends that fault-tolerant circuit designers consider a defect range of 0.01 to 0.15 defects per mm^2 since most processes run between 0.02 and 0.08 defects per mm^2.[21] These values correspond to 1 to 15 defects per cm^2 for design tolerance and 2 to 8 defects per cm^2 for current process defect densities.

5.4 1-Mb Model

The 1-Mb model simulates the combination of 16-Kb modules to form the 1-Mb module. A functional 1-Mb module requires at least 64 functional 16-Kb modules. The 1-Mb module will typically contain one or more spare 16-Kb modules to allow for the replacement of defective 16-Kb modules. The 1-Mb model will utilize the binomial model which has been used by Stapper et al.[22] and others to model redundant circuits. The yield of a module without redundancy which uses M circuits is

$$Y = (Y_{16K})^M \tag{5.7}$$

With R redundant circuits the yield is

$$Y = \sum_{n=0}^{R} \frac{(M+R)!}{(M+R-n)!\,n!} Y_{16K}^{(M+R-n)} (1 - Y_{16K})^n \tag{5.8}$$

Figure 9 shows the effectiveness of redundancy for a circuit where M=64.

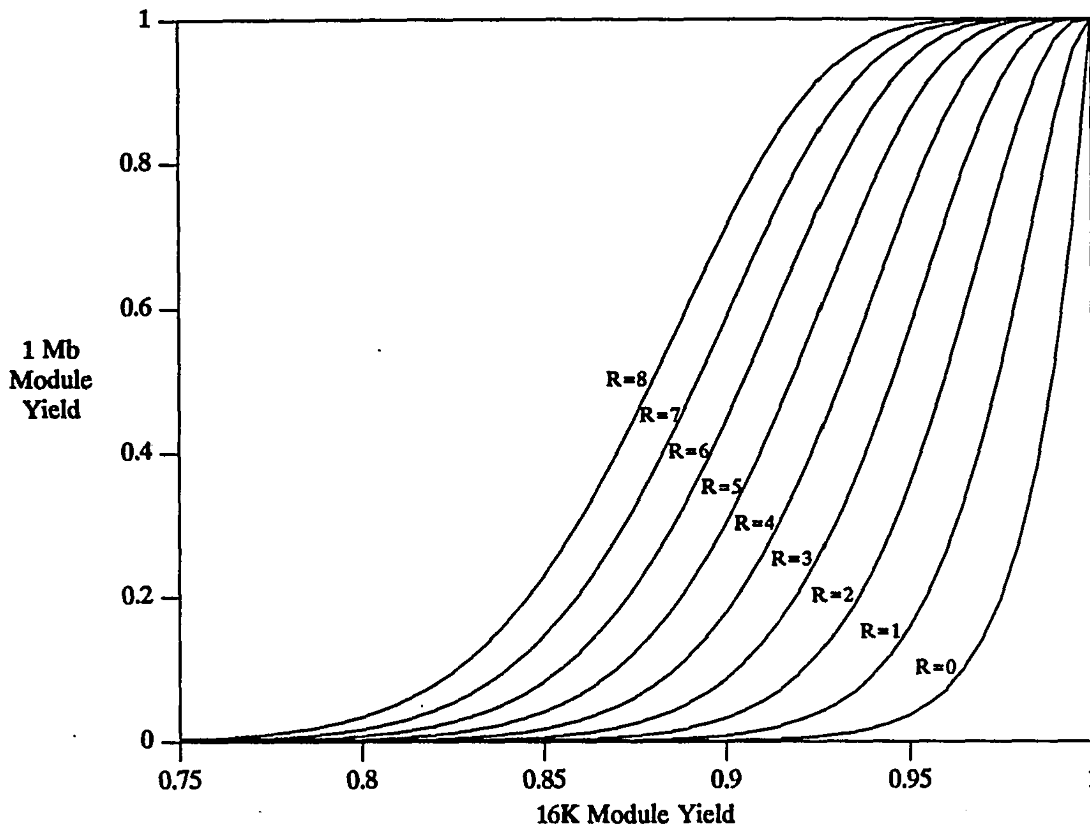

Figure 9. Effectiveness of Redundancy for M=64

5.5 Wafer Yield Model

The wafer yield model uses the 1-Mb yield and the redundancy information to calculate the capacity and yield of the wafer. The redundancy information is important because it determines how many 1-Mb modules will fit on the wafer. The redundancy at the 16-Kb and 1-Mb levels has an areal cost associated with it which detracts from the total capacity of the wafer. The number of 1-Mb modules available when no redundancy is used on the wafer is 2224. The presence of redundancy will reduce this number. The areal cost for the redundancy at each level is shown in Table 30 and 31. The cost is the number of redundant modules divided by the total number of modules. In the case of 16-Kb redundancy, the cost is multiplied by a factor of ten to include the cost of the additional circuitry to allow for sparing. At the 1-Mb level, this additional circuitry is

TABLE 30. Areal cost of redundancy at 16-Kb level

Redundant Circuits	Areal Cost (in percent)
0	0.00
1	0.98
2	1.95
3	2.92
4	3.89
5	4.86
6	5.83
7	6.79
8	7.75

TABLE 31. Areal cost of redundancy at 1-Mb level

Redundant Circuits	Areal Cost (in percent)
0	0.00
1	1.54
2	3.03
3	4.48
4	5.88
5	7.25
6	8.57
7	9.86
8	11.11

negligible.

For a redundancy of 4 at the 16-Kb level and a redundancy of 2 at the 1-Mb level which is designated 4_2, the reduction in 1 Mbit modules can be calculated as

$$\textit{Useable Area} = (1 - AC_{16K}) \cdot (1 - AC_{1M})$$

where

$$AC_{16K} = \textit{Areal cost of redundancy at the 16-Kb level.}$$
$$AC_{1M} = \textit{Areal cost of redundancy at the 1-Mb level.}$$

For 4_2,

$$AC_{16K} = 0.0389$$
$$AC_{1M} = 0.0303$$

$$\textit{Useable Area} = (1 - 0.0389) \cdot (1 - 0.0303) = 0.932$$

$$Number\ of\ 1\text{-}Mb\ Modules = (0.932)\cdot(2224) = 2072$$

The percentage of useable area and the number of 1-Mb modules can be calculated for all redundancy combinations. Tables 32 and 33 show this for a range of redundancy combinations.

TABLE 32. Percent of Usable Area at Wafer level

	16K Redundancy								
1-Mb Redundancy	0	1	2	3	4	5	6	7	8
0	1.000	0.990	0.981	0.971	0.961	0.951	0.942	0.932	0.922
1	0.985	0.975	0.965	0.956	0.946	0.937	0.927	0.918	0.908
2	0.970	0.960	0.951	0.941	0.932	0.923	0.913	0.904	0.895
3	0.955	0.946	0.937	0.927	0.918	0.909	0.900	0.890	0.881
4	0.941	0.932	0.923	0.914	0.905	0.895	0.886	0.877	0.868
5	0.928	0.918	0.909	0.900	0.891	0.882	0.874	0.865	0.856
6	0.914	0.905	0.896	0.888	0.879	0.870	0.861	0.852	0.843
7	0.901	0.893	0.884	0.875	0.866	0.858	0.849	0.840	0.832
8	0.889	0.880	0.872	0.863	0.854	0.846	0.837	0.829	0.820

TABLE 33. Number of 1-Mb Modules on Wafer

	16K Redundancy								
1-Mb Redundancy	0	1	2	3	4	5	6	7	8
0	2224	2202	2180	2159	2137	2115	2094	2073	2051
1	2189	2168	2147	2125	2104	2083	2062	2041	2020
2	2156	2135	2114	2093	2072	2051	2030	2010	1989
3	2124	2103	2083	2062	2041	2021	2000	1980	1959
4	2093	2072	2052	2032	2011	1991	1971	1951	1930
5	2062	2042	2022	2002	1982	1962	1942	1922	1902
6	2033	2013	1993	1973	1954	1934	1914	1895	1875
7	2004	1985	1965	1946	1926	1907	1887	1868	1849
8	1976	1957	1938	1919	1899	1880	1861	1842	1823

Because all the 1-Mb modules have programmable addresses, they can always be programmed into a range of addresses to provide the optimal capacity. Normally the highest capacity which can be achieved is the optimal capacity. For a set of M modules with yield Y, the probability of N modules being usable is

$$Y_{N,M} = \sum_{n=0}^{M-N} \frac{(M)!}{(M-n)!\,n!} Y^{(M-n)}(1-Y)^n \tag{5.9}$$

this is equation 5.8 modified to accommodate the proper indices. Since the wafer operates with sets of eight modules in parallel, the yield equation must be modified to incorporate this fact. The probability that the wafer will yield a usable capacity of exactly 8X megabits (which includes 8X, 8X+1, .., 8X+7) is

$$Y_{Cap=8X} = \begin{cases} \sum_{n=8X}^{8X+7} \frac{(M)!}{(M-n)!\,n!} Y^{(M-n)}(1-Y)^n & \text{for } M \geq 8X+7 \\ \sum_{n=8X}^{M} \frac{(M)!}{(M-n)!\,n!} Y^{(M-n)}(1-Y)^n & \text{for } 8X \leq M \leq 8X+7 \\ 0 & \text{for } 8X > M \end{cases} \tag{5.10}$$

The probability that M modules will yield a product with 8X or greater megabits of usable capacity is

$$Y_{Cap \geq 8X} = \sum_{i=X}^{\infty} Y_{Cap=8X} \tag{5.11}$$

A distribution of capacities can be determined for a given set of defect densities, α and redundancy. The mean and standard deviation can be easily calculated because the wafer yield model is a binomial distribution.

$$Mean = \mu = NY$$

$$Standard\ Deviation = \sigma = \sqrt{NY(1\text{-}Y)}$$

The mean of the distribution is an important figure of merit because it determines the average capacity for the wafer for the specified parameters. The standard deviation is also important since it indicates the uniformity of the yield of the wafers. When the capacity distribution is evaluated over the entire range of defect densities and alpha, the effectiveness of the redundancy can be estimated. The effectiveness of the redundancy can be determined by multiplying a 2D distribution of the manufacturing process times the 2D distribution of the capacity. Several hypothetical process distributions will be considered in Chapter Six.

6. SIMULATION RESULTS

6.1 Introduction

The simulation results are difficult to evaluate in the raw data form. The following sections will describe how to interpret the data and will discuss limitations of the model.

6.2 Limitations and Approximations

The most serious limitation in the yield simulation is the floating point accuracy of the computer. The computer used for the simulation was an HP 375 workstation with a 68030 processor and a 68882 floating point co-processor. The C programming language and the 68882 limit the simulation to 16 significant decimal digits. This limitation becomes serious when trying to calculate 16-Kb yields for five or more spares. The equation for 16-Kb yield (Eq 5.8) results in an alternating series where the combinational multipliers exceed 1×10^{17}. The resulting yield, which should be between zero and one, is inaccurate due to floating point errors. This is because floating point subtraction of nearly equal quantities can very greatly increase relative error.[23] Several software solutions were investigated but the fundamental problem of limited significant digits remained because of no method of calculating $(1+x)^{\alpha}$ could be implemented for the ranges of x and α.

An approximation for the 16-Kb yield can be made for five to eight spares if the series uses a lower module value. Using a value of 64 instead of 1024 resulted in a very accurate approximation. Values for 16-Kb yield calculated using 64 modules were within one percent of the 16-Kb yield using 1024 modules for zero to four spares. The values were compared over the ranges of $(10^{-6}\leq\lambda\leq 10^{-4})$ and $(0.1\leq\alpha\leq 10)$ which can be achieved in the manufacturing process. The software program which compares the calculated values can be found in Appendix A along with the corresponding results.

6.3 Interpreting the Results

The results of the simulation are difficult to interpret in the raw form. The data can be comprehended more readily if a simplification is made and the data are presented in a graphical format.

An important simplification occurs by neglecting the electronic circuit defect density. Table 34 shows the effects of the defect densities and α as they are varied by a factor of 10. The effects of the MR defect density and α are very large relative to the effects of electronic circuit defect density. Since λ_{EL} had such a minor impact on yield, all subsequent simulations used the worst case λ_{EL} of 10 defects per cm^2. This simplification decreased simulation time and complexity.

TABLE 34. Yields vs. Parameters

Alpha	MR element defect rate	Electronic defect density	Yield of 16-Kb Modules	Yield Impact (in percent)
Spares=0				
0.1	0.00010000	0.100	0.7495	0.00
0.1	0.00010000	0.010	0.7514	0.25
0.1	0.00001000	0.100	0.9043	20.65
1.0	0.00010000	0.100	0.3776	49.62
Spares=5				
0.1	0.00010000	0.100	0.9147	0.00
0.1	0.00010000	0.010	0.9169	0.24
0.1	0.00001000	0.100	0.9951	8.79
1.0	0.00010000	0.100	0.9447	3.28

By neglecting λ_{EL} the data can be displayed graphically for the various redundancy combinations on a two dimensional graph. The horizontal axis of the graph corresponds to α while the vertical axis corresponds to λ_{MR}. The ranges of α and λ_{MR} cover the expected manufacturing process ranges.

The effectiveness of the redundancy can be determined if the characteristics of the manufacturing process can be estimated. In such a case, the product of a 2D distribution of the process multiplied by the 2D distributions of capacity mean will determine the yield with redundancy.

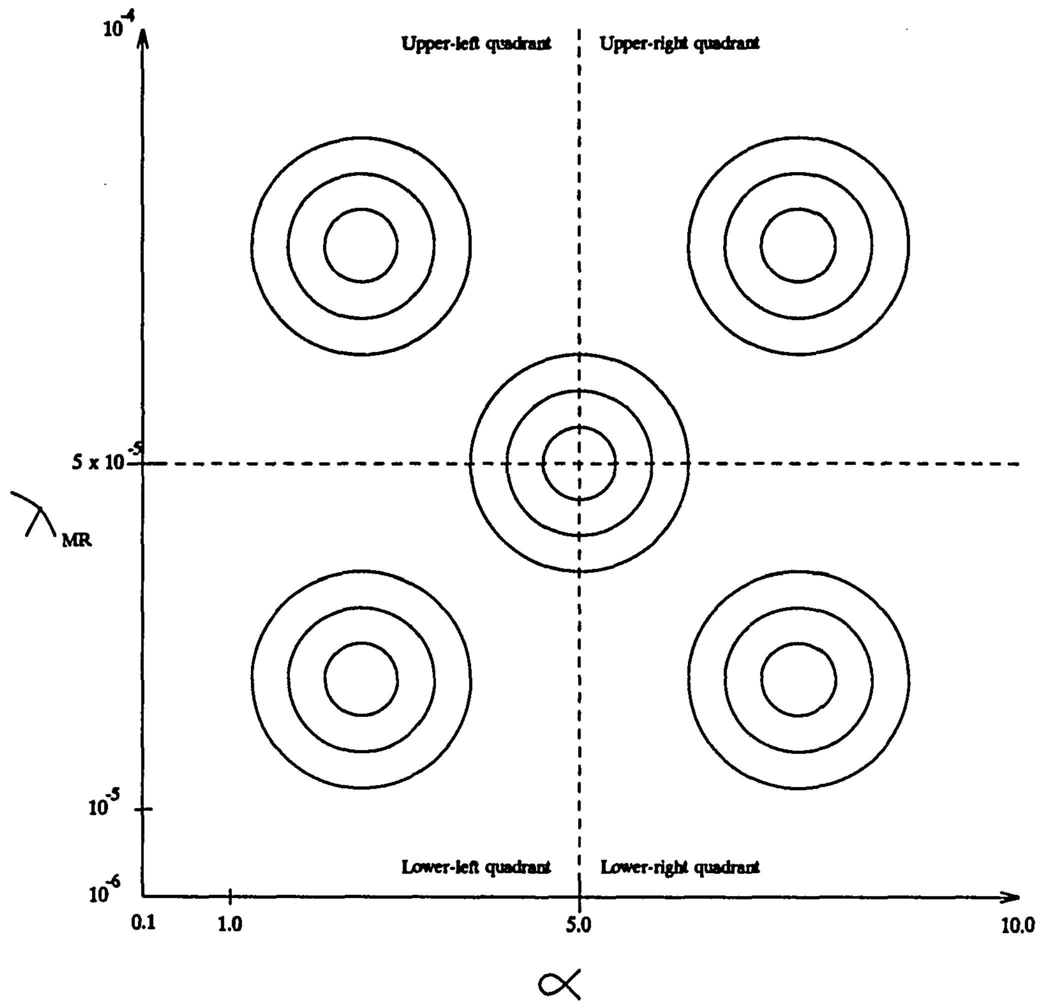

Figure 10. Manufacturing Process Range and Distributions

Figure 10 shows a graph of the manufacturing process range. This range can be divided into four quadrants. The characteristics of the quadrants are listed below.

Upper-Left Quadrant - Relatively dirty process with a high level of clustering.

Upper-Right Quadrant - Relatively dirty process with a low level of clustering.

Lower-Left Quadrant - Relatively clean process with a high level of clustering.

Lower-Right Quadrant - Relatively clean process with a low level of clustering.

Figure 10 also shows the distributions which will be used to evaluate the simulation results. The circles indicate intervals of one standard deviation from the center of the circle. The yields which occur with these distributions are listed below.

Mean yield - The yield if the manufacturing process is distributed evenly over the ranges of λ_{MR} and α.

Centered-yield - The yield if the manufacturing process is normally distributed over the process range, with the distribution centered in the middle of the range and standard deviations of $\delta_\alpha = 0.5$ and $\delta_\lambda = 5 \times 10^{-6}$.

Lower-left-centered-yield - The yield if the manufacturing process is normally distributed over the lower left quadrant of the process range, with the distribution centered in the middle of the quadrant and standard deviations of $\delta_\alpha = 0.5$ and $\delta_\lambda = 5 \times 10^{-6}$.

Lower-right-centered-yield - The yield if the manufacturing process is normally distributed over the lower right quadrant of the process range, with the distribution centered in the middle of the quadrant and standard deviations of $\delta_\alpha = 0.5$ and $\delta_\lambda = 5 \times 10^{-6}$.

Upper-left-centered-yield - The yield if the manufacturing process is normally distributed over the upper left quadrant of the process range, with the distribution centered in the middle of the quadrant and standard deviations of $\delta_\alpha = 0.5$ and $\delta_\lambda = 5 \times 10^{-6}$.

Upper-right-centered-yield - The yield if the manufacturing process is normally distributed over the upper right quadrant of the process range, with the distribution centered in the middle of the quadrant and standard deviations of $\delta_\alpha = 0.5$ and $\delta_\lambda = 5 \times 10^{-6}$.

The simulation utilizes increments of 0.1α and $10^{-6}\lambda_{MR}$ over the ranges of $(10^{-6} \leq \lambda \leq 10^{-4})$ and $(0.1 \leq \alpha \leq 10)$. This results in a matrix of 100 by 100 (or 10,000) calculated values. The mean yield is calculated by multiplying the yield of each matrix point by $\frac{1}{10000}$. The

remaining yields utilize a normal distribution and calculate the yield by multiplying the matrix of yield values by a statistical distribution matrix. The statistical distribution matrix can be thought of as a 100 x 100 matrix although only a range of the values, a 39 x 39 matrix, are non-zero. In the program, the statistical distribution matrix is modeled using a vector. The vector has 39 elements which contain the percentage of the normal distribution centered at the point $\pm 0.1\delta$. Figure 11 shows the normal distribution divided into a series of discrete values. Table 35 lists the discrete values.

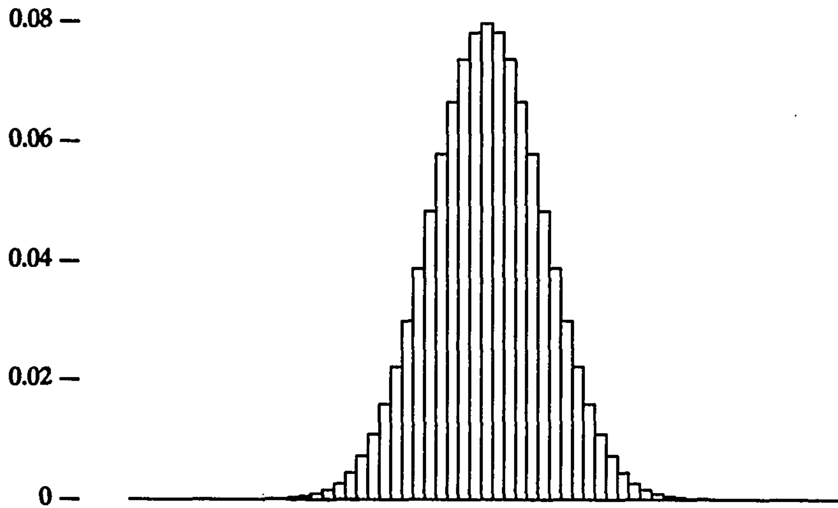

Figure 11. Normal Distribution Used in Simulation

6.4 Numerical Results

Tables 36 and 37 list the numerical results of the simulation for the level one and level two redundancies of zero through eight. The software program which simulates the yield models and calculates the yields for the distributions is listed in Appendix B. The data from the simulation are displayed graphically in Appendix C.

TABLE 35. Values of Normal Distribution

Offset from center of Distribution	Amount of Normal Distribution in Matrix Element
-20	0.0000
-19	0.0001
-18	0.0001
-17	0.0003
-16	0.0005
-15	0.0009
-14	0.0016
-13	0.0027
-12	0.0045
-11	0.0072
-10	0.0108
-9	0.0159
-8	0.0222
-7	0.0300
-6	0.0389
-5	0.0484
-4	0.0579
-3	0.0665
-2	0.0736
-1	0.0781
0	0.0796
1	0.0781
2	0.0736
3	0.0665
4	0.0579
5	0.0484
6	0.0389
7	0.0300
8	0.0222
9	0.0159
10	0.0108
11	0.0072
12	0.0045
13	0.0027
14	0.0016
15	0.0009
16	0.0005
17	0.0003
18	0.0001
19	0.0001
20	0.0000

TABLE 36. Numerical Results of the Simulation

Redund. L1_L2	Modules	Mean Capacity	Centered Capacity	LL Cent. Capacity	LR Cent. Capacity	UL Cent. Capacity	UR Cent. Capacity
0_0	2224	1.06	0.00	0.00	0.00	0.00	0.00
0_1	2189	3.09	0.00	0.00	0.00	0.00	0.00
0_2	2156	5.46	0.00	0.00	0.00	0.00	0.00
0_3	2124	7.85	0.00	0.00	0.00	0.00	0.00
0_4	2093	10.18	0.00	0.01	0.01	0.00	0.00
0_5	2062	12.43	0.00	0.03	0.03	0.00	0.00
0_6	2033	14.60	0.00	0.07	0.06	0.00	0.00
0_7	2004	16.72	0.00	0.15	0.12	0.00	0.00
0_8	1976	18.75	0.00	0.29	0.23	0.00	0.00
1_0	2202	19.70	0.00	3.03	4.37	0.00	0.00
1_1	2168	34.12	0.00	14.17	19.06	0.00	0.00
1_2	2135	44.69	0.01	34.91	44.53	0.00	0.00
1_3	2103	53.29	0.06	63.75	77.52	0.00	0.00
1_4	2072	60.70	0.32	96.66	112.96	0.00	0.00
1_5	2042	67.36	1.06	129.32	146.05	0.00	0.00
1_6	2013	73.30	2.65	158.43	173.76	0.00	0.00
1_7	1985	78.82	5.44	181.95	194.85	0.00	0.00
1_8	1957	83.78	9.96	199.48	209.56	0.00	0.00
2_0	2180	52.91	4.32	82.76	109.28	0.00	0.00
2_1	2147	82.59	21.14	169.27	198.63	0.01	0.07
2_2	2114	101.34	52.33	221.00	239.70	0.22	0.90
2_3	2083	115.70	93.11	243.56	252.60	1.39	3.67
2_4	2052	127.53	135.03	250.33	253.98	4.41	9.55
2_5	2022	137.69	171.13	250.37	251.62	10.20	19.56
2_6	1993	146.85	197.96	248.21	248.78	19.61	34.34
2_7	1965	154.79	215.45	244.89	244.99	33.09	53.56
2_8	1938	162.19	225.34	241.95	242.00	50.44	76.01
3_0	2159	93.78	77.10	182.13	202.94	3.21	11.13
3_1	2125	139.32	166.42	247.57	255.67	17.37	44.37
3_2	2093	165.45	220.94	258.36	260.18	45.69	94.04
3_3	2062	184.08	243.49	256.81	256.99	84.85	145.53
3_4	2032	198.27	249.38	253.00	253.00	126.99	186.85
3_5	2002	209.35	248.67	250.00	250.00	164.43	213.85
3_6	1973	217.02	245.85	246.00	246.00	192.88	228.43
3_7	1946	222.59	242.94	243.00	243.00	211.76	234.69
3_8	1919	225.37	239.00	239.00	239.00	222.46	235.94
4_0	2137	137.77	171.55	215.05	221.64	41.45	89.96
4_1	2104	194.91	242.29	257.04	258.54	113.02	182.14
4_2	2072	219.82	255.32	257.96	258.00	177.18	230.45
4_3	2041	232.27	254.10	254.99	255.00	217.00	246.59
4_4	2011	237.75	250.99	251.00	251.00	235.63	248.90
4_5	1982	239.30	247.00	247.00	247.00	241.68	246.81
4_6	1954	239.31	244.00	244.00	244.00	241.87	243.93
4_7	1926	237.36	240.00	240.00	240.00	239.63	240.00
4_8	1899	235.43	237.00	237.00	237.00	236.87	237.00

TABLE 37. Numerical Results of the Simulation

Redund. L1_L2	Modules	Mean Capacity	Centered Capacity	LL Cent. Capacity	LR Cent. Capacity	UL Cent. Capacity	UR Cent. Capacity
5_0	2115	174.11	209.17	220.67	222.22	113.73	172.51
5_1	2083	228.11	253.69	255.93	256.00	203.24	241.41
5_2	2051	241.94	255.30	255.99	256.00	239.44	253.18
5_3	2021	244.81	252.00	252.00	252.00	248.12	251.92
5_4	1991	244.09	248.00	248.00	248.00	247.48	248.00
5_5	1962	242.73	245.00	245.00	245.00	244.81	245.00
5_6	1934	239.79	241.00	241.00	241.00	240.99	241.00
5_7	1907	237.34	238.00	238.00	238.00	238.00	238.00
5_8	1880	233.69	234.00	234.00	234.00	234.00	234.00
6_0	2094	194.49	217.48	219.88	220.00	169.71	206.47
6_1	2062	240.39	253.37	253.99	254.00	237.94	251.02
6_2	2030	246.68	253.00	253.00	253.00	250.21	252.94
6_3	2000	245.84	249.00	249.00	249.00	248.89	249.00
6_4	1971	244.29	246.00	246.00	246.00	245.98	246.00
6_5	1942	241.14	242.00	242.00	242.00	242.00	242.00
6_6	1914	238.55	239.00	239.00	239.00	239.00	239.00
6_7	1887	234.81	235.00	235.00	235.00	235.00	235.00
6_8	1861	231.91	232.00	232.00	232.00	232.00	232.00
7_0	2073	203.52	217.63	217.99	218.21	197.45	214.81
7_1	2041	243.51	251.00	251.00	251.00	246.45	250.80
7_2	2010	247.22	251.00	251.00	251.00	249.84	250.69
7_3	1980	245.28	247.00	247.00	247.00	246.99	247.00
7_4	1951	242.18	243.00	243.00	243.00	243.00	243.00
7_5	1922	239.60	240.00	240.00	240.00	240.00	240.00
7_6	1895	235.84	236.00	236.00	236.00	236.00	236.00
7_7	1868	232.93	233.00	233.00	233.00	233.00	233.00
7_8	1842	229.97	230.00	230.00	230.00	230.00	230.00
8_0	2051	206.71	215.56	215.79	215.73	207.91	215.33
8_1	2020	243.88	248.66	248.72	248.70	247.13	248.63
8_2	1989	245.83	248.00	248.00	248.00	247.95	248.00
8_3	1959	243.03	244.00	244.00	244.00	244.00	244.00
8_4	1930	240.55	241.00	241.00	241.00	241.00	241.00
8_5	1902	236.82	237.00	237.00	237.00	237.00	237.00
8_6	1875	233.92	234.00	234.00	234.00	234.00	234.00
8_7	1849	230.97	231.00	231.00	231.00	231.00	231.00
8_8	1823	226.99	227.00	227.00	227.00	227.00	227.00

6.5 Optimizing Redundancy

Determining the optimal redundancy depends on where the manufacturing process is expected to be centered. Table 38 shows the top ten yields for each distribution.

TABLE 38. Top Ten Yields for Distributions

Mean		Centered		LL Centered	
Capacity	R1_R2	Capacity	R1_R2	Capacity	R1_R2
247.22	7_2	255.32	4_2	258.36	3_2
246.68	6_2	255.30	5_2	257.96	4_2
245.84	6_3	254.10	4_3	257.04	4_1
245.83	8_2	253.69	5_1	256.81	3_3
245.28	7_3	253.37	6_1	255.99	5_2
244.81	5_3	253.00	6_2	255.93	5_1
244.29	6_4	252.00	5_3	254.99	4_3
244.09	5_4	251.00	7_1	253.99	6_1
243.88	8_1	251.00	7_2	253.00	6_2
243.51	7_1	251.00	4_4	253.00	3_4
LR Centered		**UL Centered**		**UR Centered**	
Capacity	R1_R2	Capacity	R1_R2	Capacity	R1_R2
260.18	3_2	250.21	6_2	253.18	5_2
258.54	4_1	249.84	7_2	252.94	6_2
258.00	4_2	248.89	6_3	251.92	5_3
256.99	3_3	248.12	5_3	251.02	6_1
256.00	5_2	247.95	8_2	250.80	7_1
256.00	5_1	247.48	5_4	250.69	7_2
255.67	3_1	247.13	8_1	249.00	6_3
255.00	4_3	246.99	7_3	248.90	4_4
254.00	6_1	246.45	7_1	248.63	8_1
253.98	2_4	245.98	6_4	248.00	8_2
				248.00	5_4

The distributions in the lower quadrants, LL_centered and LR_centered, demonstrate the highest capacities. This is because the lower defect densities in the lower quadrants require less redundancy to compensate for defects. This means more 1 Mbit modules can fit on the wafer and the total capacity can be higher. The mean yield distribution demonstrates the lowest capacities because it is the only distribution which is impacted by the high defect densities which lie along the upper edges of the upper quadrants. The mean yield distribution and the upper quadrant distributions, UL_centered and UR_centered, require higher levels of redundancy to compensate for the higher

defect density. The overall wafer yield for these distributions is slightly lower due to fewer 1 Mbit modules being available because of the higher redundancy.

Figure 12 shows the statistical distribution capacities and the mean capacities for the redundancy combinations with the top ten mean capacities. The best general redundancy combination is the 6_2 combination. This combination has six redundant sense lines in the 16 Kbit module and two redundant 16 Kbit modules per 1 Mbit module. The 6_2 combination has the second highest mean capacity and is the only combination which yields capacities of 250 MBytes or greater for all the statistical distributions.

An important consideration in determining which combination is optimal is where future manufacturing process parameters will be located. Stapper states that λ and α decrease exponentially with time[20]. This means the center of the manufacturing process will move down and to the left over time. Combination 6_2 has the highest LL_centered capacity of any of the top ten mean distribution combinations.

6.6 Characteristics of the Optimized Design

The capacity distribution of the 6_2 redundancy product is shown in Figure 13 for the centered distribution. This capacity distribution is calculated using the probability function for the binomial distribution.

$$P(X=N) = \binom{M}{N} Y^N (1-Y)^{M-N}$$

Figure 13 displays the capacity by megabytes. Figure 14 shows the capacity distribution as the number of functional modules. Figures 15 and 16 show capacity versus yield for the centered 6_2 combination. The yield stays at one for all capacities below the probable capacities.

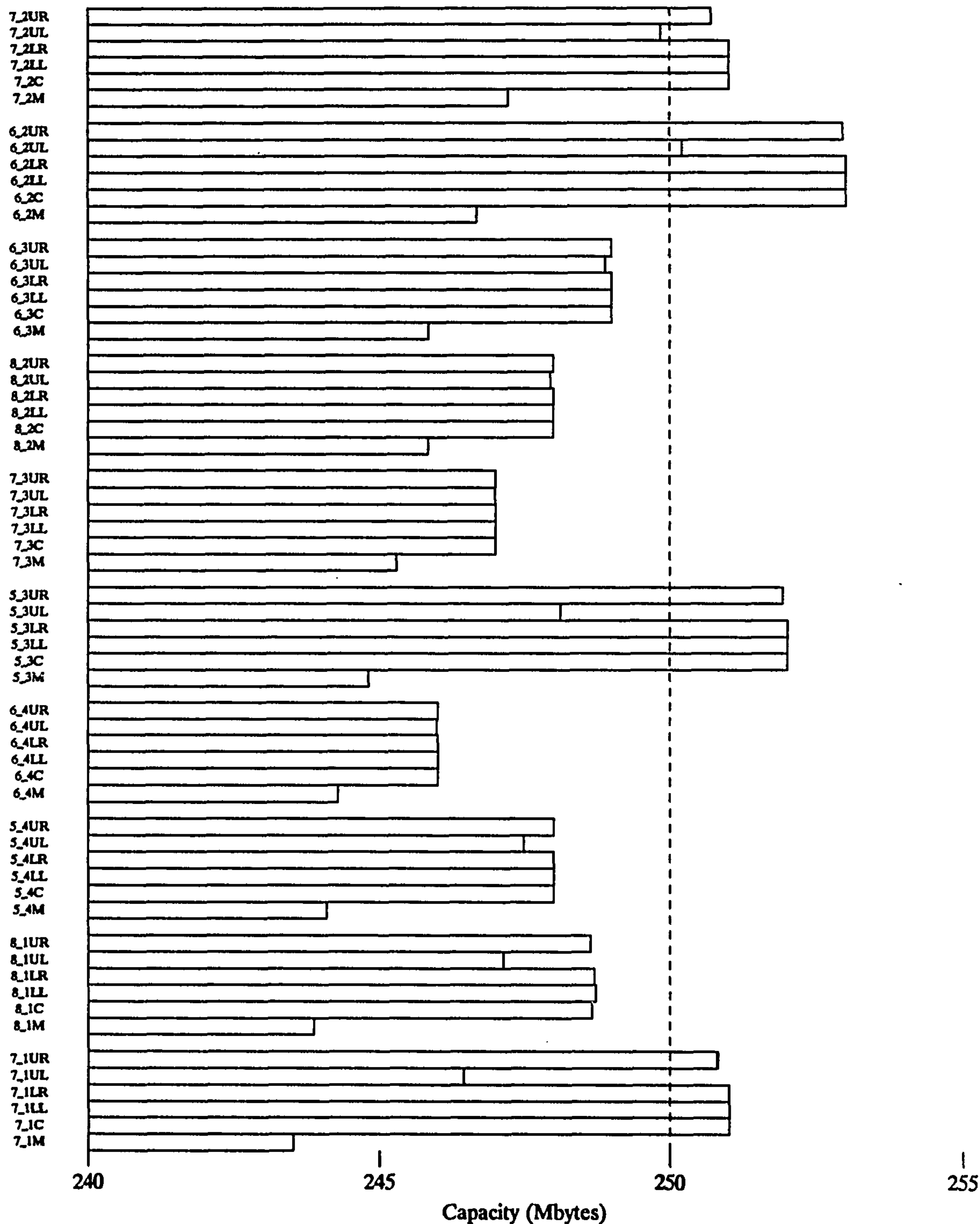

Figure 12. Distribution of Top Ten Mean Distribution Combinations

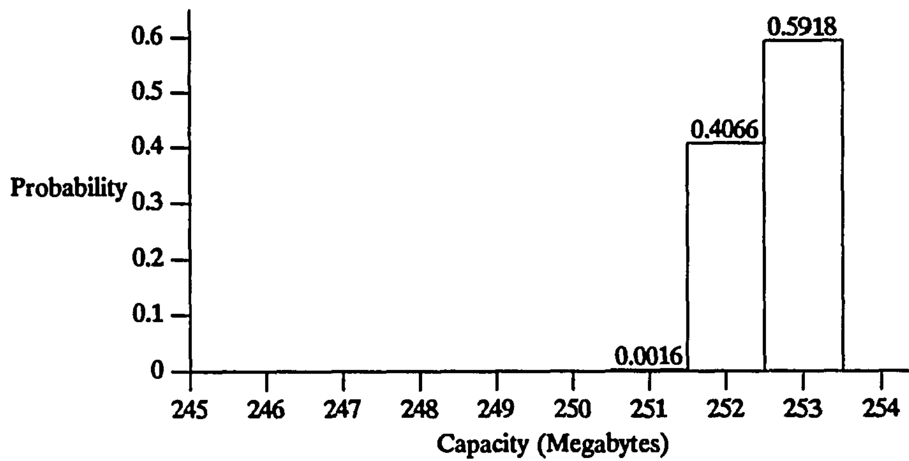

Figure 13. Capacity Probability

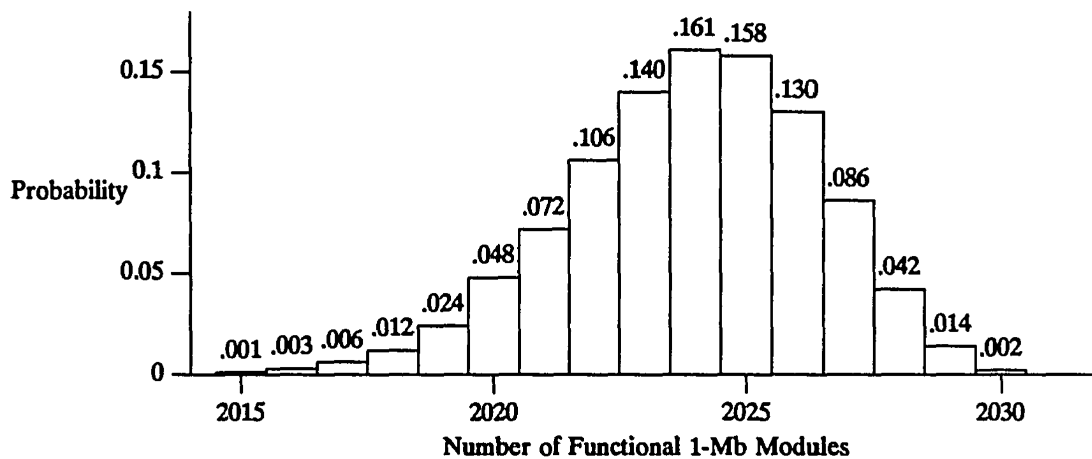

Figure 14. Number of Functional Modules Probability

Table 39 estimates the factory cost and end-user costs for the wafer scale product. Note that the factory cost is adjusted to compensate for gross yield losses which are about ten percent. Gross yield losses are due to processing errors and typically cause a large portion of the wafer to be unusable. The attributes of the wafer scale product are summarized in Table 40. The attributes of the wafer scale product are shown in Figure 17 along with the projected DRAM and winchester hard

disk attributes for 1993.

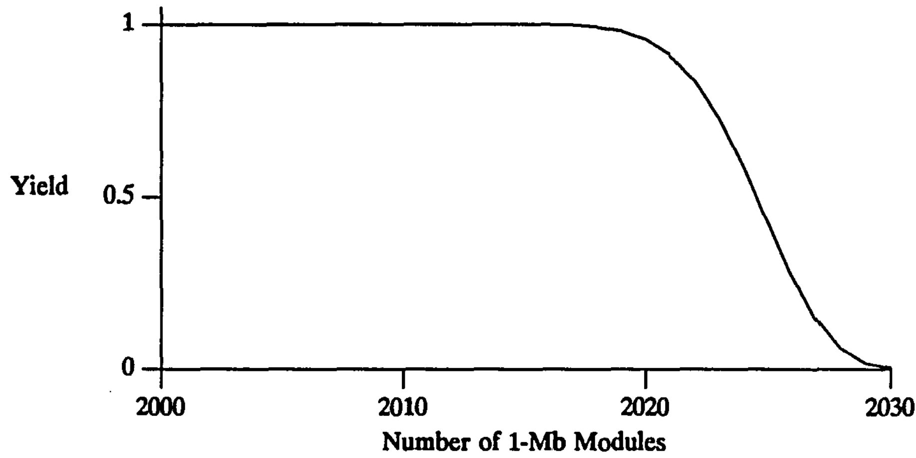

Figure 15. Yield of Capacities (Number of Functional Modules)

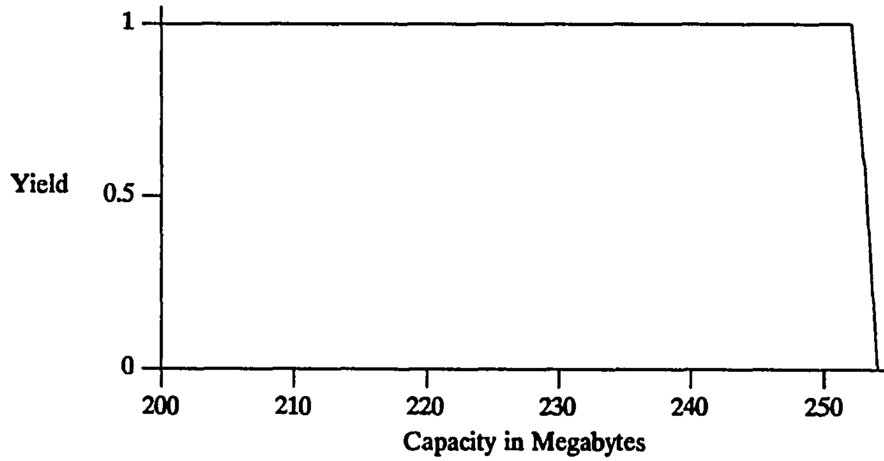

Figure 16. Yield of Capacities over 200 MB

TABLE 39. Estimated Product Costs

Item	Cost (dollars)
Processed Wafer	550
Support Electronics	50
Wafer Packaging	100
Preliminary Factory Cost	700
Cost of Gross Yield Loss	75
Corrected Factory Cost	775
Cost Multiplier	x4
End-user Cost	3100

TABLE 40. Wafer Scale Product Attributes

End-user Cost ($)	3100
Capacity (MB)	250
Approximate Cost ($/MB)	12
Access Time (usec)	1
Transfer Rate (Mb/sec)	64

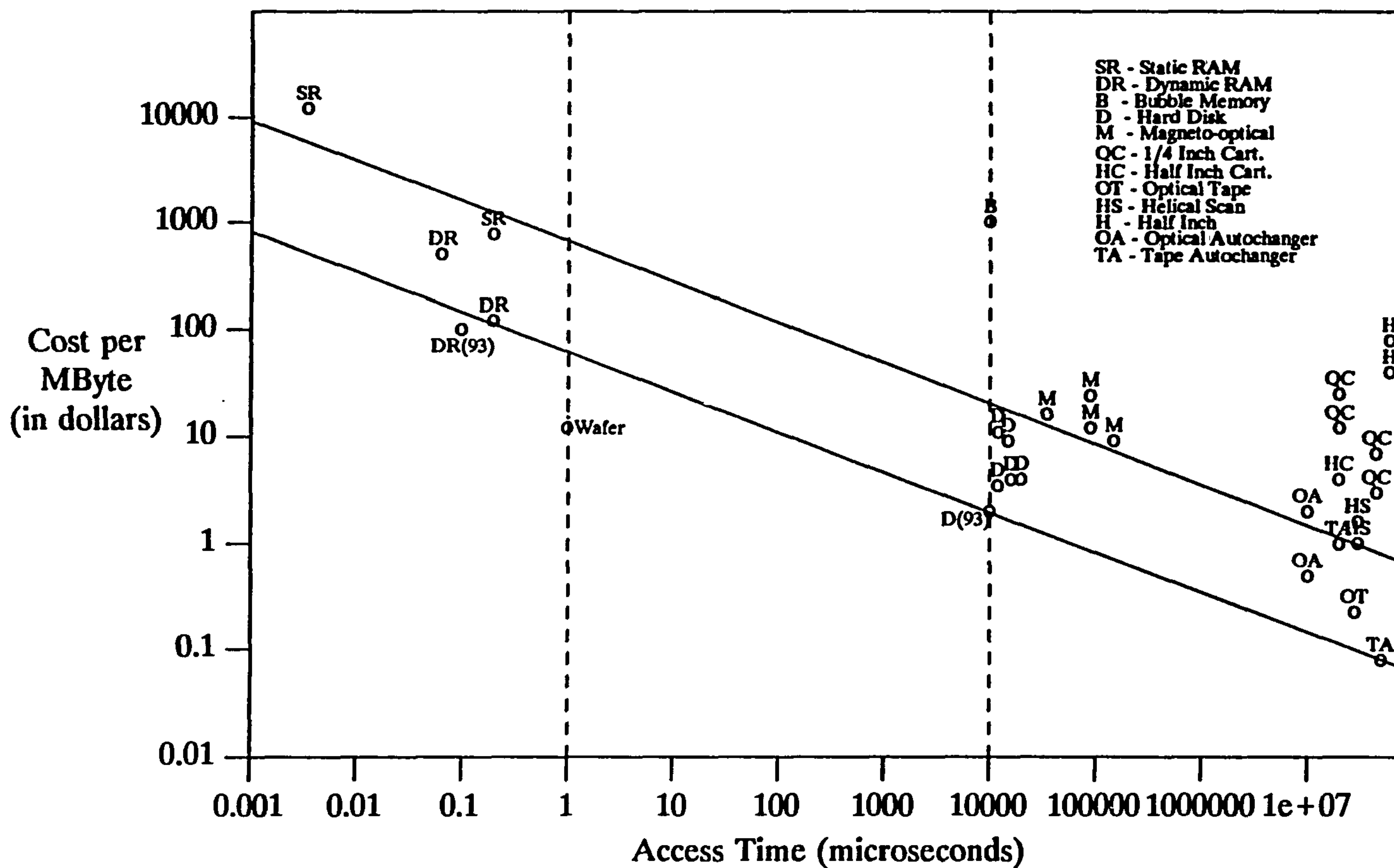

Figure 17. Cost and Access Times of Memory Technologies with 1993 Additions

7. CONCLUSIONS

The yield simulation has demonstrated that magnetoresistive memory can be effectively integrated into a wafer scale product. Efficient use of redundancy results in a high capacity product with less than 10 percent of the area dedicated to redundancy. The yield simulation indicates there are a number of possible redundancy combinations. The 6_2 combination, with six level one spares and two level two spares, appears to be the optimal combination for current manufacturing processes.

Wafer scale magnetoresistive memory presents a valuable gap technology product which could be utilized to significantly enhance hierarchical memory systems.

7.1 Future Research

The implementation of magnetoresistive memory using wafer scale integration can be further investigated in the following areas. These areas involve refining the simulation model to incorporate more detailed parameters.

The first area of refinement is to allow λ and α to vary over the surface of the wafer. The current simulations hold λ and α constant for each yield calculation. Analysis of defects in semiconductor manufacturing have shown that defects have a higher density at the edge of the wafer and a lower density in the middle of the wafer.

The second area of refinement involves the types of defects. The current model utilizes two types of defects which are interpreted as defect densities or faults per unit area. A more refined model could utilize ten or more types of defects. For each defect type, the critical area of the circuit which can be damaged by the defect type is multiplied by the defect density to determine the yield impact of the current defect type. This approach requires detailed circuit layout information and information about the defect densities of the manufacturing process.

8. REFERENCES

1. Mayer, John H. "5-1/4 in. Winchesters move toward gigabyte capacities." *Computer Design*, 27 (Sept. 1, 1988):70-6.

2. Voelcker, John. "Winchester disks reach for a gigabyte." *IEEE Spectrum*, 24 (February 1987):64-7.

3. Bajorek, Christopher. "Magnetic and Optical Devices to Quench Data-Storage Quest." *Electronic Design*, 36 (Jan. 7, 1988):86.

4. Santo, Brian. "Technology '89 - Solid State." *IEEE Spectrum*, 26 (January 1989):47-49.

5. Pohm, A.V. and O.P. Agarwal. *High Speed Memory Systems*. Reston, Va.: Reston Publishing Co., Inc., 1983.

6. Granley, Gerald and James Daughton. Magnetic Random Access Storage Technology. Private communication from Honeywell, Inc. to A. V. Pohm.

7. Smith, Alan J. "Disk Cache - Miss Ratio Analysis and Design Considerations." *ACM Transactions on Computer Systems*, 3 (August 1985):161-203.

8. Cullity, B. D. *Introduction to Magnetic Materials*. Cambridge, Mass.:Addison-Wesley Publishing Co., Inc. 1972.

9. Pohm A. V., J.S.T. Huang, J.M. Daughton, D.R. Krahn and V. Mehra. "The Design of a One Megabit Non-volatile M-R Memory Chip using 1.5 x 5 um Cells." *IEEE Transactions on Magnetics*, 24 (November 1988):311[illegible]

10. Carlson, Richard and Constantine Neugebauer. "Future Trends in Wafer Scale Integration." *Proceedings of the IEEE*, 74 (December 1986):1741-52.

11. Pohm A. V., C.S. Comstock and J.M. Daughton. "Analysis of M-R Elements For 10^8 Bit/cm2 Arrays." *IEEE Transactions on Magnetics*, 25 (September 1989):4266-8.

12. Harden, Jim C. and Noel R. Strader. "Architectural Yield Optimization for WSI." *IEEE Transactions on Computers*, 37 (January 1988):88-110.

13. Curran, Lawrence. "Wafer-Scale Integration Arrives in 'Disk' Form." *Electronic Design*, 37 (October 26, 1989):51-4.

14. Warren, Ben, Wayne Richardson, Keigi Kanegawa and Cliff Arnell. "A One Megabit SRAM Fabricated with 1.2 u Technology." *Proceedings International Conference on Wafer Scale Integration*, San Francisco ,January 1989.

15. Bourassa, Ron, Tim Coffman, and Joe Brewer. "Progress in WSI SRAM Development." *Proceedings International Conference on Wafer Scale Integration*, San Francisco ,January 1990.

16. Inoue, J., T. Matsumura, M. Tanno and J. Yamada. Parallel Testing Technology For VLSI Memories, *1987 International Test Conference*, Washington, D.C., September 1987.

17. Tuszynski A., "Memory Chip Test Economics." *1986 International Test Conference*, Washington, D.C., September 1986.

18. Baba, F and A. Sinclair. "200 Mb Wafer Scale Memory." *1990 Proceedings of the International Conference on Wafer Scale Integration*, San Francisco, January 1990.

19. Stapper C.H., A.N. McLaren and M. Dreckmann. "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product." *IBM Journal of Research and Development*, 24 (May 1980):398-409.

20. Stapper, Charles H. "Modeling Redundancy in 64K to 16Mb DRAMs." *1989 IEEE International Solid State Circuits Conference*, (1989):86-87.

21. Moore Will R. "A Review of Fault-Tolerant Techniques for the Enhancement of Integrated Circuit Yield." *Proceedings of the IEEE*, 74 (May 1986):684-98.

22. Stapper, C., F. Armstrong and K. Saji. "Integrated Circuit Yield Statistics." *Proceedings of the IEEE*, 71 (April 1983):453-70.

23. Knuth, Donald E. *The Art of Computer Programming*. Reading, MA.:Addison-Wesley Publishing Co., 1981.

9. ACKNOWLEDGEMENTS

I wish to thank Dr. A. V. Pohm and Dr. C. S. Comstock for their support and guidance on this project. I also want to thank my wife, Rebecca, and my children, Jonathan and KayLynn for their support and encouragement.

10. APPENDIX A: APPROXIMATION VERIFICATION CODE

```
#include <stdio.h>
#include <math.h>

double comb(int x, int n);
double cal_16K_yield(int spares, int rows, double MR_defect_rate,
                     double El_defect_density, double alpha);
void   output_mean(double N, double yield);

main()
{
   int spares,n,R1,R2,rows;
   double MR_def_rate, EL_def_density, alpha;
   double yield_1K, Modules, yield_64, error, max_error;

printf(" R1 Max_error0);

for (R1=1; R1<=4; R1+=1)    /* upper limit */
  {
   EL_def_density = 0.1;    /* defects per sqare mm */

   max_error = 0.00000;
   MR_def_rate = 0.00002;
   while (MR_def_rate >= 0.000001)
   {
      alpha = 0.1;
      while (alpha <= 10.0)
      {
          rows = 64;
          yield_64 = cal_16K_yield(R1, rows, MR_def_rate, EL_def_density, alpha);
          rows = 1024;
          yield_1K = cal_16K_yield(R1, rows, MR_def_rate, EL_def_density, alpha);
          error = ( yield_64 - yield_1K ) / yield_1K;
/*        printf("%1.4f %1.4f %1.6f", yield_1K, yield_64, error); */
           if (fabs(error) > fabs(max_error))
             max_error = error;
           alpha = alpha + 0.10;
      }
      MR_def_rate = MR_def_rate - 0.000001;
   }
   printf("%3d %4.4f0, R1, max_error);
  }
}

double cal_16K_yield(int spares, int rows, double MR_defect_rate,
                     double El_defect_density, double alpha)
{
       int    N, M, i;
       double factor;
       double yield_16K, yield_CK, yield_MR, yield_term, sign;
       double        lambda_MR, lambda_CK, CK_area, sense_line_kill_area;

       CK_area = 0.0265;                    /*  mm x mm */
       sense_line_kill_area = 0.0265;              /*  mm x mm */

       lambda_CK = CK_area * El_defect_density;
       lambda_MR = MR_defect_rate * 16384 + sense_line_kill_area * El_defect_density;

       M = rows;
       N = M + spares;
       yield_MR = 0 ;
       i = 0;
```

```
	while ( i <= spares )				/* sum over i */
	{
		sign = pow( -1.0 , spares + i );
		factor = comb(N,N-i);
		factor = factor * comb(N-1-i,M-1);

		yield_term = pow( (1+(N-i) * lambda_MR / ( N * alpha )), -1.0 * alpha );
		yield_MR = yield_MR + sign * factor * yield_term ;
		i = i + 1;
	}

	yield_CK = pow( 1 + lambda_CK / alpha, -1.0 * alpha );

	yield_16K = yield_CK * yield_MR;

	return yield_16K;

}

/*  ( x )       x*(x-1)*(x-2)*...*(x-n+1)   */
/*  (   )   =   -------------------------   */
/*  ( n )           n*(n-1)*...*2*1         */

double comb(int x, int n)
{
	int	limit;
	double product;

	product = 1.0;
	limit = x - n + 1;

	while ( x >= limit )
	{
		if ( x > n )
			product = product * x;
		else
			n = n - 1;
		x = x - 1;
	}

	while ( n > 1 )
	{
		product = product / n;
		n = n - 1;
	}

	return product;
}
```

Results

R1	Max_error
1	0.0007
2	0.0008
3	0.0008
4	0.0008

11. APPENDIX B: SIMULATION SOURCE CODE

```
#include <stdio.h>
#include <math.h>

const int stat_width = 19;
const double stat_vec[39] = {0.0001,0.0001,0.0003,0.0005,0.0009,0.0016,0.0027,0.0045,
                             0.0072,0.0108,0.0159,0.0222,0.0300,0.0389,0.0484,0.0579,
                             0.0665,0.0736,0.0781,0.0796,0.0781,0.0736,0.0665,0.0579,
                             0.0484,0.0389,0.0300,0.0222,0.0159,0.0108,0.0072,0.0045,
                             0.0027,0.0016,0.0009,0.0005,0.0003,0.0001,0.0001};

   struct stat_data {
                     double net_capacity;
                     int    vert_center;
                     int    horz_center;
                  };

   struct stat_log {
                     struct stat_data mean;
                     struct stat_data centered;
                     struct stat_data LL_centered;
                     struct stat_data LR_centered;
                     struct stat_data UL_centered;
                     struct stat_data UR_centered;
                  };

double comb(int x, int n);
double cal_16K_yield(int spares, double MR_defect_rate, double El_defect_density, double alpha);
double cal_1M_yield(int spares, double yield_16K);
double cal_1M_modules(int R1, int R2);
void   output_row(double *yield);
struct stat_log initialize_statistics(void);
struct stat_log update_statistics(int row, double *yield, struct stat_log stats);

main()
{

 struct stat_log statistics;
 int spares,n,R1,R2,row,column;
 double MR_def_rate, EL_def_density, alpha;
 double yield_16K, Modules, yield_1M;
 double yield_row[100];

 alpha = 0.0;
 for (n=0; n<=38; n+=1)
    alpha = alpha + stat_vec[n];

 printf("stat sum = %1.5f0, alpha);
 printf("R1  R2  Modules  Mean    Centered   LL_Centered  LR_Centered  UL_Centered  UR_Centered0)


 for (R1=0; R1<=8; R1+=1)          /* upper limit */
 {
  for (R2=8; R2<=8; R2+=1)         /* upper limit */
  {
   statistics = initialize_statistics();
   Modules = cal_1M_modules( R1, R2);
   EL_def_density = 0.1;    /* defects per sqare mm */

   MR_def_rate = 0.0001;
   row = 99;
   while (MR_def_rate >= 0.000001)
```

```
   {
      alpha = 0.1;
      column = 0;
      while (alpha <= 10.0)
      {
          yield_16K = cal_16K_yield(R1, MR_def_rate, EL_def_density, alpha);
          yield_1M = cal_1M_yield(R2, yield_16K);
/*        printf("%1.4f %1.4f %1.4f0, yield_16K, yield_1M, Modules); */
           yield_row[column] = floor(Modules * yield_1M/8.0);
           alpha = alpha + 0.10;
           column += 1;
      }
/*       output_row(&yield_row[0]); */
      statistics = update_statistics(row, &yield_row[0], statistics);
      MR_def_rate = MR_def_rate - 0.000001;
      row -= 1;
   }
      printf("%2d %2d %4.4f ", R1, R2, Modules);
      printf("%3.4f  %3.4f  ", statistics.mean.net_capacity, statistics.centered.net_capacity);
      printf("%3.4f  %3.4f  %3.4f  %3.4f0, statistics.LL_centered.net_capacity,
            statistics.LR_centered.net_capacity, statistics.UL_centered.net_capacity,
            statistics.UR_centered.net_capacity);
  }
 }
}

double cal_16K_yield(int spares, double MR_defect_rate, double El_defect_density, double alpha)
{
       int     N, M, i;
       double factor;
       double yield_16K, yield_CK, yield_MR, yield_term, sign;
       double        lambda_MR, lambda_CK, CK_area, sense_line_kill_area;

       CK_area = 0.0265;                   /*  mm x mm */
       sense_line_kill_area = 0.0265;             /*  mm x mm */

       lambda_CK = CK_area * El_defect_density;
       lambda_MR = MR_defect_rate * 16384 + sense_line_kill_area * El_defect_density;

       M = 64;
       N = M + spares;
       yield_MR = 0 ;
       i = 0;

       while ( i <= spares )                       /* sum over i */
       {
              sign = pow( -1.0 , spares + i );
              factor = comb(N,N-i);
              factor = factor * comb(N-1-i,M-1);

              yield_term = pow( (1+(N-i) * lambda_MR / ( N * alpha )), -1.0 * alpha );
              yield_MR = yield_MR + sign * factor * yield_term ;
              i = i + 1;
       }

       yield_CK = pow( 1 + lambda_CK / alpha, -1.0 * alpha );

       yield_16K = yield_CK * yield_MR;

       return yield_16K;
```

```
}

double cal_1M_yield(int spares, double yield_16K)
{
   int M,R,n;
   double yield_1M;

   M = 64;
   R = spares;
   yield_1M = 0.0;
   n = 0;
   while ( n <= R )
      {
      yield_1M = yield_1M + comb(M+R,n) * pow(yield_16K,M+R-n) * pow(1-yield_16K,n);
      n += 1;
      }
   if (yield_1M > 0.99990)
      yield_1M = 0.9999;
   return yield_1M;
}

/*  ( x )       x*(x-1)*(x-2)*...*(x-n+1)   */
/*  (   )   =   -------------------------   */
/*  ( n )           n*(n-1)*...*2*1         */

double comb(int x, int n)
{
       int    limit;
       double product;

       product = 1.0;
       limit = x - n + 1;

       while ( x >= limit )
       {
              if ( x > n )
                     product = product * x;
              else
                     n = n - 1;
              x = x - 1;
       }

       while ( n > 1 )
       {
              product = product / n;
              n = n - 1;
       }

       return product;
}


double cal_1M_modules(int R1, int R2)
{
       double AC_16K, AC_1M, Usable_area, Modules;

       AC_1M = 1.0 * R2/(64.0 + R2);

       AC_16K = 10.0 *  R1/(1024.0 + R1);

       Usable_area = ( 1.0 - AC_1M ) * ( 1.0 - AC_16K );
```

```
        Modules = floor(2224 * Usable_area);

        return Modules;
}

void    output_row(double *yield)
{
        int mean,i;
        char ch;

        for (i=0; i<=99; i+=1)
         {
        mean = floor(*(yield+i)/10.0);
        ch = 'a' + mean;
        printf("%c", ch);
         };
        printf("0);

}


struct stat_log update_statistics(int row, double *yield, struct stat_log stats)
{
        int i,center_row,center_col,start_col;
        double horz_mult, vert_mult, capacity;


        for (i=0; i<=99; i+=1)
        stats.mean.net_capacity += 0.0001 * *(yield+i);
        center_row = stats.centered.vert_center;
        if ((row >= (center_row - stat_width)) && (row <= (center_row + stat_width)))
        {
           vert_mult = stat_vec[stat_width + row - center_row];
           center_col = stats.centered.horz_center;
           start_col = center_col - stat_width;

           for (i=0; i<=38; i+=1)
           {
              horz_mult = stat_vec[i];
              capacity = *(yield + start_col + i);
              stats.centered.net_capacity += vert_mult * horz_mult * capacity;
           };
        };

        center_row = stats.LL_centered.vert_center;
        if ((row >= (center_row - stat_width)) && (row <= (center_row + stat_width)))
        {
           vert_mult = stat_vec[stat_width + row - center_row];
           center_col = stats.LL_centered.horz_center;
           start_col = center_col - stat_width;

           for (i=0; i<=38; i+=1)
           {
              horz_mult = stat_vec[i];
              capacity = *(yield + start_col + i);
              stats.LL_centered.net_capacity += vert_mult * horz_mult * capacity;
           };
        };

        center_row = stats.LR_centered.vert_center;
        if ((row >= (center_row - stat_width)) && (row <= (center_row + stat_width)))
```

```
        {
           vert_mult = stat_vec[stat_width + row - center_row];
           center_col = stats.LR_centered.horz_center;
           start_col = center_col - stat_width;

           for (i=0; i<=38; i+=1)
           {
              horz_mult = stat_vec[i];
              capacity = *(yield + start_col + i);
              stats.LR_centered.net_capacity += vert_mult * horz_mult * capacity;
           };
        };

        center_row = stats.UL_centered.vert_center;
        if ((row >= (center_row - stat_width)) && (row <= (center_row + stat_width)))
        {
           vert_mult = stat_vec[stat_width + row - center_row];
           center_col = stats.UL_centered.horz_center;
           start_col = center_col - stat_width;

           for (i=0; i<=38; i+=1)
           {
              horz_mult = stat_vec[i];
              capacity = *(yield + start_col + i);
              stats.UL_centered.net_capacity += vert_mult * horz_mult * capacity;
           };
        };

        center_row = stats.UR_centered.vert_center;
        if ((row >= (center_row - stat_width)) && (row <= (center_row + stat_width)))
        {
           vert_mult = stat_vec[stat_width + row - center_row];
           center_col = stats.UR_centered.horz_center;
           start_col = center_col - stat_width;

           for (i=0; i<=38; i+=1)
           {
              horz_mult = stat_vec[i];
              capacity = *(yield + start_col + i);
              stats.UR_centered.net_capacity += vert_mult * horz_mult * capacity;
           };
        };
        return stats;
}

struct stat_log initialize_statistics(void)
{
        struct stat_log statistics;

        statistics.mean.net_capacity = 0.0;
        statistics.centered.net_capacity = 0.0;
        statistics.LL_centered.net_capacity = 0.0;
        statistics.LR_centered.net_capacity = 0.0;
        statistics.UL_centered.net_capacity = 0.0;
        statistics.UR_centered.net_capacity = 0.0;

        statistics.mean.vert_center  = 0;
        statistics.centered.vert_center  = 50;
        statistics.LL_centered.vert_center  = 25;
        statistics.LR_centered.vert_center  = 25;
        statistics.UL_centered.vert_center  = 75;
```

```
        statistics.UR_centered.vert_center  = 75;

        statistics.mean.horz_center  = 0;
        statistics.centered.horz_center  = 50;
        statistics.LL_centered.horz_center  = 25;
        statistics.LR_centered.horz_center  = 75;
        statistics.UL_centered.horz_center  = 25;
        statistics.UR_centered.horz_center  = 75;

        return statistics;
}
```

12. APPENDIX C: GRAPHICAL SIMULATION RESULTS

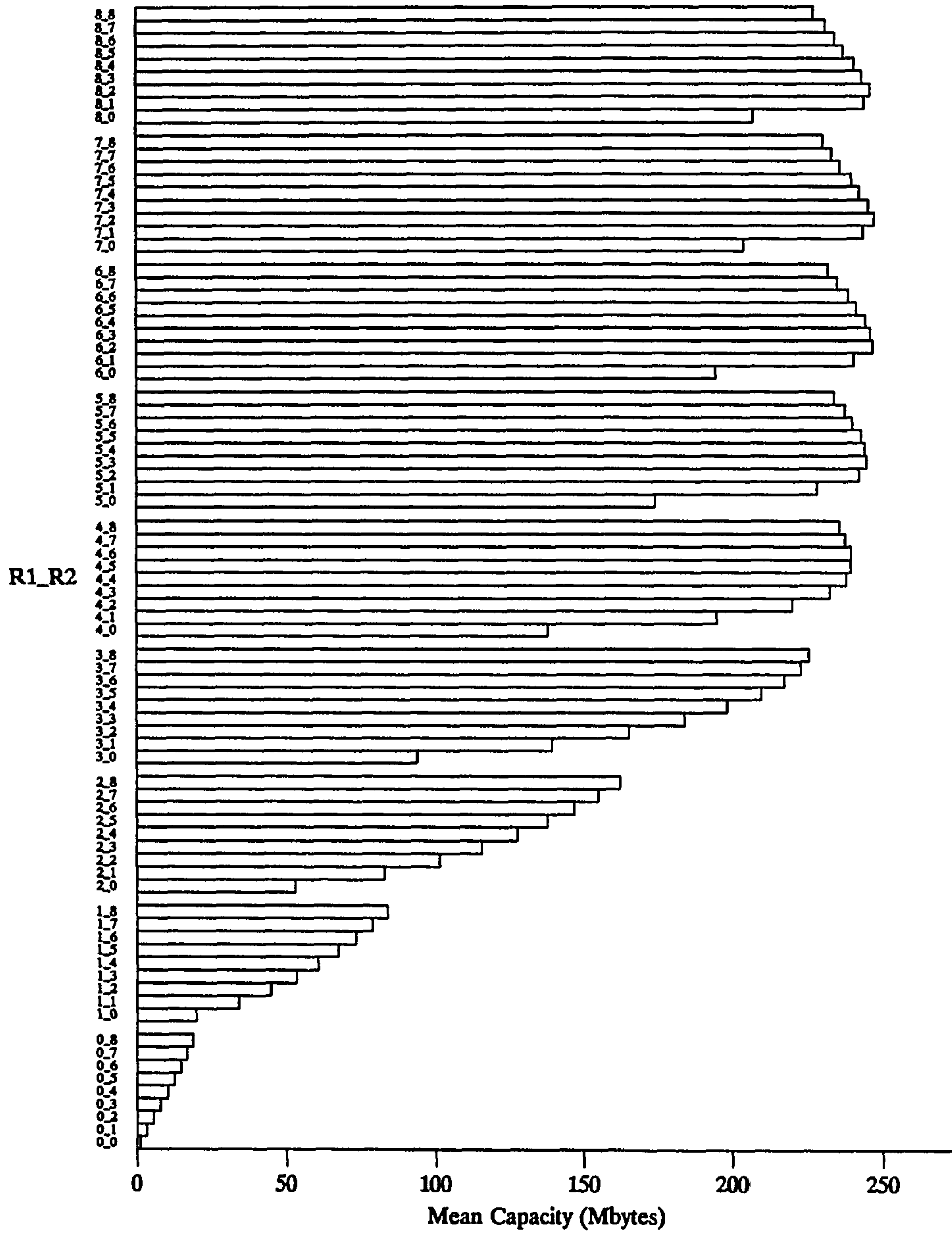

Figure 18. Mean Yield Capacities (grouped by level 1 redundancy)

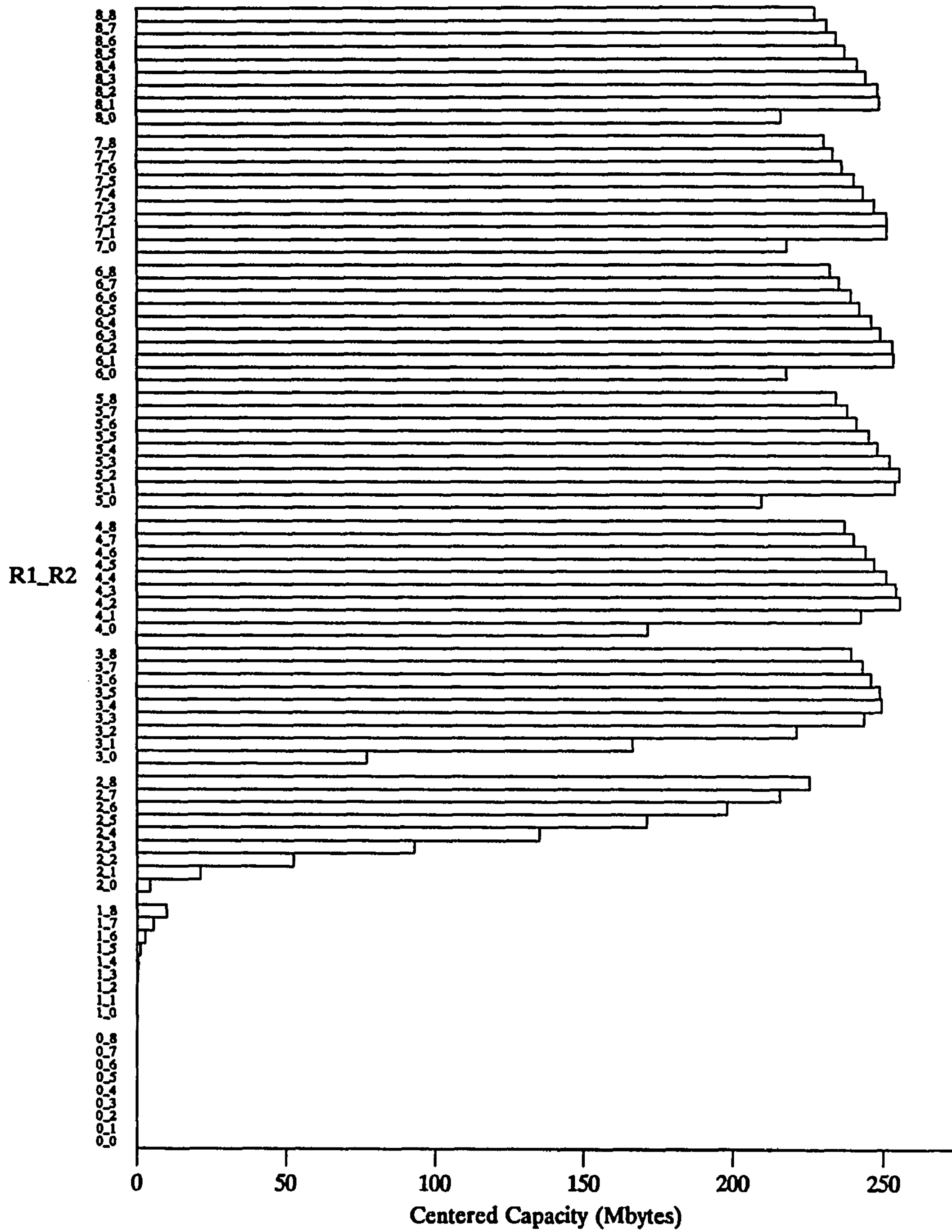

Figure 19. Centered Yield Capacities (grouped by level 1 redundancy)

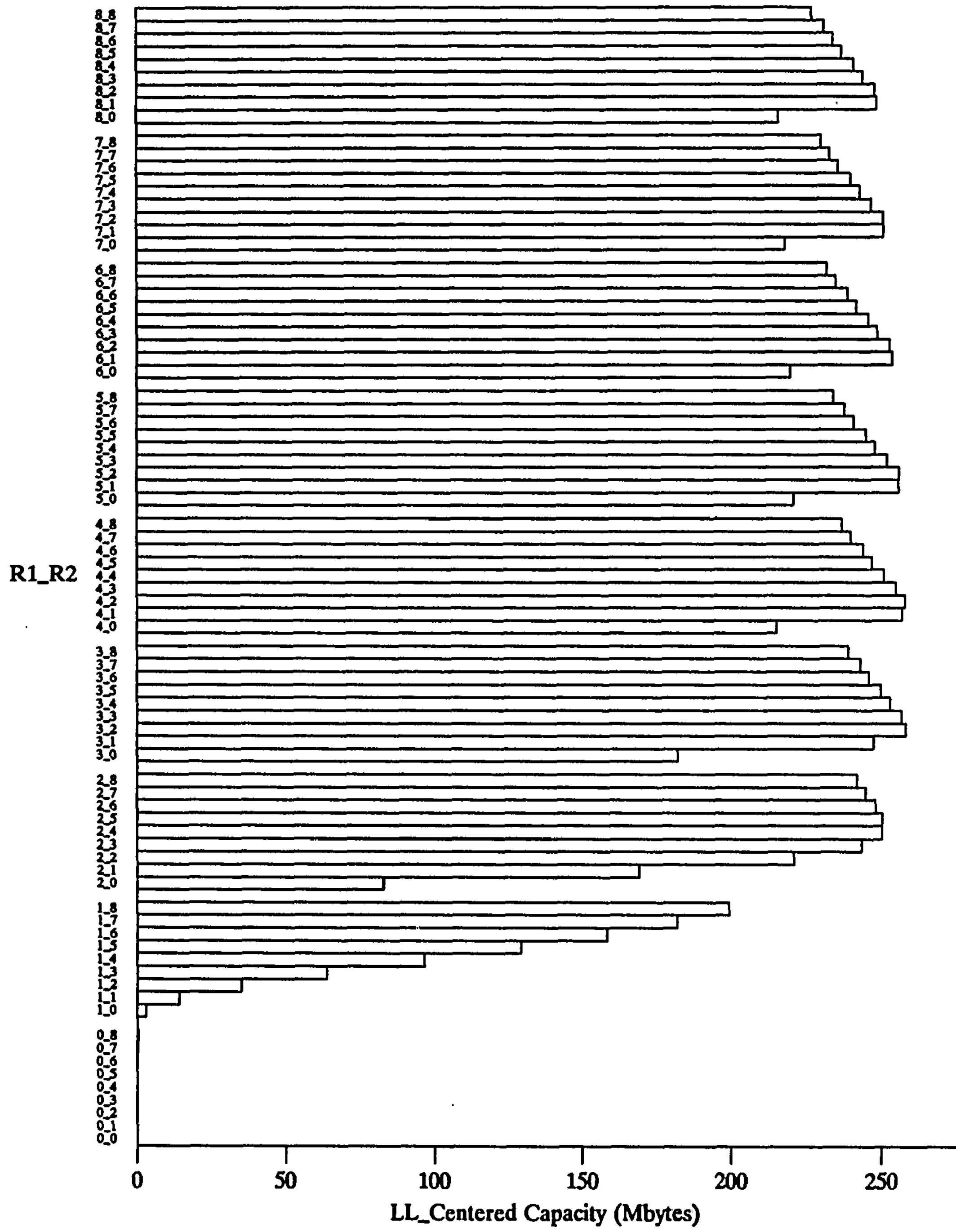

Figure 20. LL_Centered Yield Capacities (grouped by level 1 redundancy)

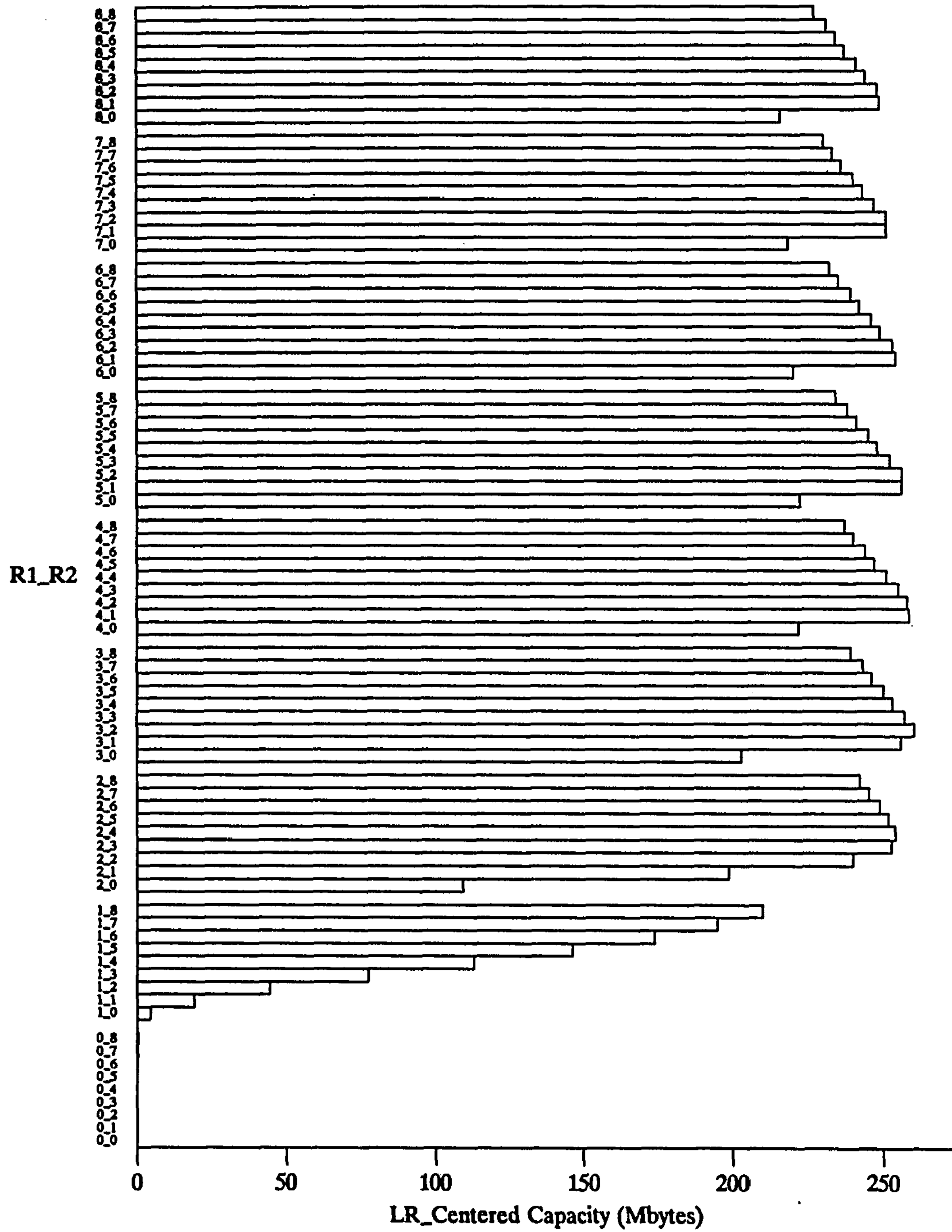

Figure 21. LR_Centered Yield Capacities (grouped by level 1 redundancy)

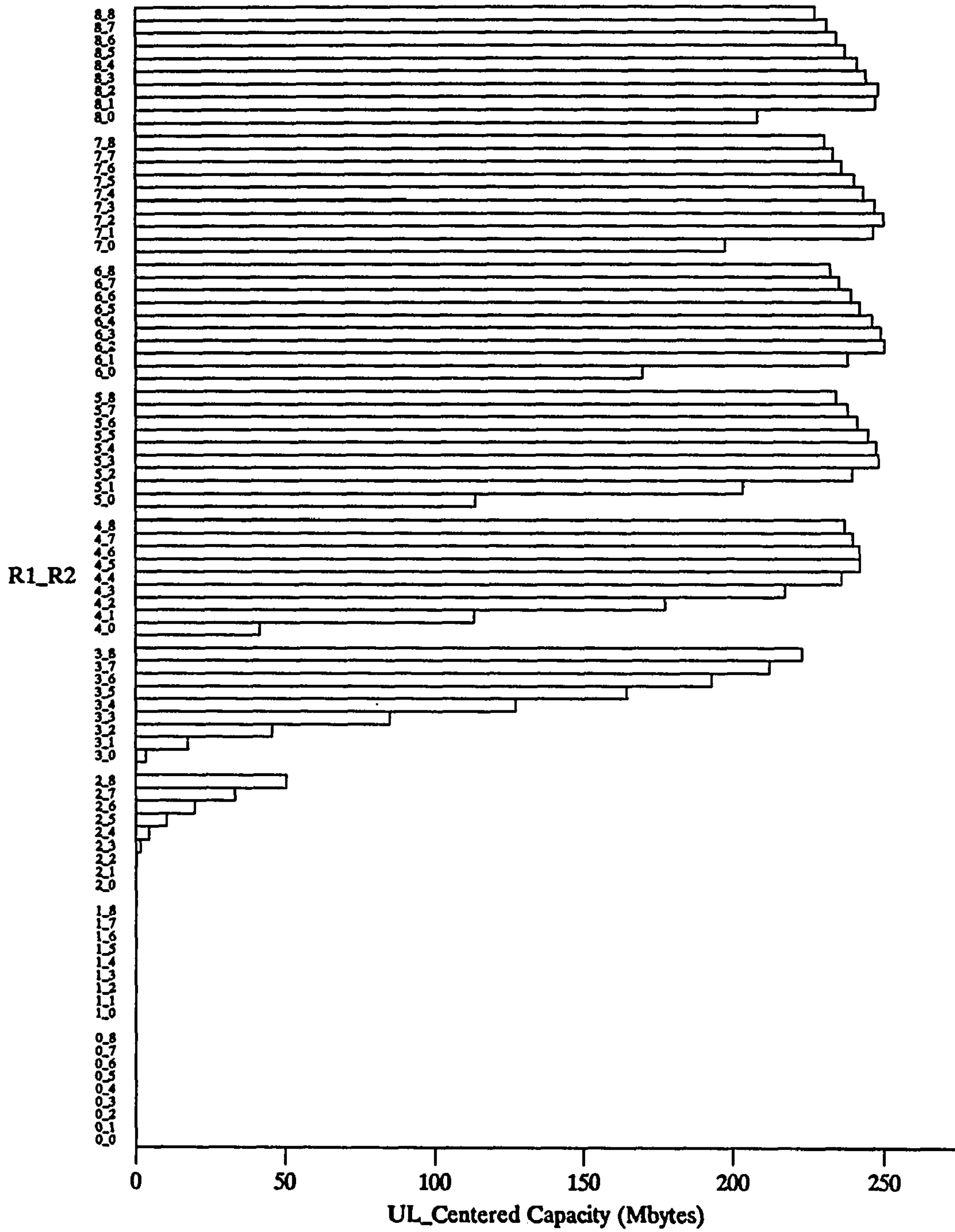

Figure 22. UL_Centered Yield Capacities (grouped by level 1 redundancy)

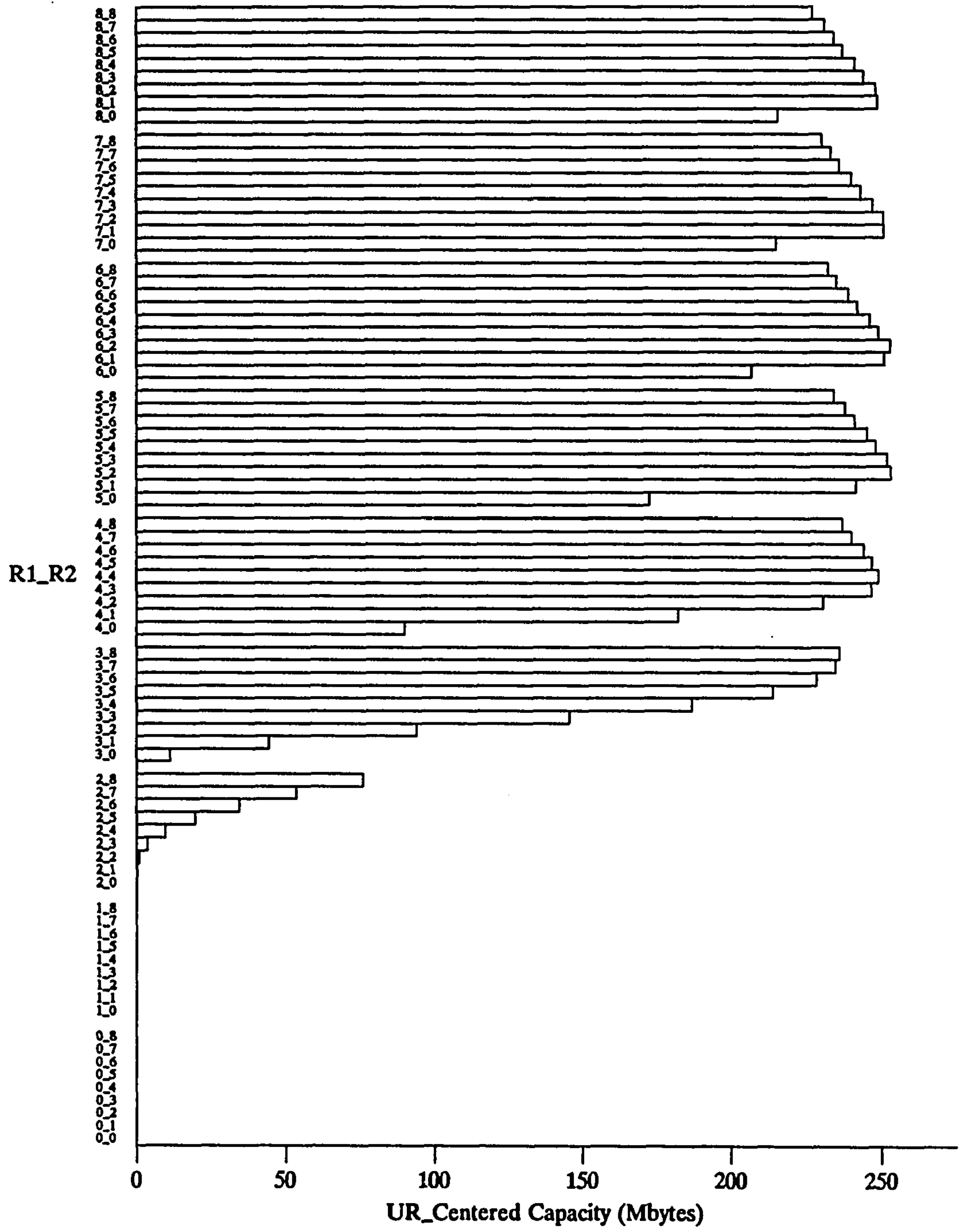

Figure 23. UR_Centered Yield Capacities (grouped by level 1 redundancy)

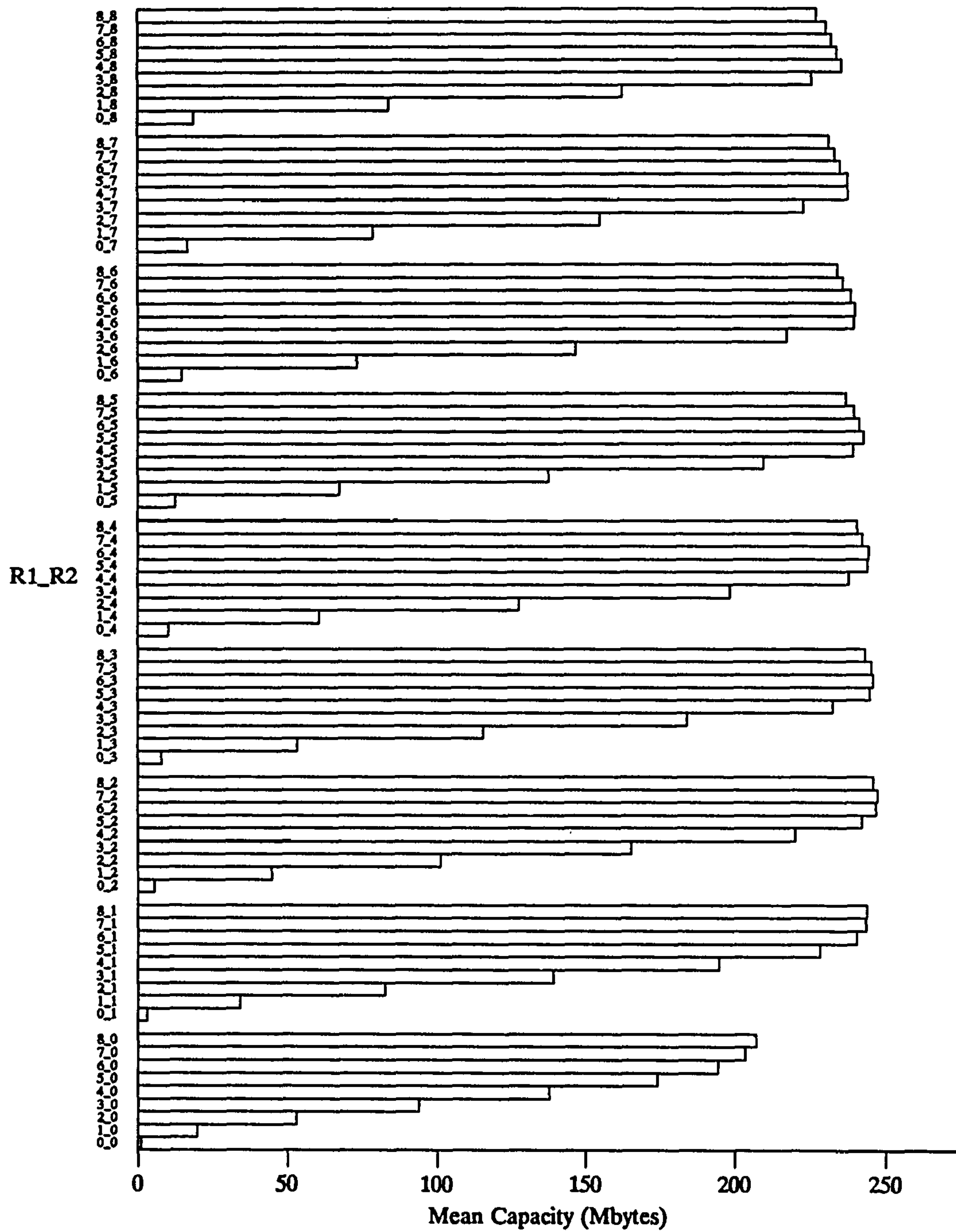

Figure 24. Mean Yield Capacities (grouped by level 2 redundancy)

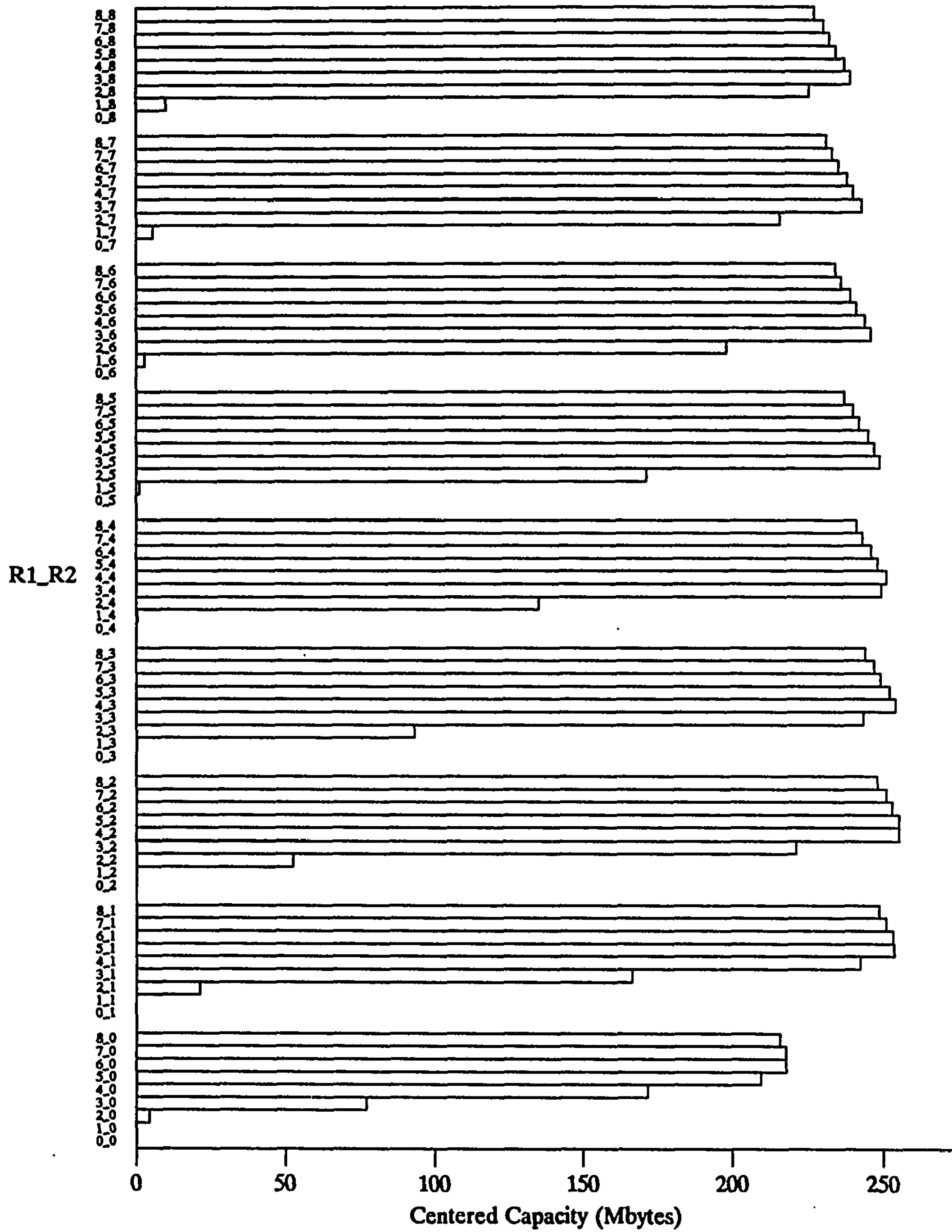

Figure 25. Centered Yield Capacities (grouped by level 2 redundancy)

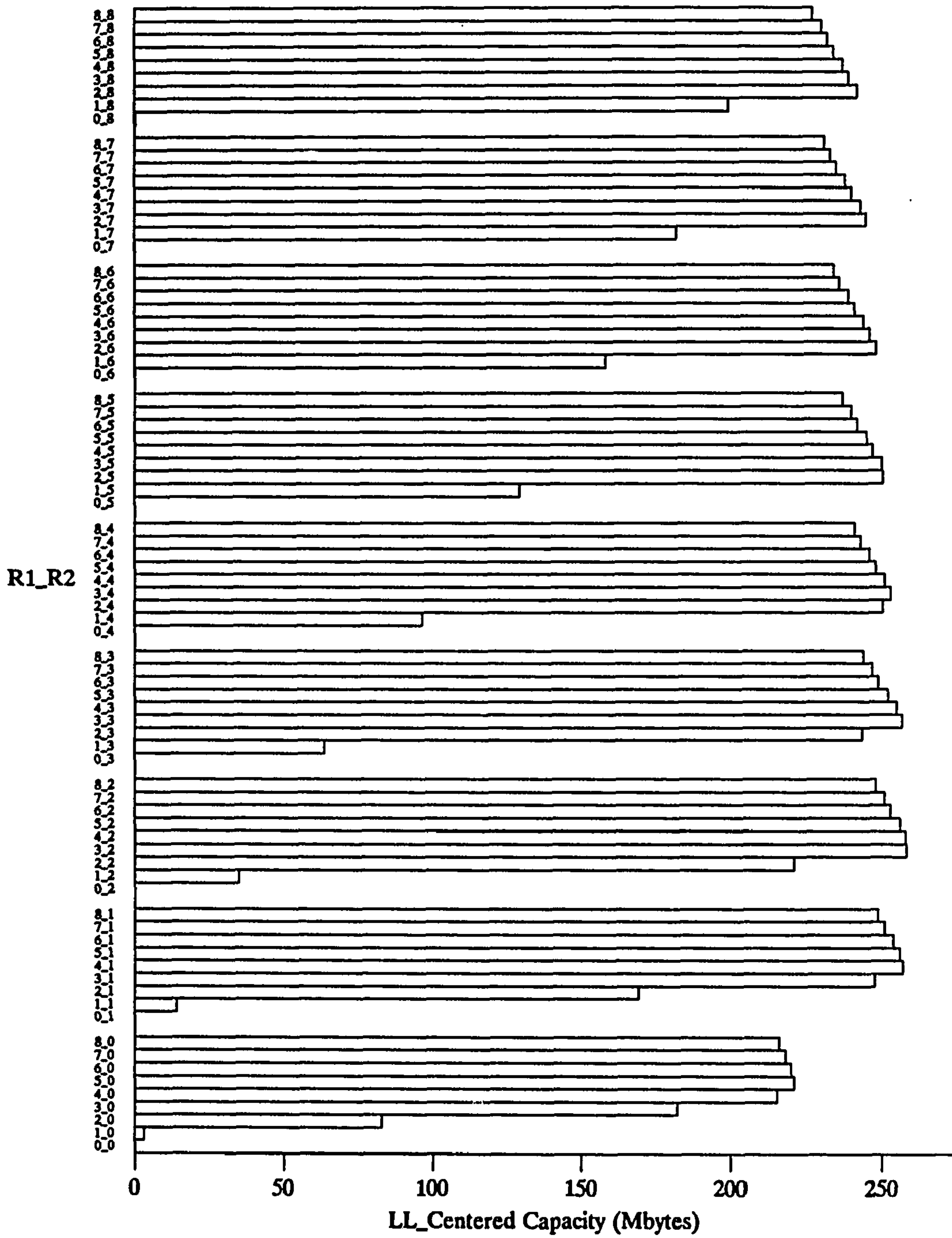

Figure 26. LL_Centered Yield Capacities (grouped by level 2 redundancy)

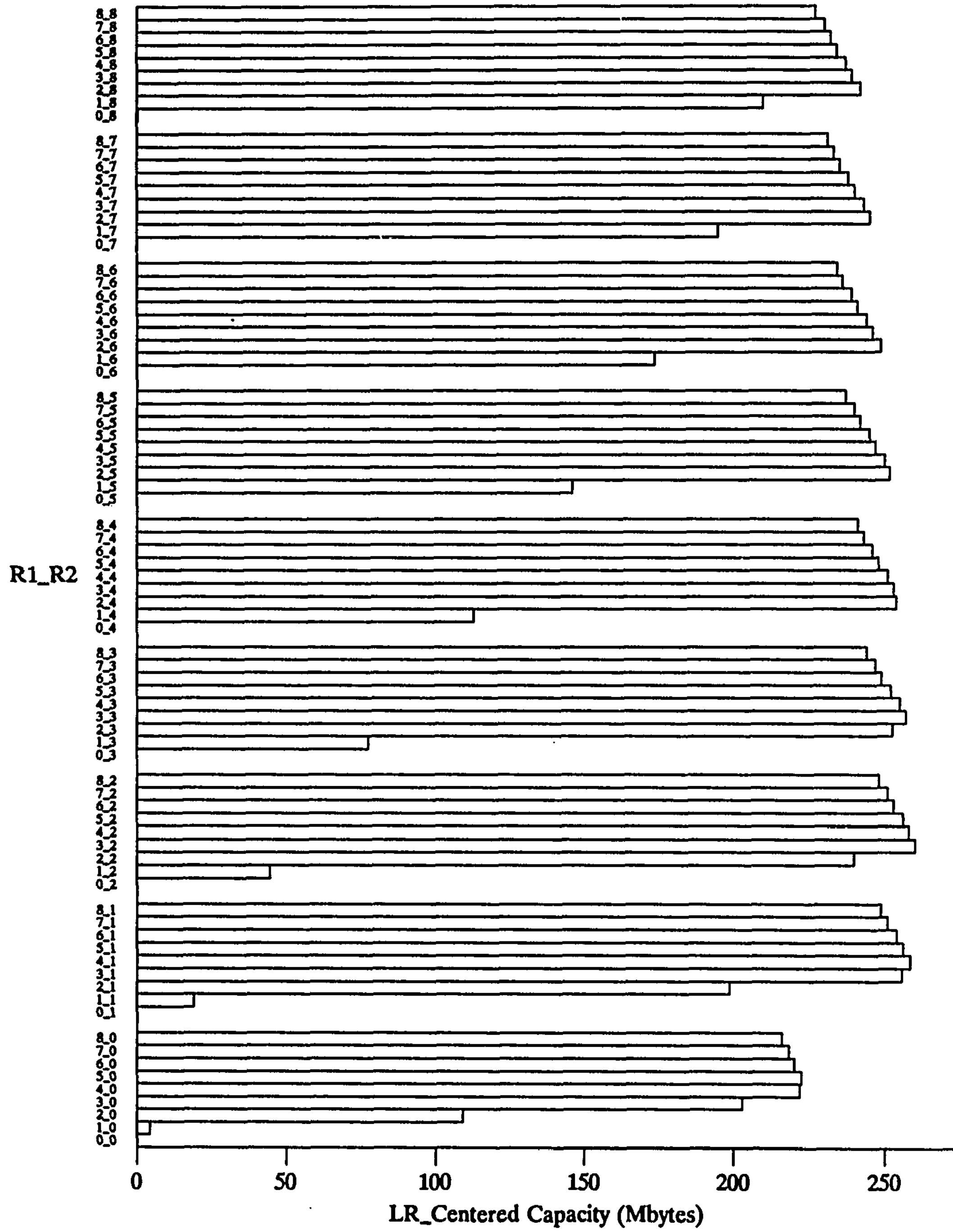

Figure 27. LR_Centered Yield Capacities (grouped by level 2 redundancy)

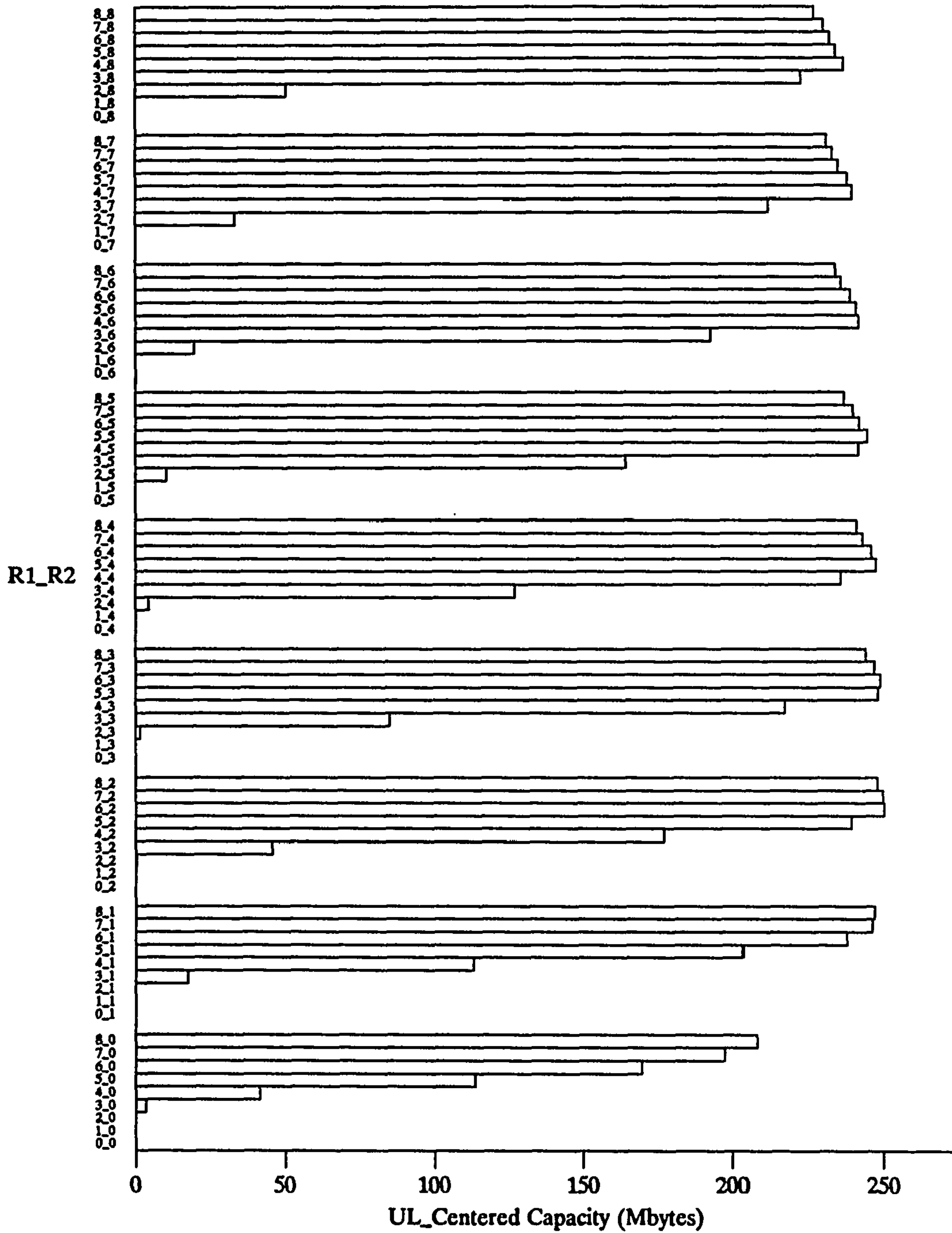

Figure 28. UL_Centered Yield Capacities (grouped by level 2 redundancy)

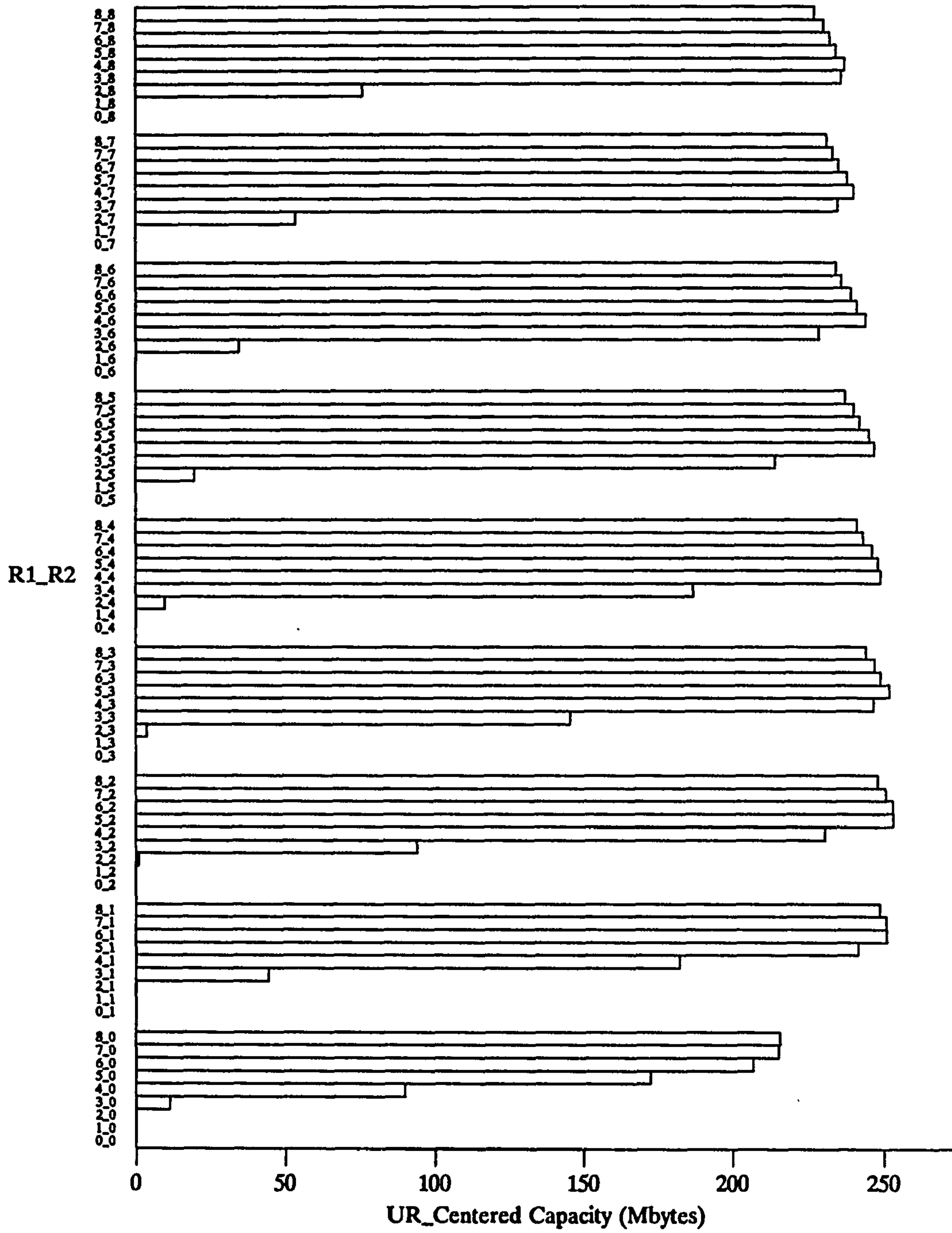

Figure 29. UR_Centered Yield Capacities (grouped by level 2 redundancy)