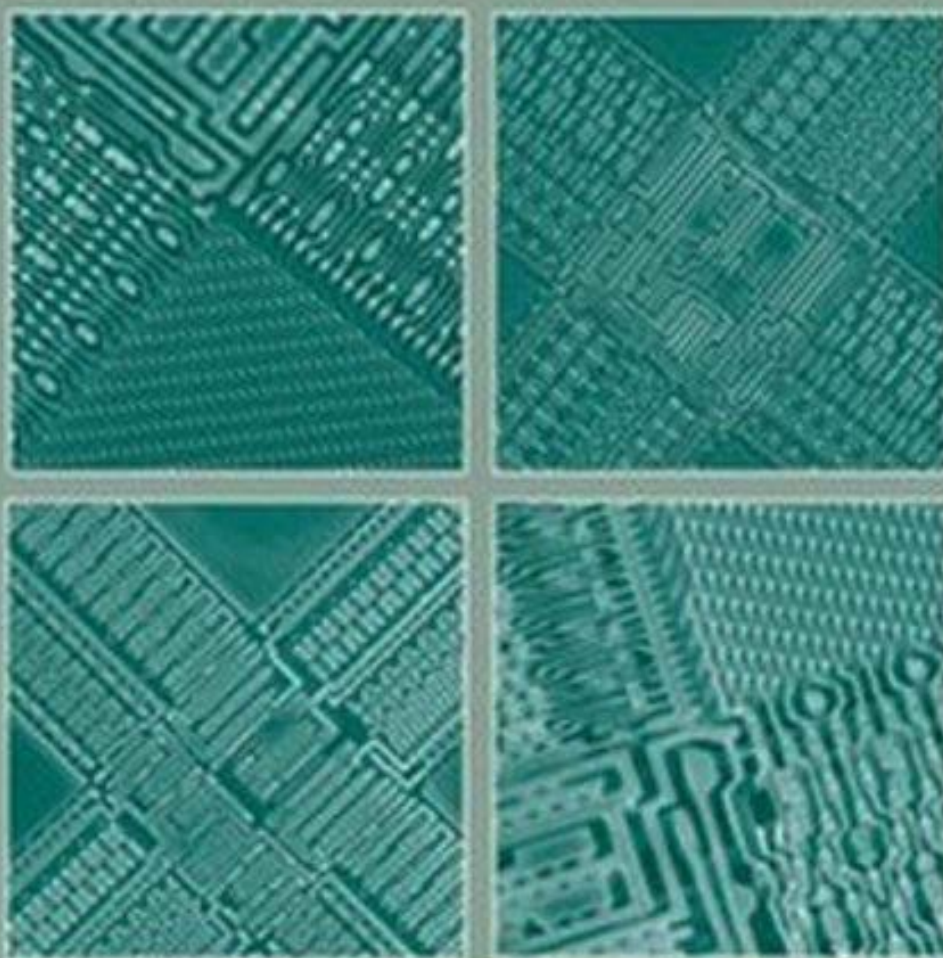


DRAM

CIRCUIT DESIGN

Fundamental and High-Speed Topics



Brent Keeth • R. Jacob Baker
Brian Johnson • Feng Lin

IEEE Press Series on Microelectronic Systems
Srinivas K. Trovati and Joe E. Duvvuri, Series Editors

DRAM Circuit Design

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

IEEE Press Editorial Board
Mohamed E. El-Hawary, *Editor in Chief*

R. Abari	T. G. Croda	S. V. Kartalopoulos
S. Basu	S. Farshchi	M. S. Newman
A. Chatterjee	B. M. Hammerli	
T. Chen	R. J. Herrick	

Kenneth Moore, *Director of IEEE Book and Information Services (BIS)*
Catherine Faduska, *Senior Acquisitions Editor*
Jeanne Audino, *Project Editor*

IEEE Solid-State Circuits Society, Sponsor
IEEE SSCS Liaison to the IEEE Press, Stuart K. Tewksbury

Technical Editing and Illustrations
Mary J. Miller, Micron Technology, Inc.

Technical Reviewers
Jim Frenzel, University of Idaho
Zhao Zhang, Iowa State University

DRAM Circuit Design

Fundamental and High-Speed Topics

Brent Keeth
R. Jacob Baker
Brian Johnson
Feng Lin

IEEE Press Series on Microelectronic Systems
Stuart K. Tewksbury and Joe E. Brewer, *Series Editors*

IEEE Solid-State Circuits Society, *Sponsor*



IEEE PRESS



WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication

Copyright © 2008 by the Institute of Electrical and Electronic Engineers, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Wiley Bicentennial Logo: Richard J. Pacifico

Library of Congress Cataloging-in-Publication Data is available.

ISBN 978-0-470-18475-2

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Preface	xi
----------------------	----

Chapter 1 An Introduction to DRAM

1.1 DRAM Types and Operation	1
1.1.1 The 1k DRAM (First Generation)	1
1.1.2 The 4k–64 Meg DRAM (Second Generation)	7
1.1.3 Synchronous DRAM (Third Generation)	15
1.2 DRAM Basics	22
1.2.1 Access and Sense Operations	24
1.2.2 Write Operation	28
1.2.3 Opening a Row (Summary)	29
1.2.4 Open/Folded DRAM Array Architectures	31

Chapter 2 The DRAM Array

2.1 The Mbit Cell	33
2.2 The Sense Amp	44
2.2.1 Equilibration and Bias Circuits	44
2.2.2 Isolation Devices	46
2.2.3 Input/Output Transistors	46
2.2.4 Nsense and Psense Amplifiers	47
2.2.5 Rate of Activation	49
2.2.6 Configurations	49
2.2.7 Operation	52
2.3 Row Decoder Elements	54
2.3.1 Bootstrap Wordline Driver	55
2.3.2 NOR Driver	57
2.3.3 CMOS Driver	58

2.3.4 Address Decode Tree	58
2.3.5 Static Tree	59
2.3.6 P&E Tree	59
2.3.7 Predecoding	60
2.3.8 Pass Transistor Tree	61
2.4 Discussion	61

Chapter 3 Array Architectures

3.1 Array Architectures	65
3.1.1 Open Digitline Array Architecture	65
3.1.2 Folded Array Architecture	75
3.2 Design Examples: Advanced Bilevel DRAM Architecture	83
3.2.1 Array Architecture Objectives	84
3.2.2 Bilevel Digitline Construction	85
3.2.3 Bilevel Digitline Array Architecture	88
3.2.4 Architectural Comparison	93

Chapter 4 The Peripheral Circuitry

4.1 Column Decoder Elements	99
4.2 Column and Row Redundancy	102
4.2.1 Row Redundancy	104
4.2.2 Column Redundancy	107

Chapter 5 Global Circuitry and Considerations

5.1 Data Path Elements	111
5.1.1 Data Input Buffer	111
5.1.2 Data Write Muxes	115
5.1.3 Write Driver Circuit	116
5.1.4 Data Read Path	118
5.1.5 DC Sense Amplifier (DCSA)	119
5.1.6 Helper Flip-Flop (HFF)	121
5.1.7 Data Read Muxes	122
5.1.8 Output Buffer Circuit	124
5.1.9 Test Modes	125
5.2 Address Path Elements	126
5.2.1 Row Address Path	126
5.2.2 Row Address Buffer	127
5.2.3 CBR Counter	127
5.2.4 Predecode Logic	128
5.2.5 Refresh Rate	128
5.2.6 Array Buffers	130
5.2.7 Phase Drivers	131

5.2.8 Column Address Path	131
5.2.9 Address Transition Detection	132
5.3 Synchronization in DRAMs	135
5.3.1 The Phase Detector	137
5.3.2 The Basic Delay Element	137
5.3.3 Control of the Shift Register	138
5.3.4 Phase Detector Operation	139
5.3.5 Experimental Results	140
5.3.6 Discussion	142
Chapter 6 Voltage Converters	
6.1 Internal Voltage Regulators	147
6.1.1 Voltage Converters	147
6.1.2 Voltage References	148
6.1.3 Bandgap Reference	153
6.1.4 The Power Stage	154
6.2 Pumps and Generators	158
6.2.1 Pumps	158
6.2.2 DVC2 Generator	165
6.3 Discussion	165
Chapter 7 An Introduction to High-Speed DRAM	
7.1 The Performance Paradigm	167
7.2 Performance for DRAM Memory Devices	169
7.3 Underlying Technology Improvements	171
Chapter 8 High-Speed Die Architectures	
8.1 Introduction: Optimizing DRAM Architecture for High Performance	173
8.2 Architectural Features: Bandwidth, Latency, and Cycle Time	175
8.2.1 Architectural Limiters: The Array Data Path	176
8.2.2 Architectural Limiters: The Read Data Path	183
8.2.3 Architectural Limiters: Latency	186
8.3 Conclusion: Designing for High Performance	190
Chapter 9 Input Circuit Paths	
9.1 Introduction	193
9.2 Input Receivers	196
9.3 Matched Routing	200
9.4 Capture Latches	203
9.5 Input Timing Adjustments	206
9.6 Current Mode Logic (CML)	212

Chapter 10 Output Circuit Paths

10.1 Transmission Line, Impedance, and Termination.	220
10.2 Impedance Control	224
10.3 Simultaneous Switching Noise (SSN)	230
10.4 Signal Return Path Shift (SRPS).	238
10.5 Electrostatic Discharge (ESD)	242
10.6 Parallel-to-Serial Conversion	244
10.7 Emerging Memory I/O Features	248

Chapter 11 Timing Circuits

11.1 Introduction.	251
11.2 All-Digital Clock Synchronization Design.	254
11.2.1 Timing Analysis	255
11.2.2 Digital Delay Line.	259
11.2.3 Phase Detector (PD)	267
11.2.4 Test and Debug	273
11.2.5 Dual-Loop Architecture	274
11.3 Mixed-Mode Clock Synchronization Design	275
11.3.1 Analog Delay Line	276
11.3.2 Charge-Pump Phase Detector (CPPD)	283
11.3.3 Dual-Loop Analog DLL	286
11.3.4 Mixed-Mode DLL and Its Applications	289
11.4 What's Next for Timing	291

Chapter 12 Control Logic Design

12.1 Introduction.	295
12.2 DRAM Logic Styles	298
12.2.1 Process Limitations	298
12.2.2 Array Operation.	301
12.2.3 Performance Requirements	304
12.2.4 Delay-Chain Logic Style.	306
12.2.5 Domino Logic	309
12.2.6 Testability	314
12.3 Command and Address Control	317
12.3.1 Command Decoder	318
12.3.2 Read and Write Data Address Registers	324
12.3.3 Column Access Control.	328
12.4 Write Data Latency Timing and Data Demultiplexing.	330
12.4.1 Write Latency Timing	332
12.4.2 Write Data Demultiplexing	338
12.5 Read Data Latency Timing and Data Multiplexing	346
12.5.1 Read Data FIFO.	350

12.5.2 Read Latency (CL) Tracking 359

12.6 Comments on Future Direction for DRAM Logic Design ... 367

Chapter 13 Power Delivery

13.1 Power Delivery Network Design..... 373

13.2 Device/Package Co-design 376

13.3 Full-Chip Simulations 380

Chapter 14 Future Work in High-Performance Memory 385

Appendix..... 391

Glossary 407

Index 413

The DRAM Array

This chapter begins a more detailed examination of standard DRAM array elements. This examination is necessary for a clear understanding of fundamental DRAM elements and how they are used in memory block construction. A common point of reference is required before considering the analysis of competing array architectures. Included in this chapter is a detailed discussion of mbits, array configurations, sense amplifier elements, and row decoder elements.

2.1 THE MBIT CELL

The primary advantage of DRAM over other types of memory technology is low cost. This advantage arises from the simplicity and scaling characteristics of its 1T1C memory cell [1]. Although the DRAM mbit is simple conceptually, its actual design and implementation are highly complex. Therefore, successful, cost-effective DRAM designs require a tremendous amount of process technology.

Figure 2.1 presents the layout of a modern buried capacitor DRAM mbit pair. (*Buried* means that the capacitor is below the digitline.) This type of mbit is also referred to as a *bitline over capacitor* (BOC) cell. Because sharing a contact significantly reduces overall cell size, DRAM mbits are constructed in pairs. In this way, a digitline contact can be shared. The mbits comprise an active area rectangle (in this case, an n⁺ active area), a pair of polysilicon wordlines, a single digitline contact, a metal or polysilicon digitline, and a pair of cell capacitors formed with an oxide-nitride-oxide dielectric between two layers of polysilicon. For most processes, the wordline polysilicon is silicided to reduce sheet resistance, permitting longer wordline segments without reducing speed. The mbit layout, as shown in

Figure 2.1, is essentially under the control of process engineers, for every aspect of the mbit must meet stringent performance and yield criteria.

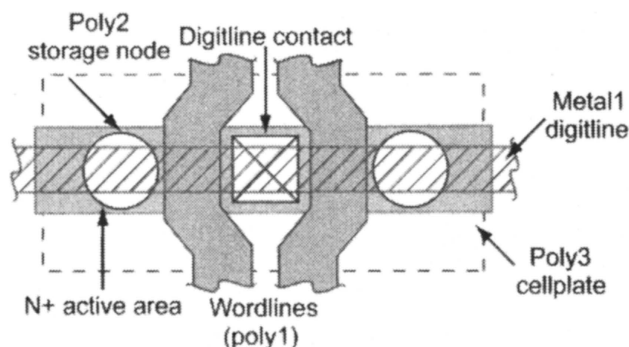


Figure 2.1 Mbit pair layout.

A small array of mbits appears in Figure 2.2. This figure is useful to illustrate several features of the mbit. First, note that the digitline pitch (width plus space) dictates the active area pitch and the capacitor pitch. Process engineers adjust the active area width and the field oxide width to maximize transistor drive and minimize transistor-to-transistor leakage. Field oxide technology greatly impacts this balance. A thicker field oxide or a shallower junction depth affords a wider transistor active area. Second, the wordline pitch (width plus space) dictates the space available for the digitline contact, transistor length, active area, field poly width, and capacitor length. Optimization of each of these features by process engineers is necessary to maximize capacitance, minimize leakage, and maximize yield. Contact technology, subthreshold transistor characteristics, photolithography, and etch and film technology dictate the overall design.

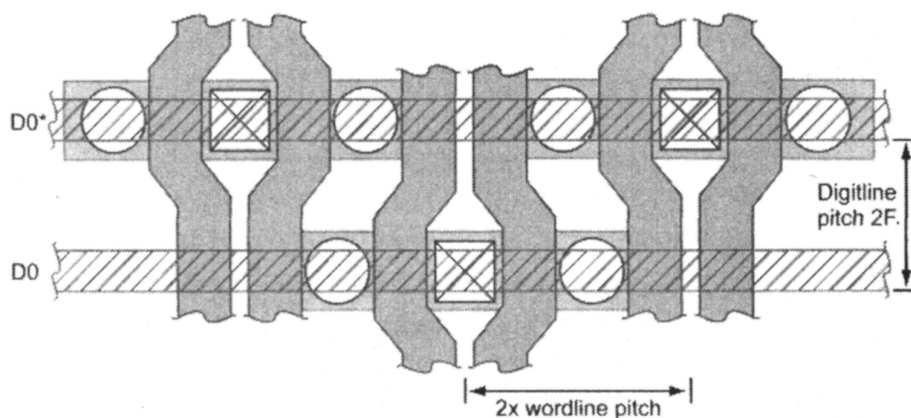


Figure 2.2 Layout to show array pitch.

At this point in the discussion, it is appropriate to introduce the concept of feature size and how it relates to cell size. The mbit shown in Figures 2.1

and 2.2 is by definition an eight-square feature cell ($8F^2$) [2][3]. The intended definition of feature (F) in this case is minimum realizable process dimension but in fact equates to a dimension that is one-half the wordline (row) or digitline (column) pitch. A 45 nm process having wordline and digitline pitches of 100 nm yields an mbit size that is

$$(8) \cdot 50nm = 0.02\mu m^2 \quad (2.1)$$

It is easier to explain the $8F^2$ designation with the aid of Figure 2.3. An imaginary box drawn around the mbit defines the cell's outer boundary. Along the x-axis, this box includes one-half digitline contact feature, one wordline feature, one capacitor feature, one field poly feature, and one-half poly space feature for a total of four features. Along the y-axis, this box contains two one-half field oxide features and one active area feature for a total of two features. The area of the mbit is therefore

$$4F \cdot 2F = 8F^2 \quad (2.2)$$

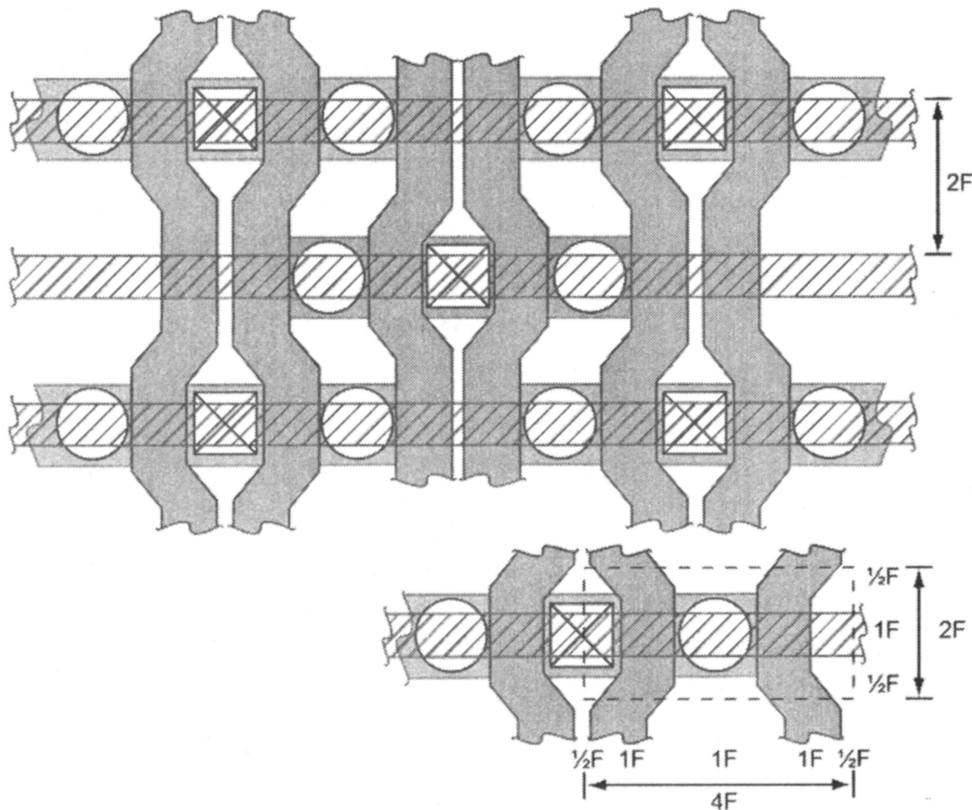


Figure 2.3 Layout to show $8F^2$ derivation.

The folded array architecture, as shown in Figure 2.2, always produces an $8F^2$ mbit. This results from the fact that each wordline connects or forms a crosspoint with an mbit transistor on every other digitline and must pass around mbit transistors as field poly on the remaining digitlines. The field poly in each mbit cell adds two square features to what would have been a $6F^2$ cell. Although the folded array yields a cell that is 25% larger than other array architectures, it also produces superior signal-to-noise performance, especially when combined with some form of digitline twisting [4]. Superior low-noise performance made folded array architecture the architecture of choice since the 64kbit generation [5].

A folded array is schematically depicted in Figure 2.4. Sense amplifier circuits placed at the edge of each array connect to both true and complement digitlines (D and D^*) coming from a single array. Optional digitline pair twisting in one or more positions reduces and balances the coupling to adjacent digitline pairs and improves overall signal-to-noise characteristics [4]. Figure 2.5 shows the variety of twisting schemes used throughout the DRAM industry [6].

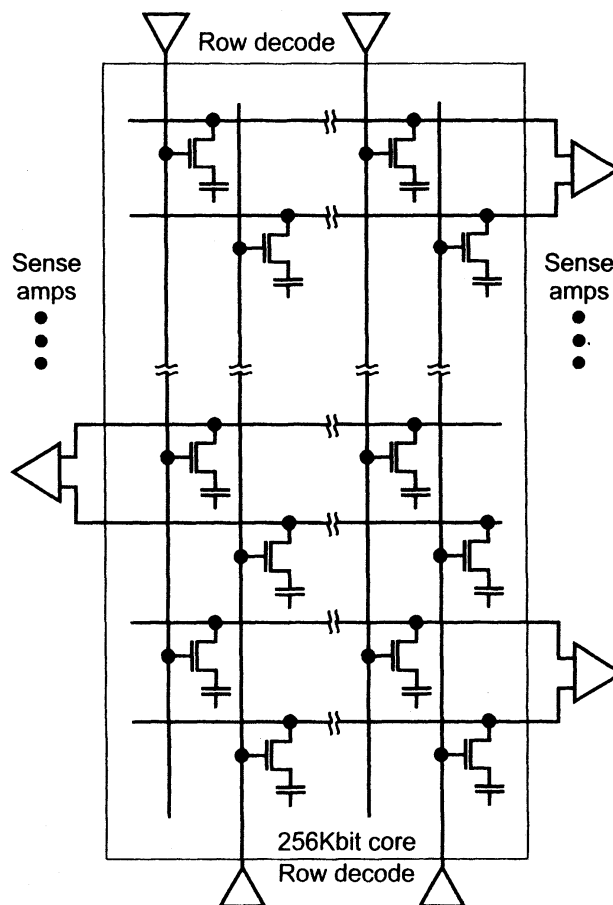


Figure 2.4 Folded digitline array schematic.

Ideally, a twisting scheme equalizes the coupling terms from each digitline to all other digitlines, both true and complement. If implemented properly, the noise terms cancel or at least produce only common-mode noise to which the differential sense amplifier is more immune.

Each digitline twist region consumes valuable silicon area. Thus, design engineers resort to the simplest and most efficient twisting scheme to get the job done. Because the coupling between adjacent metal lines is inversely proportional to the line spacing, the signal-to-noise problem gets increasingly worse as DRAMs scale to smaller and smaller dimensions. Hence, the industry trend is toward use of more complex twisting schemes on succeeding generations [6][7].

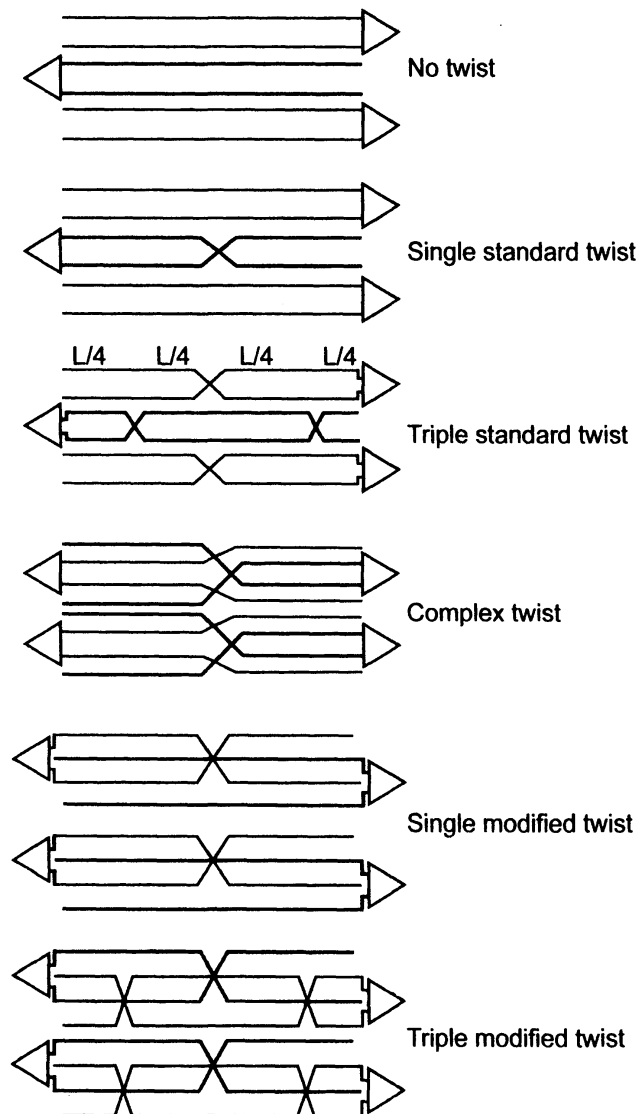


Figure 2.5 Digitline twisting scheme.

An alternative to the folded array architecture, popular prior to the 64kbit generation [1], was the open digitline architecture. Seen schematically in Figure 2.6, this architecture also features the sense amplifier circuits between two sets of arrays [8]. Unlike the folded array, however, true and complement digitlines (D and D^*) connected to each sense amplifier pair come from separate arrays [9]. This arrangement precludes using digitline twisting to improve signal-to-noise performance, which is the prevalent reason why the industry switched to folded arrays. Note that unlike the folded array architecture, each wordline in an open digitline architecture connects to mbit transistors on every digitline, creating crosspoint-style arrays.

This feature permits a 25% reduction in mbit size to only $6F^2$ because the wordlines do not have to pass alternate mbits as field poly. The layout for an array of standard $6F^2$ mbit pairs is shown in Figure 2.7 [2]. A box is drawn around one of the mbits to show the $6F^2$ cell boundary. Again, two mbits share a common digitline contact to improve layout efficiency. While folded array architectures were used extensively in second-generation DRAMs [8] and the open array architectures, as just mentioned were used extensively in first-generation DRAMs, the third-generation designs are seeing a return to the open architecture. As the device/array size continue to shrink with scaling, the effects of coupled noise on the arrays is reduced, which makes the open array architecture both viable and preferable (due to the reduced size).

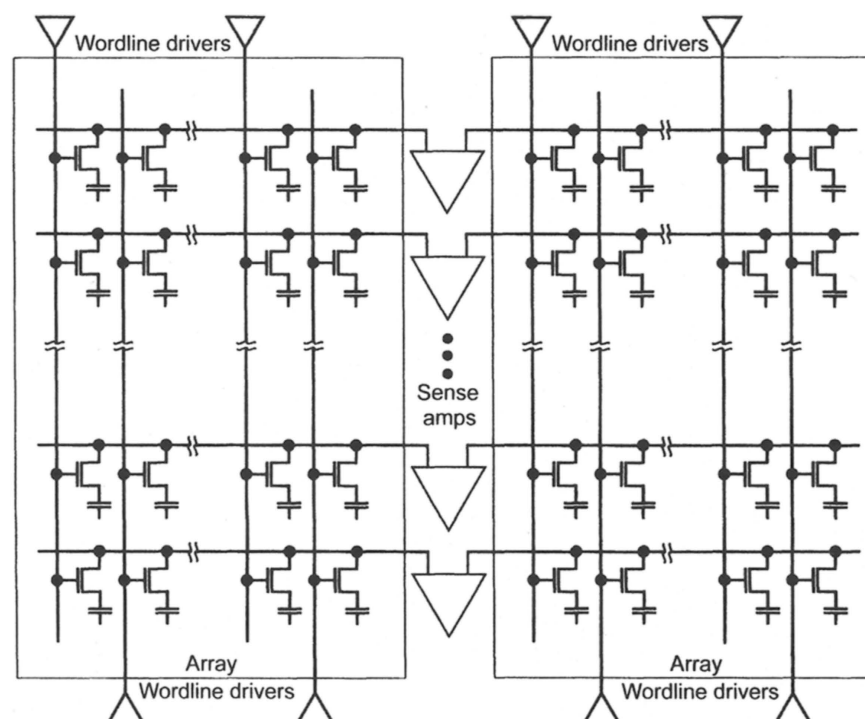


Figure 2.6 Open digitline array schematic.

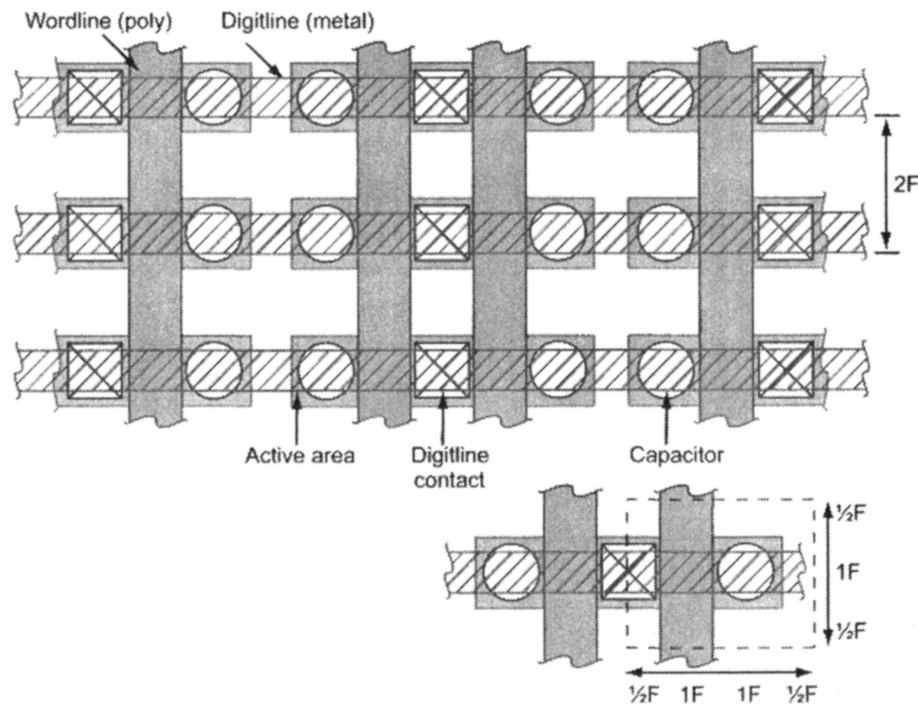


Figure 2.7 Open digitline array layout.
(Feature size, F , is equal to one-half digitline pitch.)

Digitline capacitive components, contributed by each mbit, include junction capacitance, digitline-to-cellplate (poly3), digitline-to-wordline, digitline-to-digitline, digitline-to-substrate, and, in some cases, digitline-to-storage cell (poly2) capacitance. Therefore, each mbit connected to the digitline adds a specific amount of capacitance to the digitline. Most modern DRAM designs have no more than 512 mbits connected to a digitline segment.

Two factors dictate this quantity. First, for a given cell size, as determined by row and column pitches, a maximum storage capacitance can be achieved without resorting to exotic processes or excessive cell height. For processes in which the digitline is above the storage capacitor (buried capacitor), contact technology determines the maximum allowable cell height. This fixes the volume available (cell area multiplied by cell height) in which to build the storage capacitor. Second, as the digitline capacitance increases, the power associated with charging and discharging this capacitance during Read and Write operations increases. Any given wordline essentially accesses (crosses) all of the columns within a DRAM. For a 256-Meg DRAM, each wordline crosses 16,384 columns. With a multiplier such as that, it is easy to appreciate why limits to digitline capacitance are necessary to keep power dissipation low.

Figure 2.8 presents a process cross section for the buried capacitor mbit depicted in Figure 2.2, and Figure 2.9 shows a SEM image of the buried capacitor mbit. This type of mbit, employing a buried capacitor structure, places the digitline physically above the storage capacitor [10]. The digitline is constructed from either metal or polycide, while the digitline contact is formed using a metal or polysilicon plug technology. The mbit capacitor is formed with polysilicon (poly2) as the bottom plate, an oxide-nitride-oxide (ONO) dielectric, and a sheet of polysilicon (poly3). This top sheet of polysilicon becomes a common node shared by all mbit capacitors. The capacitor shape can be simple, such as a rectangle, or complex, such as concentric cylinders or stacked discs. The most complex capacitor structures are the topic of many DRAM process papers [11][12][13].

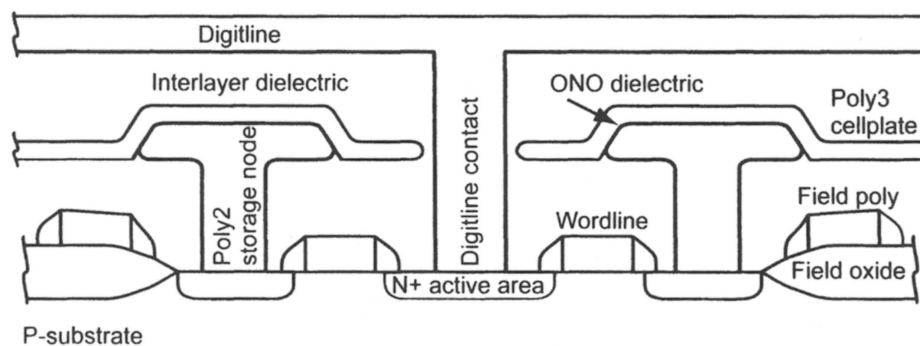


Figure 2.8 Buried capacitor cell process cross section.

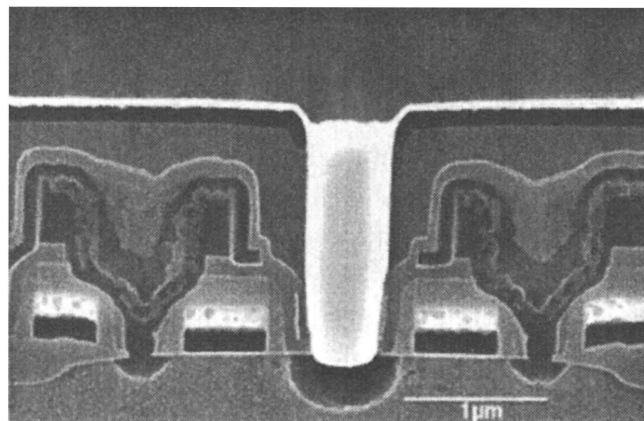


Figure 2.9 Buried capacitor cell process SEM image.

The ONO dielectric undergoes optimization to achieve maximum capacitance with minimum leakage. It must also tolerate the maximum DRAM operating voltage without breaking down. For this reason, the cellplate (poly3) is normally biased at $+V_{CC}/2$ V, ensuring that the dielectric

has no more than $V_{CC}/2$ V across it for either stored logic state, a logic one at $+V_{CC}/2$ V or a logic zero at $-V_{CC}/2$ V.

Two other basic mbit configurations are used in the DRAM industry. The first, shown in Figures 2.10, 2.11, and 2.12, is referred to as a *buried digitline* or *capacitor over bitline* (COB) cell [14][15]. The digitline in this cell is almost always formed of polysilicon rather than of metal.

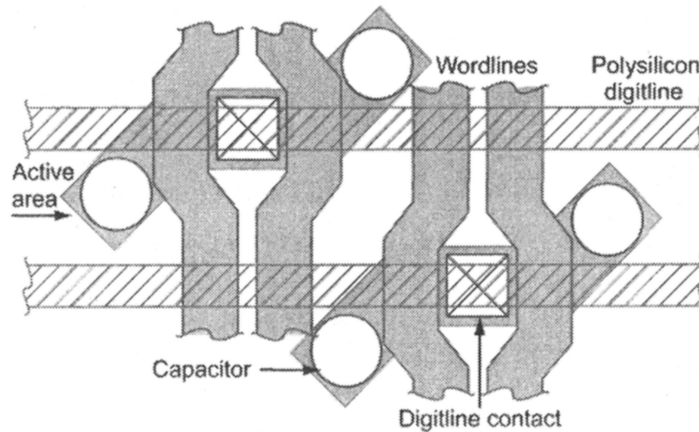


Figure 2.10 Buried digitline Mbit cell layout.

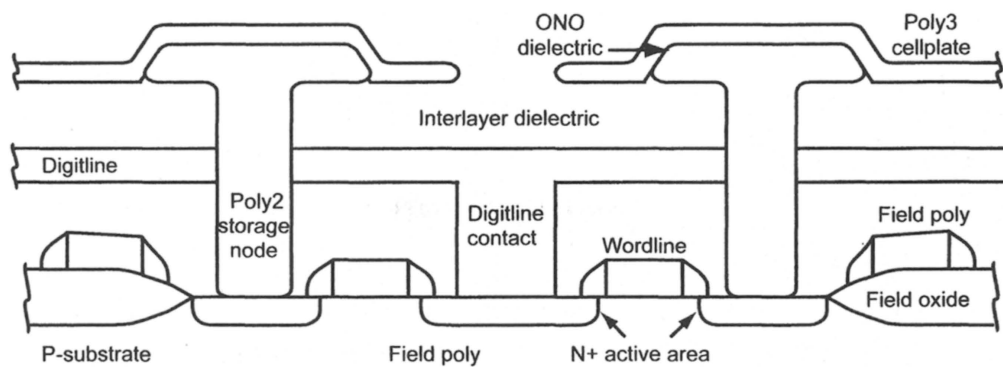


Figure 2.11 Buried digitline mbit process cross section.

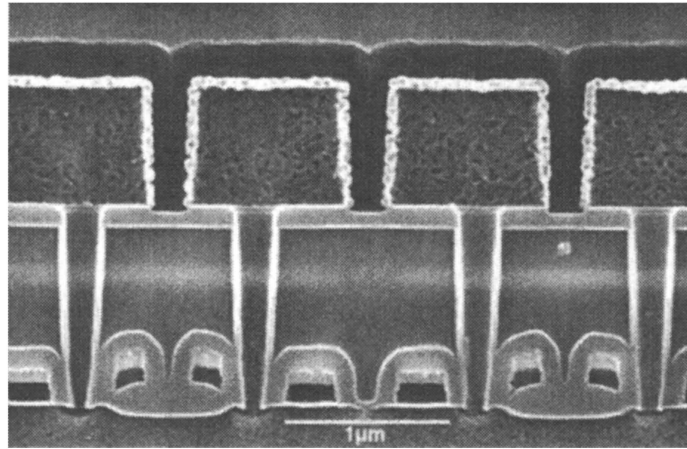


Figure 2.12 Buried digitline Mbit process SEM image.

As viewed from the top, the active area is normally bent or angled to accommodate the storage capacitor contact that must drop between digitlines. An advantage of the buried digitline cell over the buried capacitor cell of Figure 2.8 is that its digitline is physically very close to the silicon surface, making digitline contacts much easier to produce. The angled active area, however, reduces the effective active area pitch, constraining the isolation process even further. In buried digitline cells, it is also very difficult to form the capacitor contact. Because the digitline is at or near minimum pitch for the process, insertion of a contact between digitlines can be difficult.

Figures 2.13 and 2.14 present a process cross section of the third type of mbit used in the construction of DRAMs. Using trench storage capacitors, this cell is accordingly called a *trench cell* [12][13]. Trench capacitors are formed in the silicon substrate, rather than above the substrate, after etching deep holes into the wafer. The storage node is a doped polysilicon plug, which is deposited in the hole following growth or deposition of the capaci-

tor dielectric. Contact between the storage node plug and the transistor drain is usually made through a poly strap.

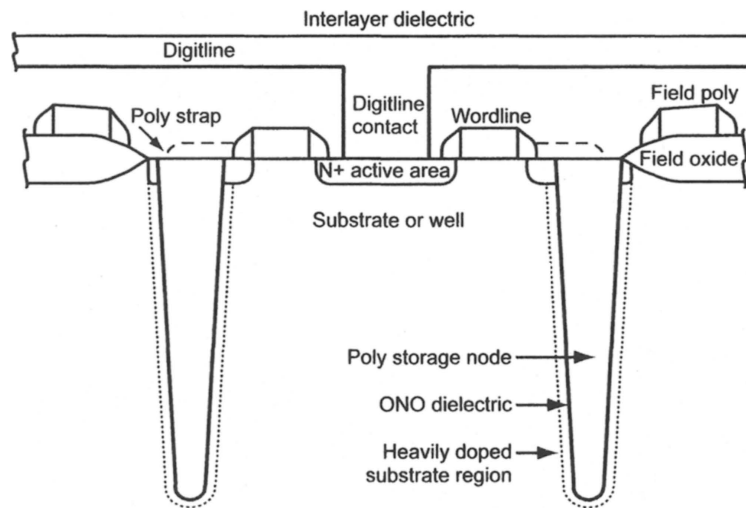


Figure 2.13 Trench capacitor Mbit process cross section.

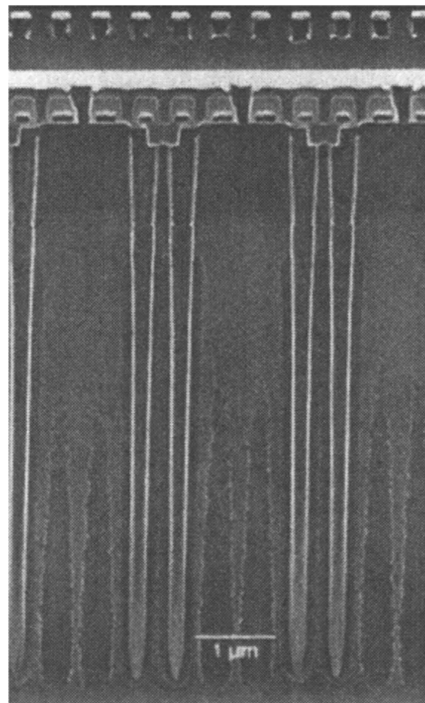


Figure 2.14 Trench capacitor Mbit process SEM image.

With most trench capacitor designs, the substrate serves as the common-node connection to the capacitors, preventing the use of $+V_{CC}/2$ bias and thinner dielectrics. The substrate is heavily doped around the capacitor to reduce resistance and improve the capacitor's CV characteristics. A real

advantage of the trench cell is that the capacitance can be increased by merely etching a deeper hole into the substrate [16].

Furthermore, the capacitor does not add stack height to the design, greatly simplifying contact technology. The disadvantage of trench capacitor technology is the difficulty associated with reliably building capacitors in deep silicon holes and connecting the trench capacitor to the transistor drain terminal.

2.2 THE SENSE AMP

The term *sense amplifier* actually refers to a collection of circuit elements that pitch up to the digitlines of a DRAM array. This collection most generally includes isolation transistors, devices for digitline equilibration and bias, one or more Nsense amplifiers, one or more Psense amplifiers, and devices connecting selected digitlines to *I/O* signal lines. All of the circuits along with the wordline driver circuits, as discussed in Section 2.3, are called *pitch cells*. This designation comes from the requirement that the physical layout for these circuits is constrained by the digitline and wordline pitches of an array of mbits. For example, the sense amplifiers for a specific digitline pair (column) are generally laid out within the space of four digitlines. With one sense amplifier for every four digitlines, this is commonly referred to as quarter pitch or four pitch.

2.2.1 Equilibration and Bias Circuits

The first elements analyzed are the equilibration and bias circuits. As we can recall from the discussions of DRAM operation in Section 1.1, the digitlines start at $V_{CC}/2$ V prior to cell Access and Sensing [17]. It is vitally important to the Sensing operation that both digitlines, which form a column pair, are at the same voltage before the wordline is driven HIGH. Any offset voltage appearing between the pair directly reduces the effective signal produced during the Access operation [5]. Equilibration of the digitlines is accomplished with one or more NMOS transistors connected between the digitline conductors. NMOS are used because of their higher drive capability and the resulting faster equilibration.

An equilibration transistor, together with bias transistors, is shown in Figure 2.15. The gate terminal is connected to a signal called *equilibrate* (*EQ*). *EQ* is held to V_{CC} whenever the external row address strobe signal *RAS** is HIGH. This indicates an inactive or *PRECHARGE* state for the DRAM. After *RAS** has fallen, *EQ* transitions LOW, turning the equilibration transistor OFF just prior to a wordline going HIGH. *EQ* will again transition HIGH at the end of a *RAS** cycle to force equilibration of the

digitlines. The equilibration transistor is sized large enough to ensure rapid equilibration of the digitlines to prepare the part for a subsequent access.

As shown in Figure 2.15, two more NMOS transistors accompany the *EQ* transistor to provide a bias level of $V_{CC}/2$ V. These devices operate in conjunction with equilibration to ensure that the digitline pair remains at the prescribed voltage for Sensing. Normally, digitlines that are at V_{CC} and ground equilibrate to $V_{CC}/2$ V [5]. The bias devices ensure that this occurs and also that the digitlines remain at $V_{CC}/2$, despite leakage paths that would otherwise discharge them. Again, for the same reasons as for the equilibration transistor, NMOS transistors are used. Most often, the bias and equilibration transistors are integrated to reduce their overall size. $V_{CC}/2$ V *PRECHARGE* is used on most modern DRAMs because it reduces power consumption and Read-Write times and improves Sensing operations. Power consumption is reduced because a $V_{CC}/2$ *PRECHARGE* voltage can be obtained by equilibrating the digitlines (which are at V_{CC} and ground, respectively) at the end of each cycle.

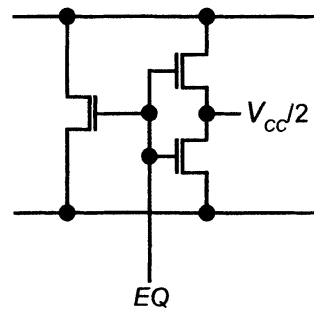


Figure 2.15 Equilibration schematic.

The charge-sharing between the digitlines produces $V_{CC}/2$ without additional I_{CC} current. Some designs, however, equilibrate the digitlines to V_{DD} [18]. Because the wordlines and digitlines are both at V_{CC} when the part is inactive, row-to-column shorts that exist do not contribute to increased standby current. $V_{CC}/2$ *PRECHARGE* DRAMs, on the other hand, suffer higher standby current with row-to-column shorts because the wordlines and digitlines are at different potentials when the part is inactive. A layout for the equilibration and bias circuits is shown in Figure 2.16.

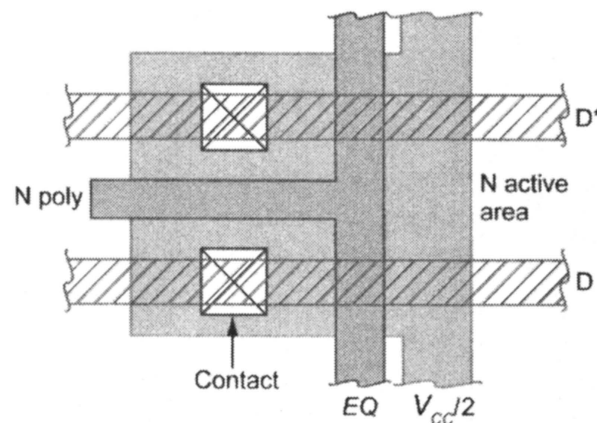


Figure 2.16 Equilibration and bias circuit layout.

2.2.2 Isolation Devices

Isolation devices are important to the sense amplifier circuits. These devices are NMOS transistors placed between the array digitlines and the sense amplifiers (see Figure 2.19). Isolation transistors are physically located on both ends of the sense amplifier layout. In quarter-pitch sense amplifier designs, there is one isolation transistor for every two digitlines. Although this is twice the active area width and space of an array, it nevertheless sets the limit for isolation processing in the pitch cells.

The isolation devices provide two functions. First, if the sense amps are positioned between and connected to two arrays, they electrically isolate one of the two arrays. This is necessary whenever a wordline fires in one array because isolation of the second array reduces the digitline capacitance driven by the sense amplifiers, thus speeding Read-Write times, reducing power consumption, and extending Refresh for the isolated array. Second, the isolation devices provide resistance between the sense amplifier and the digitlines. This resistance stabilizes the sense amplifiers and speeds up the Sensing operation by somewhat isolating the highly capacitive digitlines from the low-capacitance sense nodes [19]. Capacitance of the sense nodes between isolation transistors is generally less than 15fF, permitting the sense amplifier to latch much faster than if it were solidly connected to the digitlines. The isolation transistors slow Write-Back to the mbits, but this is far less of a problem than initial Sensing.

2.2.3 Input/Output Transistors

The input/output (*I/O*) transistors allow data to be read from and written to specific digitline pairs. A single *I/O* transistor is connected to each sense node as shown in Figure 2.17. The outputs of each *I/O* transistor are connected to *I/O* signal pairs. Commonly, there are two pairs of *I/O* signal lines,

which permit four *I/O* transistors to share a single *column select* (*CSEL*) control signal. DRAM designs employing two or more metal layers run the column select lines across the arrays in either Metal2 or Metal3. Each column select can activate four *I/O* transistors on each side of an array to connect four digitline pairs (columns) to peripheral data path circuits. The *I/O* transistors must be sized carefully to ensure that instability is not introduced into the sense amplifiers by the *I/O* bias voltage or remnant voltages on the *I/O* lines.

Although designs vary significantly as to the numerical ratio, *I/O* transistors are designed to be two to eight times smaller than the *Nsense* amplifier transistors. This is sometimes referred to as *beta ratio*. A beta ratio between five and eight is considered standard; however, it can only be verified with silicon. Simulations may fail to adequately predict sense amplifier instability, although theory would predict better stability with higher beta ratio and better Write times with lower beta ratio. During Write, the sense amplifier remains ON and must be overdriven by the Write driver (see Section 1.2.2).

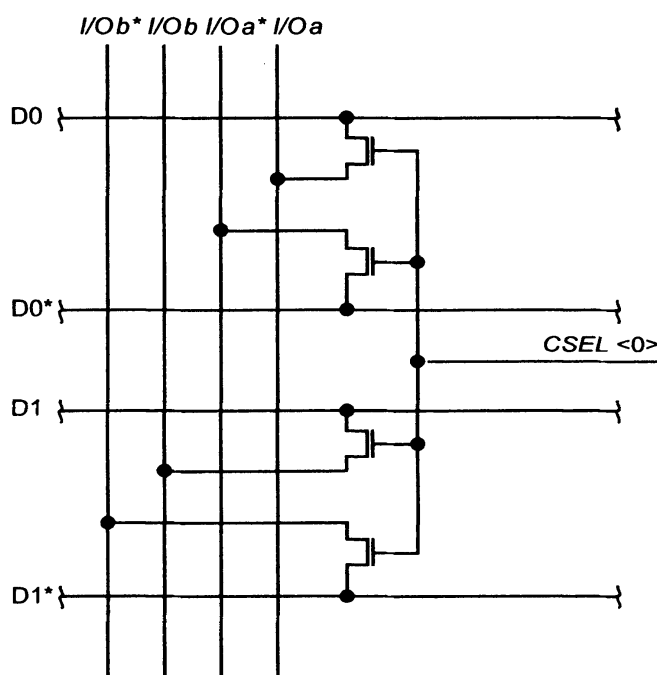


Figure 2.17 *I/O* transistors.

2.2.4 *Nsense* and *Psense* Amplifiers

The fundamental elements of any sense amplifier block are the *Nsense* amplifier and the *Psense* amplifier. These amplifiers, as previously discussed, work together to detect the access signal voltage and drive the digitlines, accordingly to V_{CC} and ground. The *Nsense* amplifier depicted in Figure 2.18 consists of cross-coupled NMOS transistors. The *Nsense* ampli-

fier drives the low-potential digitline to ground. Similarly, the Psense amplifier consists of cross-coupled PMOS transistors and drives the HIGH-potential digitline to V_{CC} .

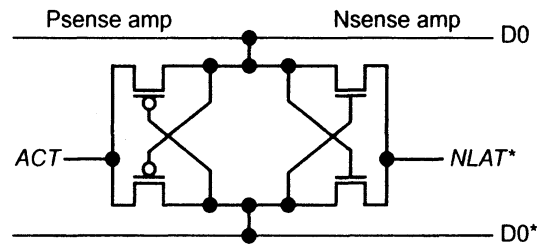


Figure 2.18 Basic sense amplifier block.

The sense amplifiers are carefully designed to guarantee correct detection and amplification of the small signal voltage produced during cell access (less than 200mV) [5]. Matching of transistor V_{TH} , transconductance, and junction capacitance within close tolerances helps ensure reliable sense amplifier operation. Ultimately, the layout dictates the overall balance and performance of the sense amplifier block. As a result, a tremendous amount of time is spent ensuring that the sense amplifier layout is optimum. Symmetry and exact duplication of elements are critical to a successful design. This includes balanced coupling to all sources of noise, such as *I/O* lines and latch signals (*NLAT** and *ACT*). Balance is especially critical for layout residing inside the isolation transistors. And because the sense node capacitance is very low, it is more sensitive to noise and circuit imbalances.

While the majority of DRAM designs latch the digitlines to V_{CC} and ground, a certain number of designs choose to reduce these levels. Various technical papers report improved Refresh times and lower power dissipation through reductions in latch voltages [20][21]. At first, this appears contradictory: writing a smaller charge into the memory cell would require refreshing the cell more often. However, by maintaining lower *drain-to-source voltages* (V_{DS}) and negative *gate-to-source voltages* (V_{GS}) across non-accessed mbit transistors, substantially lower subthreshold leakage and longer Refresh times can be realized, despite the smaller stored charge. An important concern that we are not mentioning here is the leakage resulting from defects in the silicon crystal structure. These defects can result in an increase in the drain/substrate diode saturation current and can practically limit the Refresh times.

Most designs that implement reduced latch voltages generally raise the Nsense amplifier latch voltage without lowering the Psense amplifier latch voltage. Designated as boosted sense ground designs, they write data into each mbit using full V_{CC} for a logic one and boosted ground for a logic zero. The sense ground level is generally a few hundred millivolts above true ground. In standard DRAMs, which drive digitlines fully to ground, the V_{GS}

of nonaccessed mbits becomes zero when the digitlines are latched. This results in high subthreshold leakage for a stored one level because full V_{CC} exists across the mbit transistor while the V_{GS} is held to zero. Stored zero levels do not suffer from prolonged subthreshold leakage: any amount of cell leakage produces a negative V_{GS} for the transistor. The net effect is that a stored one level leaks away much faster than a stored zero level. One's level retention, therefore, establishes the maximum Refresh period for most DRAM designs. Boosted sense ground extends Refresh by reducing subthreshold leakage for stored ones. This is accomplished by guaranteeing negative gate-to-source bias on nonaccessed mbit transistors. The benefit of extended Refresh from these designs is somewhat diminished, though, by the added complexity of generating boosted ground levels and the problem of digitlines that no longer equilibrate at $V_{CC}/2$ V.

2.2.5 Rate of Activation

The rate at which the sense amplifiers are activated has been the subject of some debate. A variety of designs use multistage circuits to control the rate at which $NLAT^*$ fires. Especially prevalent with boosted sense ground designs are two-stage circuits that initially drive $NLAT^*$ quickly toward true ground to speed sensing and then bring $NLAT^*$ to the boosted ground level to reduce cell leakage. An alternative to this approach, again using two-stage drivers, drives $NLAT^*$ to ground, slowly at first to limit current and digitline disturbances. This is followed by a second phase in which $NLAT^*$ is driven more strongly toward ground to complete the Sensing operation. This phase usually occurs in conjunction with ACT activation. Although these two designs have contrary operation, each meets specific performance objectives: trading off noise and speed.

2.2.6 Configurations

Figure 2.19 shows a sense amplifier block commonly used in double- or triple-metal designs. It features two Psense amplifiers outside the isolation transistors, a pair of $EQ/bias$ (EQb) devices, a single Nsense amplifier, and a single I/O transistor for each digitline. Because only half of the sense amplifiers for each array are on one side, this design is quarter pitch, as are the designs in Figures 2.20 and 2.21. Placement of the Psense amplifiers outside the isolation devices is necessary because a full one level (V_{CC}) cannot pass through unless the gate terminal of the ISO transistors is driven above V_{CC} . $EQ/bias$ transistors are placed outside of the ISO devices to permit continued equilibration of digitlines in arrays that are isolated. The I/O transistor gate terminals are connected to a common $CSEL$ signal for four adjacent digitlines. Each of the four I/O transistors is tied to a separate I/O

bus. This sense amplifier, though simple to implement, is somewhat larger than other designs due to the presence of two Psense amplifiers.

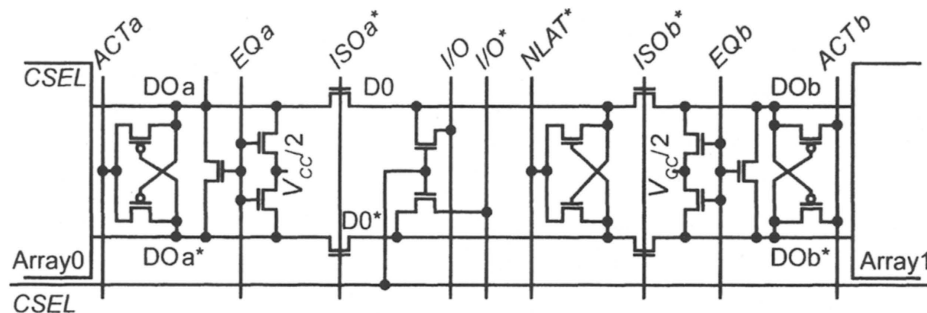


Figure 2.19 Standard sense amplifier block.

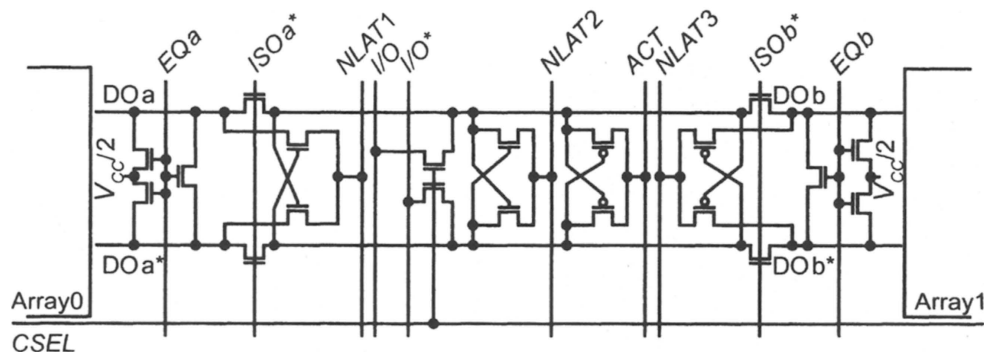


Figure 2.20 Complex sense amplifier block.

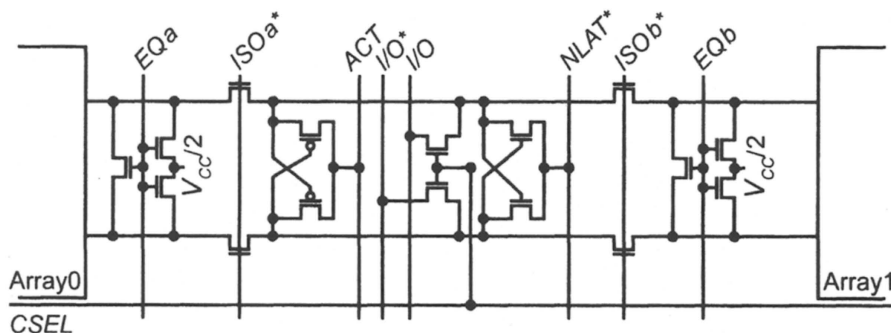


Figure 2.21 Reduced sense amplifier block.

Figure 2.20 shows a second, more complicated style of sense amplifier block. This design employs a single Psense amplifier and three sets of Nsense amplifiers. Because the Psense amplifier is within the isolation transistors, the isolation devices must be the NMOS depletion type, the PMOS enhancement type, or the NMOS enhancement type with boosted gate drive

to permit writing a full logic one into the array mbits. The triple Nsense amplifier is suggestive of PMOS isolation transistors; it prevents full zero levels to be written unless the Nsense amplifiers are placed adjacent to the arrays. In this more complicated style of sense amplifier block, using three Nsense amplifiers guarantees faster sensing and higher stability than a similar design using only two Nsense amplifiers. The inside Nsense amplifier is fired before the outside Nsense amplifiers. However, this design will not yield a minimum layout, an objective that must be traded off against performance needs.

The sense amplifier block of Figure 2.21 can be considered a reduced configuration. This design has only one Nsense-amp and one Psense-amp, both of which are placed within the isolation transistors. To write full logic levels, either the isolation transistors must be depletion mode devices or the gate voltage must be boosted above V_{CC} by at least one V_{TH} . This design still uses a pair of EQ /bias circuits to maintain equilibration on isolated arrays.

Only a handful of designs operates with a single EQ /bias circuit inside the isolation devices, as shown in Figure 2.22. Historically, DRAM engineers tended to shy away from designs that permitted digitlines to float for extended periods of time. However, as of this writing, at least three manufacturers in volume production have designs using this scheme.

A sense amplifier design for single-metal DRAMs is shown in Figure 2.23. Prevalent on 1-Meg and 4-Meg designs, single-metal processes conceded to multi-metal processes at the 16-Meg generation. Unlike the sense amplifiers shown in Figures 2.19, 2.20, 2.21, and 2.22, single-metal sense amps are laid out at half pitch: one amplifier for every two array digitlines. This type of layout is extremely difficult and places tight constraints on process design margins. With the loss of Metal2, the column select signals are not brought across the memory arrays. Generating column select signals locally for each set of I/O transistors requires a full column decode block.

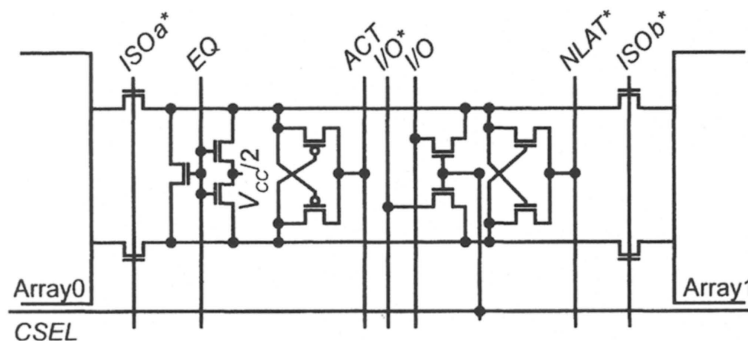


Figure 2.22 Minimum sense amplifier block.

As shown in Figure 2.23, the Nsense-amp and Psense-amps are placed on separate ends of the arrays. Sharing sense amps between arrays is especially beneficial for single-metal designs. As illustrated in Figure 2.23, two Psense-amps share a single Nsense-amp. In this case, with the *I/O* devices on only one end, the right Psense-amp is activated only when the right array is accessed. Conversely, the left Psense-amp is always activated, regardless of the accessed array, because all Read and Write data must pass through the left Psense-amp to get to the *I/O* devices.

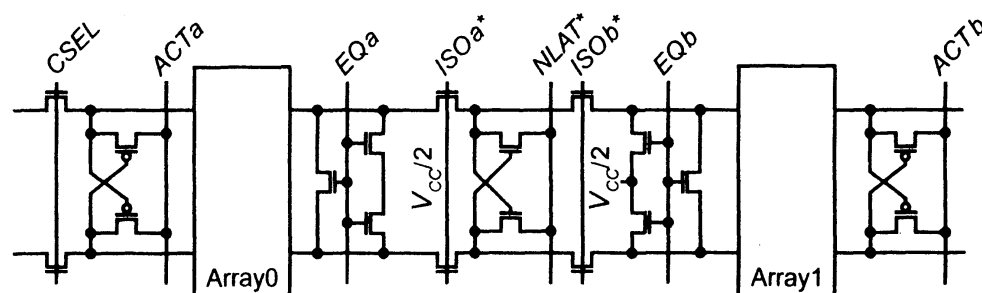


Figure 2.23 Single-metal sense amplifier block.

2.2.7 Operation

A set of signal waveforms is illustrated in Figure 2.24 for the sense amplifier of Figure 2.19. These waveforms depict a Read-Modify-Write cycle (Late Write) in which the cell data is first read out and then new data is written back. In this example, a one level is read out of the cell, as indicated by *D0* rising above *D0* during cell access. A one level is always $+V_{CC}/2$ in the mbit cell, regardless of whether it is connected to a true or complement digitline. The correlation between mbit cell data and the data appearing at the DRAM's data terminal (*DQ*) is a function of the data topology and the presence of data scrambling. Data or topo scrambling is implemented at the circuit level: it ensures that the mbit data state and the *DQ* logic level are in agreement. An mbit one level ($+V_{CC}/2$) corresponds to a logic one at the *DQ*, and an mbit zero level ($-V_{CC}/2$) corresponds to a logic zero at the *DQ* terminal.

Writing specific data patterns into the memory arrays is important to DRAM testing. Each type of data pattern identifies the weaknesses or sensitivities of each cell to the data in surrounding cells. These patterns include solids, row stripes, column stripes, diagonals, checkerboards, and a variety of moving patterns. Test equipment must be programmed with the data topology of each type of DRAM to correctly write each pattern. Often the tester itself guarantees that the pattern is correctly written into the arrays, unscrambling the complicated data and address topology as necessary to write a specific pattern. On some newer DRAM designs, part of this task is

implemented on the DRAM itself, in the form of a topo scrambler, such that the mbit data state matches the DQ logic level. This implementation somewhat simplifies tester programming.

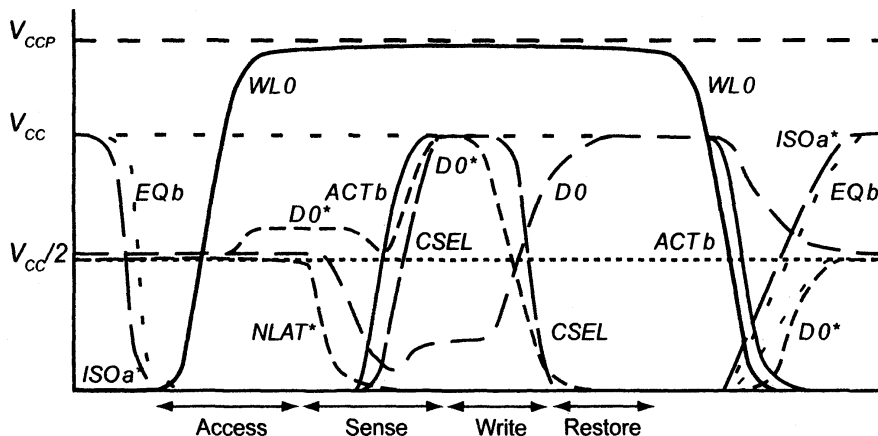


Figure 2.24 Waveforms for the Read-Modify-Write cycle.

Returning to Figure 2.24, we see that a wordline is fired in Array1. Prior to this, $ISOa^*$ will go LOW to isolate Array0, and EQb will go LOW to disable the $EQ/bias$ transistors connected to Array1. The wordline then fires HIGH, accessing an mbit, which dumps charge onto $D0^*$. $NLAT^*$, which is initially at $V_{CC}/2$, drives LOW to begin the Sensing operation and pulls $D0$ toward ground. Then, ACT fires, moving from ground to V_{CC} , activating the Psense amplifier and driving $D0^*$ toward V_{CC} . After separation has commenced, $CSEL0$ rises to V_{CC} , turning ON the I/O transistors so that the cell data can be read by peripheral circuits. The I/O lines are biased at a voltage close to V_{CC} , which causes $D1$ to rise while the column is active. After the Read is complete, Write drivers in the periphery turn ON and drive the I/O lines to opposite data states (in our example).

This new data propagates through the I/O devices and writes over the existing data stored in the sense amplifiers. Once the sense amplifiers latch the new data, the Write drivers and the I/O devices can be shut down, allowing the sense amplifiers to finish restoring the digitlines to full levels. Following this restoration, the wordline transitions LOW to shut OFF the mbit transistor. Finally, EQb and $ISOa^*$ fire HIGH to equilibrate the digitlines back to $V_{CC}/2$ and reconnect Array0 to the sense amplifiers. The timing for each event of Figure 2.24 depends on circuit design, transistor sizes, layout, device performance, parasitics, and temperature. While timing for each event must be minimized to achieve optimum DRAM performance, it cannot be pushed so far as to eliminate all timing margins. Margins are necessary to ensure proper device operation over the expected range of process variations and the wide range of operating conditions.

Again, there is not one set of timing waveforms that covers all design options. The sense amps of Figures 2.19–2.23 all require slightly different signals and timings. Various designs actually fire the Psense amplifier prior to or coincident with the Nsense amplifier. This obviously places greater constraints on the Psense amplifier design and layout, but these constraints are balanced by potential performance benefits. Similarly, the sequence of events as well as the voltages for each signal can vary. There are almost as many designs for sense amplifier blocks as there are DRAM design engineers. Each design reflects various influences, preconceptions, technologies, and levels of understanding. The bottom line is to maximize yield and performance and minimize everything else.

2.3 ROW DECODER ELEMENTS

Row decode circuits are similar to sense amplifier circuits in that they pitch up to mbit arrays and have a variety of implementations. A row decode block is comprised of two basic elements: a wordline driver and an address decoder tree. There are three basic configurations for wordline driver circuits: the NOR driver, the inverter (CMOS) driver, and the bootstrap driver. In addition, the drivers and associated decode trees can be configured either as local row decodes for each array section or as global row decodes that drive a multitude of array sections.

Global row decodes connect to multiple arrays through metal wordline straps. The straps are stitched to the polysilicon wordlines at specific intervals dictated by the polysilicon resistance and the desired RC wordline time constant. Some processes that strap wordlines with metal do not silicide the polysilicon, although doing so would reduce the number of stitch regions required. Strapping wordlines and using global row decoders obviously reduces die size [22], very dramatically in some cases. The disadvantage of strapping is that it requires an additional metal layer at minimum array pitch. This puts a tremendous burden on process technologists in which three conductors are at minimum pitch: wordlines, digitlines, and wordline straps.

Local row decoders, on the other hand, require additional die size rather than metal straps. It is highly advantageous to reduce the polysilicon resistance in order to stretch the wordline length and lower the number of row decodes needed. This is commonly achieved with silicided polysilicon processes. On large DRAMs, such as the 1Gb, the area penalty can be prohibitive, making low-resistance wordlines all the more necessary.

2.3.1 Bootstrap Wordline Driver

The bootstrap wordline driver shown in Figure 2.25 is built exclusively from NMOS transistors [22], resulting in the smallest layout of the three types of driver circuits. The absence of PMOS transistors eliminates large Nwell regions from the layout. As the name denotes, this driver relies on bootstrapping principles to bias the output transistor's gate terminal. This bias voltage must be high enough to allow the NMOS transistor to drive the wordline to the boosted wordline voltage V_{CCP} .

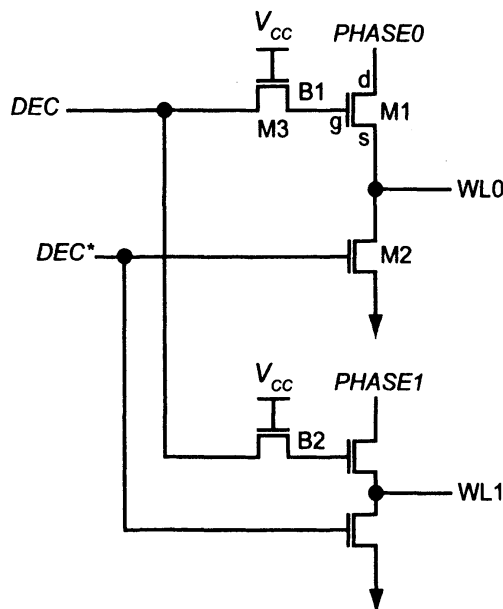


Figure 2.25 Bootstrap wordline driver.

Operation of the bootstrap driver is described with help from Figure 2.26. Initially, the driver is OFF with the wordline and *PHASE* terminals at ground. The wordline is held at ground by transistor M2 because the decoder output signal *DEC** is at V_{CC} . The gate of M3, the pass transistor, is fixed at V_{CC} . The signals *DEC* and *DEC** are fed from a decode circuit that will be discussed later. As a complement pair, *DEC* and *DEC** represent the first of two terms necessary to decode the correct wordline. *PHASE0*, which is also fed from a decode circuit, represents the second term. If *DEC* rises to V_{CC} and *DEC** drops to ground, as determined by the decoder, the boot node labeled B1 will rise to $V_{CC} - V_{TN}$ V, and M2 will turn OFF. The wordline continues to be held to ground by M1 because *PHASE0* is still grounded. After B1 rises to $V_{CC} - V_{TN}$, the *PHASE0* signal fires to the boosted wordline voltage V_{CCP} . Because of the gate-to-drain and gate-to-source capacitance of M1, the gate of M1 boots to an elevated voltage, V_{boot} . This voltage is determined by the parasitic capacitance of node B1, C_{GS1} , C_{GD1} , V_{CCP} , and the initial voltage at B1, $V_{CC} - V_{TN}$. Accordingly,

$$V_{boot} \equiv \frac{(V_{CCP} \cdot C_{GD1})}{(C_{GS1} + C_{GD1} + C_{B1})} + (V_{CC} - V_{TN}) \quad (2.3)$$

In conjunction with the wordline voltage rising from ground to V_{CCP} , the gate-to-source capacitance of M1 provides a secondary boost to the boot node. The secondary boost helps to ensure that the boot voltage is adequate to drive the wordline to a full V_{CCP} level.

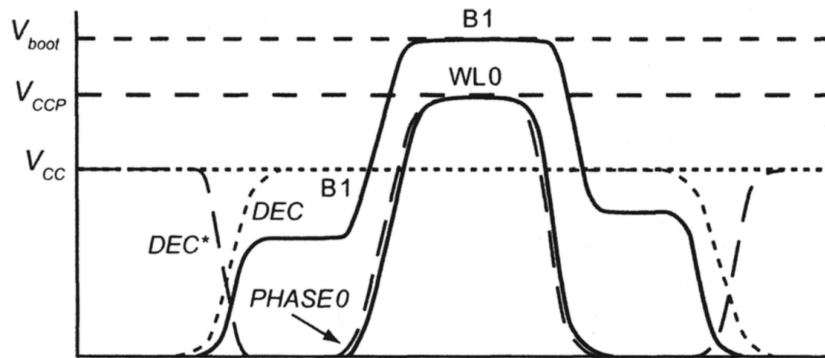


Figure 2.26 Bootstrap operation waveforms.

The layout of the boot node is very important to the bootstrap wordline driver. First, the parasitic capacitance of node B1, which includes routing, junction, and overlap components, must be minimized to achieve maximum boot efficiency. Second, charge leakage from the boot node must be minimized to ensure adequate V_{GS} for transistor M1 such that the wordline remains at V_{CCP} for the maximum RAS^* LOW period. Low leakage is often achieved by minimizing the source area for M3 or by using donut gate structures that surround the source area, as illustrated in Figure 2.27.

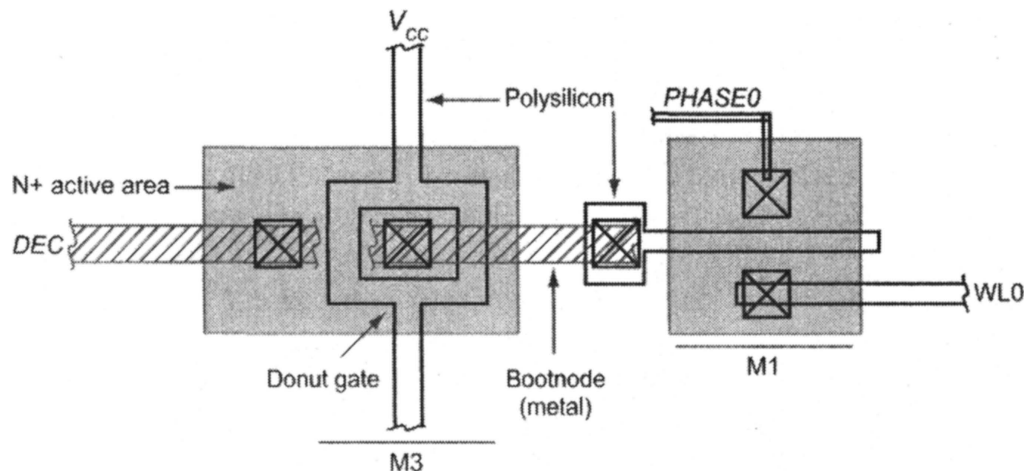


Figure 2.27 Donut gate structure layout.

The bootstrap driver is turned OFF by first driving the *PHASE0* signal to ground. M1 remains ON because node *B1* cannot drop below $V_{CC} - V_{TH}$; M1 substantially discharges the wordline toward ground. This is followed by the address decoder turning OFF, bringing *DEC* to ground and *DEC** to V_{CC} . With *DEC** at V_{CC} , transistor M2 turns ON and fully clamps the wordline to ground. A voltage level translator is required for the *PHASE* signal because it operates between ground and the boosted voltage V_{CCP} . For a global row decode configuration, this requirement is not much of a burden. For a local row decode configuration, however, the requirement for level translators can be very troublesome. Generally, these translators are placed either in the array gap cells at the intersection of the sense amplifier blocks and row decode blocks or distributed throughout the row decode block itself. The translators require both PMOS and NMOS transistors and must be capable of driving large capacitive loads. Layout of the translators is exceedingly difficult, especially because the overall layout needs to be as small as possible.

2.3.2 NOR Driver

The second type of wordline driver is similar to the bootstrap driver in that two decode terms drive the output transistor from separate terminals. The NOR driver, as shown in Figure 2.28, uses a PMOS transistor for M1 and does not rely on bootstrapping to derive the gate bias. Rather, the gate is driven by a voltage translator that converts *DEC* from V_{CC} to V_{CCP} voltage levels. This conversion is necessary to ensure that M1 remains OFF for unselected wordlines as the *PHASE* signal, which is shared by multiple drivers, is driven to V_{CCP} .

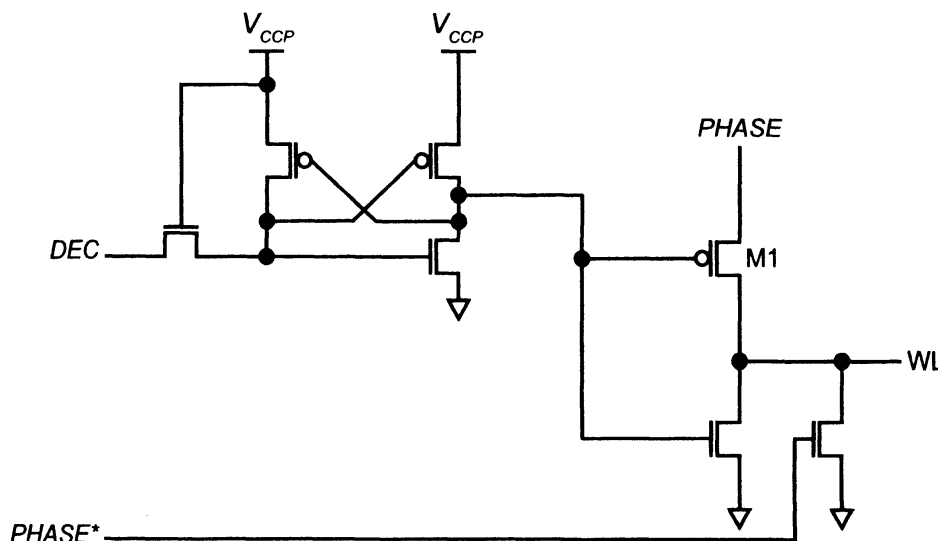


Figure 2.28 NOR driver.

To fire a specific wordline, *DEC* must be HIGH and the appropriate *PHASE* must fire HIGH. Generally, there are four to eight *PHASE* signals per row decoder block. The NOR driver requires one level translator for each *PHASE* and *DEC* signal. By comparison, the bootstrap driver only requires level translators for the *PHASE* signal.

2.3.3 CMOS Driver

The final wordline driver configuration is shown in Figure 2.29. In general, it lacks a specific name: it is sometimes referred to as a CMOS inverter driver or a CMOS driver. Unlike the first two drivers, the output transistor M1 has its source terminal permanently connected to V_{CCP} . This driver, therefore, features a voltage translator for each and every wordline. Both decode terms *DEC* and *PHASE** combine to drive the output stage through the translator. The advantage of this driver, other than simple operation, is low power consumption. Power is conserved because the translators drive only the small capacitance associated with a single driver. The *PHASE* translators of both the bootstrap and NOR drivers must charge considerable junction capacitance. The disadvantages of the CMOS driver are layout complexity and standby leakage current. Standby leakage current is a product of V_{CCP} voltage applied to M1 and its junction and subthreshold leakage currents. For a large DRAM with high numbers of wordline drivers, this leakage current can easily exceed the entire standby current budget unless great care is exercised in designing output transistor M1.

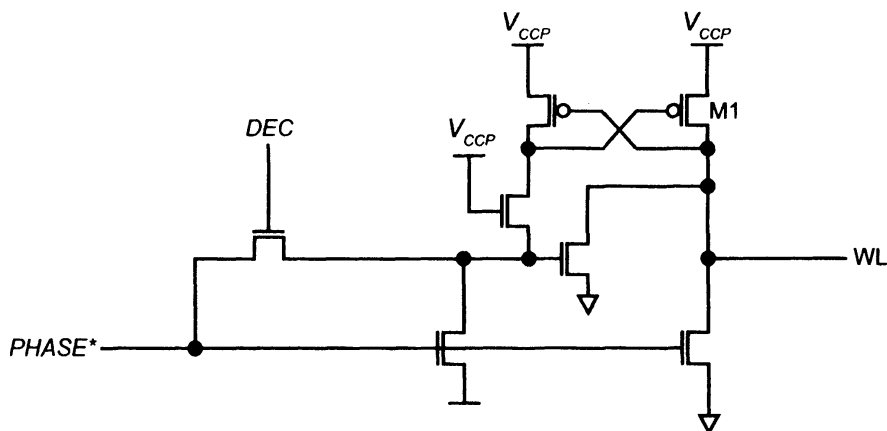


Figure 2.29 CMOS driver.

2.3.4 Address Decode Tree

With the wordline driver circuits behind us, we can turn our attention to the address decoder tree. There is no big secret to address decoding in the row decoder network. Just about any type of logic suffices: static, dynamic,

pass gate, or a combination thereof. With any type of logic, however, the primary objectives in decoder design are to maximize speed and minimize die area. Because a great variety of methods have been used to implement row address decoder trees, it is next to impossible to cover them all. Instead, we will give an insight into the possibilities by discussing a few of them.

Regardless of the type of logic with which a row decoder is implemented, the layout must completely reside beneath the row address signal lines to constitute an efficient, minimized design. In other words, the metal address tracks dictate the die area available for the decoder. Any additional tracks necessary to complete the design constitute wasted silicon. For DRAM designs requiring global row decode schemes, the penalty for inefficient design may be insignificant; however, for distributed local row decode schemes, the die area penalty may be significant. As with mbits and sense amplifiers, time spent optimizing row decode circuits is time well spent.

2.3.5 Static Tree

The most obvious form of address decode tree uses static CMOS logic. As shown in Figure 2.30, a simple tree can be designed using two-input NAND gates. While easy to design schematically, static logic address trees are not popular. They waste silicon and are very difficult to lay out efficiently. Static logic requires two transistors for each address term, one NMOS and one PMOS, which can be significant for many address terms. Furthermore, static gates must be cascaded to accumulate address terms, adding gate delays at each level. For these and other reasons, static logic gates are not used in row decode address trees in today's state-of-the-art DRAMs.

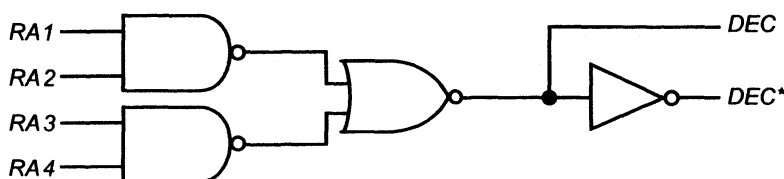


Figure 2.30 Static decode tree.

2.3.6 P&E Tree

The second type of address tree uses dynamic logic, the most prevalent being *precharge and evaluate* (P&E) logic. Used by the majority of DRAM manufacturers, P&E address trees come in a variety of configurations, although the differences between them can be subtle. Figure 2.31 shows a simplified schematic for one version of a P&E address tree designed for use with bootstrapped wordline drivers. P&E address tree circuits feature one or

more PMOS *PRECHARGE* transistors and a cascade of NMOS *ENABLE* transistors M2–M4. This P&E design uses half of the transistors required by the static address tree of Figure 2.30. As a result, the layout of the P&E tree is much smaller than that of the static tree and fits more easily under the address lines. The *PRE* transistor is usually driven by a *PRECHARGE** signal under the control of the *RAS** chain logic. *PRECHARGE** and transistor M1 ensure that *DEC** is precharged HIGH, disabling the wordline driver and preparing the tree for row address activation.

M7 is a weak PMOS transistor driven by the *DEC* inverter (M5 and M6). Together, M7 and the inverter form a latch to ensure that *DEC** remains HIGH for all decoders that are not selected by the row addresses. At the beginning of any *RAS** cycle, *PRECHARGE** is LOW and the row addresses are all disabled (LOW). After *RAS** falls, *PRECHARGE** transitions HIGH to turn OFF M1; then the row addresses are enabled. If *RA1*–*RA3* all go HIGH, then M2–M4 turn ON, overpowering M7 and driving *DEC** to ground and subsequently *DEC* to V_{CC} . The output of this tree segment normally drives four bootstrapped wordline drivers, each connected to a separate *PHASE* signal. Therefore, for an array with 256 wordlines, there will be 64 such decode trees.

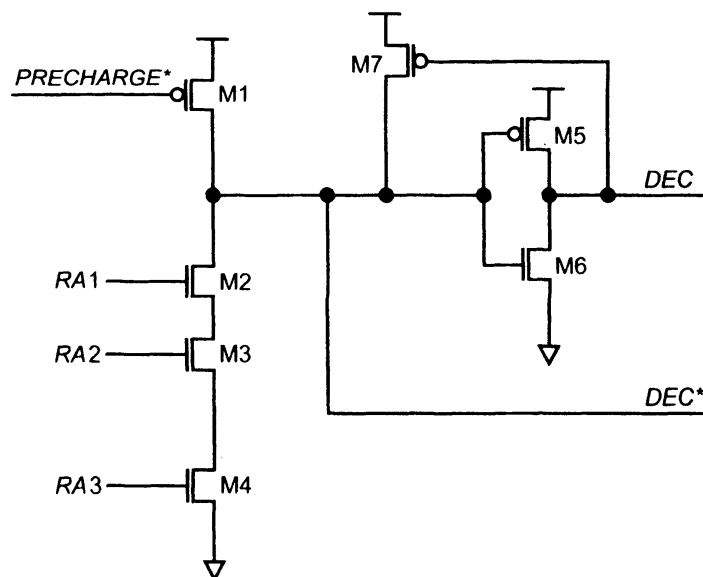


Figure 2.31 P&E decode tree.

2.3.7 Predecoding

The row address lines shown as *RA1*–*RA3* can be either true and complement or predecoded. Predecoded address lines are formed by logically combining (AND) addresses as shown in Table 2.1.

The advantages of using predecoded addresses include lower power (fewer signals make transitions during address changes) and higher efficiency (only three transistors are necessary to decode six addresses for the circuit of Figure 2.31). Predecoding is especially beneficial in redundancy circuits. In fact, predecoded addresses are used throughout most DRAM designs today.

Table 2.1 Predecoded address truth table.

RA0	RA1	PR01<n>	PR01<0>	PR01<1>	PR01<2>	PR01<3>
0	0	0	1	0	0	0
1	0	1	0	1	0	0
0	1	2	0	0	1	0
1	1	3	0	0	0	1

2.3.8 Pass Transistor Tree

The final type of address tree to be examined, shown in Figure 2.32, uses pass transistor logic. Pass transistor address trees are similar to P&E trees in numerous ways. Both designs use PMOS *PRECHARGE* transistors and NMOS address *ENABLE* transistors. Unlike P&E logic, however, the NMOS cascade does not terminate at ground. Rather, the cascade of M2–M4 goes to a *PHASE** signal, which is HIGH during *PRECHARGE* and LOW during *EVALUATE*. The address signals operate the same as in the P&E tree: HIGH to select and LOW to deselect. The pass transistor tree is shown integrated into a CMOS wordline driver. This is necessary because the pass transistor tree and the CMOS wordline driver are generally used together and their operation is complementary. The cross-coupled PMOS transistors of the CMOS level translator provide a latch necessary to keep the final interstage node biased at V_{CC} . Again the latch has a weak pull-up, which is easily overpowered by the cascaded NMOS *ENABLE* transistors. A pass transistor address tree is not used with bootstrapped wordline drivers because the *PHASE* signal feeds into the address tree logic rather than into the driver, as required by the bootstrap driver.

2.4 DISCUSSION

We have briefly examined the basic elements required in DRAM row decoder blocks. Numerous variations are possible. No single design is best for all applications. As with sense amplifiers, design depends on technology and performance and cost trade-offs.

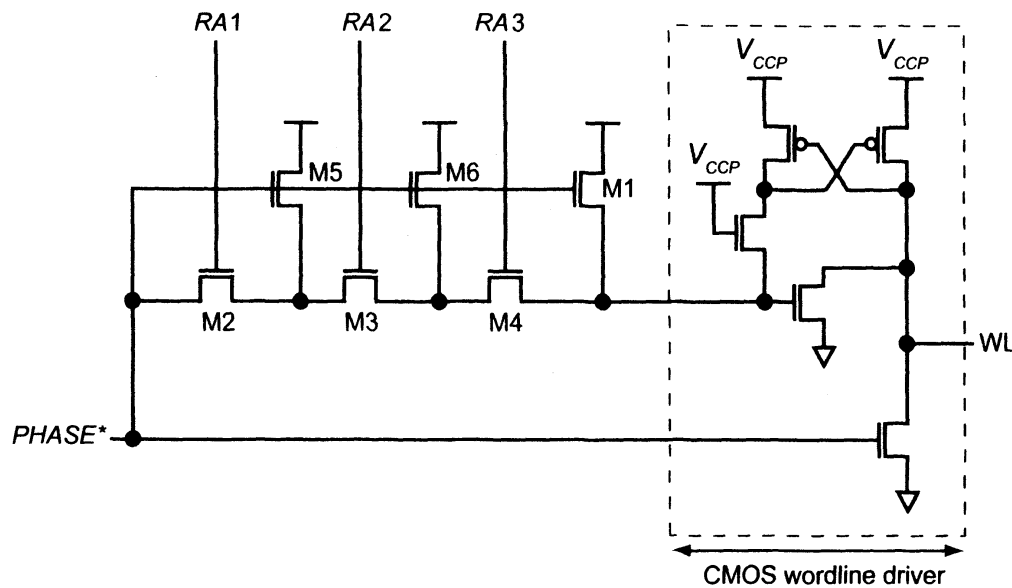


Figure 2.32 Pass transistor decode tree.

REFERENCES

- [1] K. Itoh, "Trends in megabit DRAM circuit design," *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 778–791, June 1990.
- [2] D. Takashima, S. Watanabe, H. Nakano, Y. Oowaki, and K. Ohuchi, "Open/folded bit-line arrangement for ultra-high-density DRAMs," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 539–542, April 1994.
- [3] Hideto Hidaka, Yoshio Matsuda, and Kazuyasu Fujishima, "A divided/shared bit-line sensing scheme for ULSI DRAM cores," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 473–477, April 1991.
- [4] M. Aoki, Y. Nakagome, M. Horiguchi, H. Tanaka, S. Ikenaga, J. Etoh, Y. Kawamoto, S. Kimura, E. Takeda, H. Sunami, and K. Itoh, "A 60-ns 16-Mbit CMOS DRAM with a transposed data-line structure," *IEEE Journal of Solid-State Circuits*, vol. 23, pp. 1113–1119, October 1988.
- [5] R. Kraus and K. Hoffmann, "Optimized sensing scheme of DRAMs," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 895–899, August 1989.
- [6] T. Yoshihara, H. Hidaka, Y. Matsuda, and K. Fujishima, "A twisted bitline technique for multi-Mb DRAMs," in *1988 IEEE ISSCC Digest of Technical Papers*, pp. 238–239.
- [7] Yukihiro Oowaki, Kenji Tsuchida, Yohji Watanabe, Daisaburo Takashima, Masako Ohta, Hiroaki Nakano, Shigeyoshi Watanabe, Akihiro Nitayama, Fumio Horiguchi, Kazunori Ohuchi, and Fujio Masuoka, "A 33-ns 64Mb DRAM," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 1498–1505, November 1991.

- [8] M. Inoue, H. Kotani, T. Yamada, H. Yamauchi, A. Fujiwara, J. Matsushima, H. Akamatsu, M. Fukumoto, M. Kubota, I. Nakao, N. Aoi, G. Fuse, S. Ogawa, S. Odanaka, A. Ueno, and Y. Yamamoto, "A 16Mb DRAM with an open bit-line architecture," in *1988 IEEE ISSCC Digest of Technical Papers*, pp. 246–247.
- [9] Y. Kubota, Y. Iwase, K. Iguchi, J. Takagi, T. Watanabe, and K. Sakiyama, "Alternatively-activated open bitline technique for high density DRAM's," *IEICE Trans. Electron.*, vol. E75-C, pp. 1259–1266, October 1992.
- [10] T. Hamada, N. Tanabe, H. Watanabe, K. Takeuchi, N. Kasai, H. Hada, K. Shibahara, K. Tokashiki, K. Nakajima, S. Hirasawa, E. Ikawa, T. Saeki, E. Kakehashi, S. Ohya, and T. Kunio, "A split-level diagonal bit-line (SLDB) stacked capacitor cell for 256Mb DRAMs," *1992 IEDM Technical Digest*, pp. 799–802.
- [11] Toshinori Morihara, Yoshikazu Ohno, Takahisa Eimori, Toshiharu Katayama, Shinichi Satoh, Tadashi Nishimura, and Hirokazu Miyoshi, "Disk-shaped stacked capacitor cell for 256Mb dynamic random-access memory," *Japan Journal of Applied Physics*, vol. 33, Part 1, pp. 4570–4575, August 1994.
- [12] J. H. Ahn, Y. W. Park, J. H. Shin, S. T. Kim, S. P. Shim, S. W. Nam, W. M. Park, H. B. Shin, C. S. Choi, K. T. Kim, D. Chin, O. H. Kwon, and C. G. Hwang, "Micro villus patterning (MVP) technology for 256Mb DRAM stack cell," in *1992 Symposium on VLSI Tech. Digest of Technical Papers*, pp. 12–13.
- [13] Kazuhiko Sagara, Tokuo Kure, Shoji Shukuri, Jiro Yugami, Norio Hasegawa, Hidekazu Goto, and Hisaomi Yamashita, "Recessed memory array technology for a double cylindrical stacked capacitor cell of 256M DRAM," *IEICE Trans. Electron.*, vol. E75-C, pp. 1313–1322, November 1992.
- [14] S. Ohya. 1994. Semiconductor memory device having stacked-type capacitor of large capacitance. US Patent 5,298,775, March 29, 1994.
- [15] M. Sakao, N. Kasai, T. Ishijima, E. Ikawa, H. Watanabe, K. Terada, and T. Kikkawa, "A capacitor-over-bit-line (COB) cell with hemispherical-grain storage node for 64Mb DRAMs," in *1990 IEDM Technical Digest*, pp. 655–658.
- [16] G. Bronner, H. Aochi, M. Gall, J. Gambino, S. Gernhardt, E. Hammerl, H. Ho, J. Iba, H. Ishiuchi, M. Jaso, R. Kleinhenz, T. Mii, M. Narita, L. Nesbit, W. Neumueller, A. Nitayama, T. Ohiwa, S. Parke, J. Ryan, T. Sato, H. Takato, and S. Yoshikawa, "A fully planarized 0.25 μ m CMOS technology for 256Mbit DRAM and beyond," in *1995 Symposium on VLSI Tech. Digest of Technical Papers*, pp. 15–16.
- [17] N. C.-C. Lu and H. H. Chao, "Half- V_{DD} / bit-line sensing scheme in CMOS DRAMs," in *IEEE Journal of Solid-State Circuits*, vol. SC19, p. 451, August 1984.
- [18] E. Adler; J. K. DeBrosse; S. F. Geissler; S. J. Holmes; M. D. Jaffe; J. B. Johnson; C. W. Koburger, III; J. B. Lasky; B. Lloyd; G. L. Miles;

- J. S. Nakos; W. P. Noble, Jr.; S. H. Voldman; M. Armacost; and R. Ferguson; "The evolution of IBM CMOS DRAM technology," *IBM Journal of Research and Development*, vol. 39, pp. 167–188, March 1995.
- [19] R. Kraus, "Analysis and reduction of sense-amplifier offset," in *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1028–1033, August 1989.
- [20] M. Asakura, T. Ohishi, M. Tsukude, S. Tomishima, H. Hidaka, K. Arimoto, K. Fujishima, T. Eimori, Y. Ohno, T. Nishimura, M. Yasunaga, T. Kondoh, S. I. Satoh, T. Yoshihara, and K. Demizu, "A 34ns 256Mb DRAM with boosted sense-ground scheme," in *1994 IEEE ISSCC Digest of Technical Papers*, pp. 140–141.
- [21] T. Ooishi, K. Hamade, M. Asakura, K. Yasuda, H. Hidaka, H. Miyamoto, and H. Ozaki, "An automatic temperature compensation of internal sense ground for sub-quarter micron DRAMs," in *1994 Symposium on VLSI Circuits Digest of Technical Papers*, pp. 77–78.
- [22] K. Noda, T. Saeki, A. Tsujimoto, T. Murotani, and K. Koyama, "A boosted dual word-line decoding scheme for 256 Mb DRAMs," in *1992 Symposium on VLSI Circuits Digest of Technical Papers*, pp. 112–113.