"...continually revised and updated to reflect the ever-changing PC landscape. A resource from my PC repair days, Scott Mueller's tome is one of the books that has stayed on my bookshelf throughout the years."

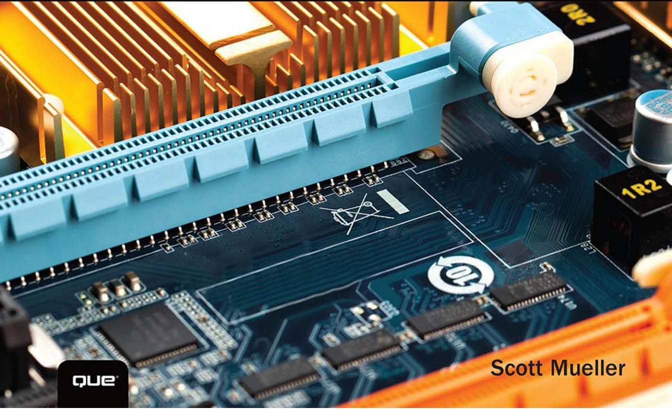
—**Adam Overa, Managing Editor, *Tom's Hardware***

DVD-ROM
Includes
3 hours of new
how-to video

Scott Mueller's

UPGRADING AND REPAIRING PCs

22nd EDITION



QUE

Scott Mueller

FREE SAMPLE CHAPTER

SHARE WITH OTHERS

UPGRADING AND REPAIRING PCs

22nd Edition

Scott M. Mueller

que®

800 East 96th Street,
Indianapolis, Indiana 46240 USA

Contents at a Glance

Upgrading and Repairing PCs, 22nd Edition

ISBN-13: 978-0-7897-5610-7
ISBN-10: 0-7897-5610-2

Library of Congress Control Number: 2015938404

Printed in the United States of America

First Printing: June 2015

Trademarks

Warning and Disclaimer

Special Sales

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact international@pearsoned.com.

Editor-in-Chief
Greg Wiegand

Executive Editor
Rick Kughen

Development Editor
Todd Brakke

Managing Editor
Sandra Schroeder

Project Editor
Mandie Frank

Copy Editor
Megan Wade-Taxter

Indexer
Erika Millen

Proofreader
Debbie Williams

Technical Editor
Karen Weinstein

Editorial Assistant
Kristen Watterson

Media Producer
Dan Scherf

Designer
Mark Shirar

Compositor
Nonie Ratcliff

Contents

Accessing the Online Media Included with this Book

Don't forget about the free bonus content available online! You'll find a cache of helpful material to go along with this book, including 2 hours of video, newly recorded for the 22nd edition. You'll also find complete PDF copies of the 19th and 20th editions, as well as many pages of valuable reference material that is particularly useful for those maintaining legacy equipment.

Register this eBook to unlock the companion files that are included in the Print edition DVD. Follow the steps below:

1. Go to quepublishng.com/register and log in or create a new account.
2. Enter the ISBN: 9780134057743
3. Enter the following code when prompted: URPCSDVD22E
4. Click on the "Access Bonus Content" link in the Registered Products section of your account page, to be taken to the page where your content is available.

The video files will play in your browser. Click the links to the 19th and 20th edition PDFs, and other materials to view them, or right-click and choose to save the file to your computer.

About the Authors

Scott Mueller is the president of Mueller Technical Research (MTR), an international research and corporate training firm. Since 1982, MTR has produced the industry's most in-depth, accurate, and effective seminars, books, articles, videos, and FAQs covering PC hardware and data recovery. MTR maintains a client list that includes Fortune 500 companies, U.S. and foreign governments, major software and hardware corporations, as well as PC enthusiasts and entrepreneurs. Scott's seminars have been presented to several thousands of PC support professionals throughout the world.

Scott personally teaches seminars nationwide covering all aspects of PC hardware (including troubleshooting, maintenance, repair, and upgrade), A+ Certification, and data recovery/forensics. He has a knack for making technical topics not only understandable, but entertaining; his classes are never boring! If you have ten or more people to train, Scott can design and present a custom seminar for your organization.

Although he has taught classes virtually nonstop since 1982, Scott is best known as the author of the longest-running, most popular, and most comprehensive PC hardware book in the world, *Upgrading and Repairing PCs*, which has become the core of an entire series of books, including *Upgrading and Repairing PCs*, *Upgrading and Repairing Laptops*, and *Upgrading and Repairing Windows*.

Scott's premiere work, *Upgrading and Repairing PCs*, has sold more than two million copies, making it by far the most popular and longest-running PC hardware book on the market today. Scott has been featured in *Forbes* magazine and has written several articles for *PC World* magazine, *Maximum PC* magazine, the Scott Mueller Forum, various computer and automotive newsletters, and the *Upgrading and Repairing PCs* website.

Contact MTR directly if you have a unique book, article, or video project in mind or if you want Scott to conduct a custom PC troubleshooting, repair, maintenance, upgrade, or data-recovery seminar tailored for your organization:

Mueller Technical Research

Web: www.muellertech.com
Email: info@muellertech.com
Forum: www.forum.scottmueller.com

Scott has a forum exclusively for his readers at www.forum.scottmueller.com. Anybody can view the forum, but posting is available only to registered members.

If you have suggestions or comments about the book or new book or article topics you would like to see covered, send them to info@muellertech.com.

Mark Edward Soper has helped users deal with problems with computers, digital cameras, and other personal tech devices for more than 30 years. He is the author of *PC and Gadgets Help Desk in a Book*, author of *Easy Windows 10*, as well as more than two dozen other books on Windows, digital imaging, networking, broadband Internet, CompTIA A+ Certification, and computer troubleshooting and upgrading. With this level of experience, combined with years of teaching technical classes, Mark is experienced at helping readers understand and use creative solutions to connectivity, configuration issues, data recovery, and other types of problems that can beset users of personal technology.

We Want to Hear from You!

As the reader of this book, you are our most important critic and commentator. We value your opinion and want to know what we're doing right, what we could do better, what areas you'd like to see us publish in, and any other words of wisdom you're willing to pass our way.

We welcome your comments. You can email or write to let us know what you did or didn't like about this book—as well as what we can do to make our books better.

Please note that we cannot help you with technical problems related to the topic of this book.

When you write, please be sure to include this book's title and author as well as your name and email address. We will carefully review your comments and share them with the author and editors who worked on the book.

Email: feedback@quepublishing.com

Mail: Que Publishing
ATTN: Reader Feedback
800 East 96th Street
Indianapolis, IN 46240 USA

Reader Services

Visit our website and register this book at quepublishing.com/register for convenient access to any updates, downloads, or errata that might be available for this book.

Introduction

Welcome to *Upgrading and Repairing PCs,* 22nd Edition. Since debuting as the first book of its kind on the market in 1988, no other book on PC hardware has matched the depth and quality of the information found in this tome. This edition continues *Upgrading and Repairing PCs'* role as not only the best-selling book of its type, but also the most comprehensive and complete PC hardware reference available. This book examines PCs in depth, outlines the differences among them, and presents options for configuring each system.

The 22nd edition of *Upgrading and Repairing PCs* provides you with the in-depth knowledge you need to work with the most recent systems and components and gives you an unexcelled resource for understanding older systems. As with previous editions, we worked to make this book keep pace with the rapid changes in the PC industry so that it continues to be the most accurate, complete, and in-depth book of its kind on the market today.

I wrote this book for all PC enthusiasts who want to know everything about their PCs: how they originated; how they've evolved; how to upgrade, troubleshoot, and repair them; and everything in between. This book covers the full gamut of PC-compatible systems, from the oldest 8-bit machines to the latest high-end, 64-bit multicore processors and systems. If you need to know everything about PC hardware from the original to the latest technology on the market today, this book and the accompanying information-packed disc is definitely for you.

Upgrading and Repairing PCs also doesn't ignore the less glamorous PC components. Every part of your PC plays a critical role in its stability and performance. Over the course of this book, you'll find out exactly why your motherboard's chipset might just be the most important part of your PC and what can go wrong when you settle for a run-of-the-mill power supply that can't get enough juice to that monster graphics card you just bought. You'll also find in-depth coverage of technologies such as the latest Intel Haswell processors (and the Haswell Refresh!), forthcoming Intel Broadwell desktop processors and AMD Kaveri APUs (and the latest about the graphics built into Intel and AMD processors); easy-to-use software to overclock your processor; the latest Intel and AMD chipsets; the latest BIOS settings for new processors; how helium-filled hard disks and shingled magnetic recording are leading to huge new hard disk capacities; the latest CFast and XQD flash memory cards; the newest high-performance graphics cards based on AMD and NVIDIA GPUs for the fastest 3D gaming; the latest developments in OpenGL, OpenCL, and DirectX 3D APIs; how AMD's Mantle improves 3D gaming; USB charge while sleeping support and the new USB 3.1 interface; SATAExpress, mSATA, and M.2 interfaces; the latest satellite Internet speed boosts; choosing the right keyswitches for high-performance gaming or data entry keyboards; new HomeGrid and G.hn home network standards; and more—it's all in here, right down to the guts-level analysis of every port on the back of or inside your system.

Book Objectives

Upgrading and Repairing PCs focuses on several objectives. The primary objective is to help you learn how to maintain, upgrade, and troubleshoot your PC system. To that end, *Upgrading and Repairing PCs* helps you fully understand the family of computers that has grown from the original IBM PC, including all PC-compatible systems. This book discusses all areas of system improvement, such as motherboards, processors, memory, and even case and power-supply improvements. It covers proper system and component care, specifies the most failure-prone items in various PC systems, and tells you how to locate and identify a failing component. You'll learn about powerful diagnostics hardware and software that help you determine the cause of a problem and know how to repair it.

As always, PCs are moving forward rapidly in power and capabilities. Processor performance increases with every new chip design. Upgrading and Repairing PCs helps you gain an understanding of all the processors used in PC-compatible computer systems.

This book covers the important differences between major system architectures, from the original Industry Standard Architecture (ISA) to the latest PCI Express interface standards. *Upgrading and Repairing PCs* covers each of these system architectures and their adapter boards to help you make decisions about which type of system you want to buy in the future and help you upgrade and troubleshoot such systems.

The amount of storage space available to modern PCs is increasing geometrically. *Upgrading and Repairing PCs* covers storage options ranging from larger, faster hard drives to state-of-the-art solid-state storage devices.

When you finish reading this book, you will have the knowledge to upgrade, troubleshoot, and repair almost any system and component.

The 22nd Edition Online Content

Make sure to check out Que's dedicated *Upgrading and Repairing PCs* website! Here, you'll find a cache of helpful material to go along with the book you're holding.

The 22nd edition of *Upgrading and Repairing PCs* includes all-new video that delivers a complete seminar on PC troubleshooting, teaching you how to identify and resolve an array of common and not-so-common PC problems.

From detailed explainers on all the tools that should be a basic part of any PC toolkit, to all the critical rules you should follow to safely operate on your PC's internal components, in these videos Scott Mueller ensures you are armed with everything you need to know to successfully operate on your PC.

From there, Scott takes you through a complete disassembly of an All-in-One (AiO) system, showing just what you can do to keep these specialized systems running smoothly.

Finally, there is a detailed look at today's ultra-fast solid state disk drives (SSD) and the benefits they bring to modern systems.

In addition to the all-new video you'll find Technical Reference material, a repository of reference information that has appeared in previous editions of *Upgrading and Repairing PCs* but has been moved to the web to make room for coverage of newer technologies. You'll also find the complete 19th edition of this book, the complete 20th edition of the book, a detailed list of acronyms, and much more available in printable PDF format. There's more PC hardware content and knowledge here than you're likely to find from any other single source.

I also have a private forum (www.forum.scottmueller.com) designed exclusively to support those who have purchased my recent books and DVDs. I use the forum to answer questions and otherwise help my loyal readers. If you own one of my current books or DVDs, feel free to join in and post questions. I endeavor to answer each question personally, but I also encourage knowledgeable members to respond. Anybody can view the forum without registering, but to post a question of your own you need to join. Even if you don't join in, the forum is a tremendous resource because you can still benefit from all the reader questions I have answered over the years.

My Online Forum

Be sure to check out my private forum (www.forum.scottmueller.com) designed exclusively to support those who have purchased my recent books and DVDs. I use the forum to answer questions and otherwise help my loyal readers. If you own one of my current books or DVDs, feel free to join in and post questions. I endeavor to answer each question personally, but I also encourage knowledgeable members to respond. Anybody can view the forum without registering, but to post a question of your own you need to join. Even if you don't join in, the forum is a tremendous resource because you can still benefit from all the reader questions I have answered over the years.

A Personal Note

When asked which was his favorite Corvette, Dave McLellan, former manager of the Corvette platform at General Motors, always said, "Next year's model." Now with the new 22^{nd} edition, next year's model has just become this year's model, until *next* year that is....

I believe that this book is absolutely the best book of its kind on the market, and that is due in large part to the extensive feedback I have received from both my seminar attendees and book readers. I am so grateful to everyone who has helped me with this book through each edition, as well as all the loyal readers who have been using this book, many of you since the first edition was published. I have had personal contact with many thousands of you in the seminars I have been teaching since 1982, and I enjoy your comments and even your criticisms tremendously. Using this book in a teaching environment has been a major factor in its development. Some of you might be interested to know that I originally began writing this book in early 1985; back then it was self-published and used exclusively in my PC hardware seminars before being professionally published by Que in 1988.

In one way or another, I have been writing and rewriting this book for 30 years! In that time, *Upgrading and Repairing PCs* has proven to be not only the first, but also the most comprehensive and yet approachable and easy-to-understand book of its kind. With this new edition, it is even better than ever. Your comments, suggestions, and support have helped this book to become the best PC hardware book on the market. I look forward to hearing your comments after you see this exciting new edition.

—Scott

In one way or another, I have been writing and rewriting this book for 30 years! In that time, *Upgrading and Repairing PCs* has proven to be not only the first, but also the most comprehensive and yet approachable and easy-to-understand book of its kind. With this new edition, it is even better than ever. Your comments, suggestions, and support have helped this book to become the best PC hardware book on the market. I look forward to hearing your comments after you see this exciting new edition.

—Scott

This page intentionally left blank

Chapter

6

Memory

Memory Basics

This chapter discusses memory from both a physical and a logical point of view. First, you'll examine what memory is, where it fits into the PC architecture, and how it works. Then, you'll look at the various types of memory, speeds, and packaging of the chips and memory modules you can buy and install.

This chapter also covers the logical layout of memory, defining the various areas of memory and their uses from the system's point of view. Because the logical layout and uses are within the "mind" of the processor, memory mapping and logical layout remain perhaps the most difficult subjects to grasp in the PC universe. This chapter contains useful information that removes the mysteries associated with memory and enables you to get the most out of your system.

Memory is the workspace for the processor. It is a temporary storage area where the programs and data being operated on by the processor must reside. Memory storage is considered temporary because the data and programs remain there only as long as the computer has electrical power or is not reset. Before the computer is shut down or reset, any data that has been changed in memory should be saved to a more permanent storage device (usually a hard disk) so it can be reloaded into memory in the future.

Main memory is normally called *random access memory (RAM)* because you can randomly (as opposed to sequentially) access any location. This designation is somewhat misleading and often misinterpreted. Read-only memory (ROM), for example, is also randomly accessible, yet it is usually differentiated from the system RAM because it maintains data without power and can't normally be written to. Although a hard disk can be used as virtual random access memory, we don't consider that RAM either.

Over the years, the definition of RAM has changed from a simple acronym to become something that means the primary memory workspace the processor uses to run programs, which usually is constructed out of a type of chip called *dynamic RAM (DRAM)*. One of the characteristics of DRAM chips (and therefore most types of RAM in general) is that they store data dynamically, which really has two meanings. One meaning is that the information can be written to RAM repeatedly at any time. The other has to do with the fact that DRAM requires the data to be refreshed (essentially rewritten) every few milliseconds or so; faster RAM requires refreshing more often than slower RAM. A type of RAM called *static RAM (SRAM)* does not require the periodic refreshing.

An important characteristic of RAM in general is that data is stored only as long as the memory has electrical power.

Note

Both DRAM and SRAM memory maintain their contents only as long as power is present. However, a different type of memory known as *flash memory* can retain its contents without power, and it is most commonly used today in solid-state drives (SSDs), digital camera and player media, and USB flash drives. As far as the PC is concerned, a flash memory device emulates a disk drive (not RAM) and is accessed by a drive letter, just as with any other disk or optical drive. For more information about flash memory, see Chapter 9, "Flash and Removable Storage."

When we talk about a computer's memory, we usually mean the RAM or physical memory in the system, which are the memory chips or modules the processor uses to store primary active programs and data. This often is confused with the term *storage*, which should be used when referring to things such as disk drives (although they can be used as a form of RAM called virtual memory).

RAM can refer to both the physical chips that make up the memory in the system and the logical mapping and layout of that memory. *Logical mapping* and *layout* refer to how the memory addresses are mapped to actual chips and what address locations contain which types of system information.

People new to computers often confuse main memory (RAM) with disk storage because both have capacities that are expressed in similar megabyte or gigabyte terms. The best analogy I've found to explain the relationship between memory and disk storage is to think of an office with a desk and a file cabinet.

In this popular analogy, the file cabinet represents the system's hard disk, where both programs and data are stored for long-term safekeeping. The desk represents the system's main memory, which allows the person working at the desk (acting as the processor) direct access to any files placed on it. Files represent the programs and documents you can "load" into the memory. To work on a particular file, you first must retrieve it from the cabinet and place it on the desk. If the desk is large enough, you might be able to have several files open on it at one time; likewise, if your system has more memory, you can run more or larger programs and work on more or larger documents.

Adding hard disk space to a system is similar to putting a bigger file cabinet in the office—more files can be permanently stored. And adding more memory to a system is like getting a bigger desk—you can work on more programs and data at the same time.

One difference between this analogy and the way things really work in a computer is that when a file is loaded into memory, it is a copy of the file that is actually loaded; the original still resides on the hard disk. Because of the temporary nature of memory, any files that have been changed after being loaded into memory must then be saved back to the hard disk before the system is powered off (which erases the memory). If the changed file in memory is not saved, the original copy of the file on the hard disk remains unaltered. This is like saying that any changes made to files left on the desktop are discarded when the office is closed, although the original files are still preserved in the cabinet.

Memory temporarily stores programs when they are running, along with the data being used by those programs. RAM chips are sometimes termed *volatile storage* because when you turn off your computer or an electrical outage occurs, whatever is stored in RAM is lost unless you saved it to your hard drive. Because of the volatile nature of RAM, many computer users make it a habit to save their work frequently—a habit I recommend. Many software applications perform periodic saves automatically to minimize the potential for data loss.

Physically, the *main memory* in a system is a collection of chips or modules containing chips that are usually plugged into the motherboard. These chips or modules vary in their electrical and physical designs and must be compatible with the system into which they are being installed to function properly. This chapter discusses the various types of chips and modules that can be installed in different systems.

To better understand physical memory in a system, you should understand which types of memory are found in a typical PC and what the role of each type is. Three main types of physical memory are used in modern PCs. (Remember, I'm talking about the type of memory chip, not the type of module that memory is stored on.)

- **ROM**—Read-only memory
- **DRAM**—Dynamic random access memory
- **SRAM**—Static RAM

The only type of memory you normally need to purchase and install in a system is DRAM. The other types are built in to the motherboard (ROM), processor (SRAM), and other components such as the video card, hard drives, and so on.

ROM

Read-only memory (ROM) is a type of memory that can permanently or semi-permanently store data. It is called *read-only* because it is either impossible or difficult to write to. ROM also is often referred to as *nonvolatile* memory because any data stored in ROM remains there, even if the power is turned off. As such, ROM is an ideal place to put the PC's startup instructions—that is, the software that boots the system.

Note that ROM and RAM are not opposites, as some people seem to believe. Both are simply types of memory. In fact, ROM technically could be classified as a subset of the system's RAM. In other words, a portion of the system's random access memory address space is mapped into one or more ROM chips. This is necessary to contain the software that enables the PC to boot; otherwise, the processor would have no program in memory to execute when it was powered on.

The main ROM BIOS is contained in a ROM chip on the motherboard, but there are also adapter cards with ROMs on them. ROMs on adapter cards contain auxiliary BIOS routines and drivers needed by the particular card, especially for those cards that must be active early in the boot process, such as video cards. Cards that don't need drivers active at boot time typically don't have a ROM because those drivers can be loaded from the hard disk later in the boot process.

Most systems today use a type of ROM called *electrically erasable programmable ROM (EEPROM)*, which is a form of flash memory. Flash is a truly nonvolatile memory that is rewritable, enabling users to easily update the ROM or firmware in their motherboards or any other components (video cards, SCSI cards, and so on).

◄◄ For more information on BIOS upgrades, see the Chapter 5 section "Upgrading the BIOS," **p. 292**.

DRAM

Dynamic RAM (DRAM) is the type of memory chip used for most of the main memory in a modern PC. The main advantages of DRAM are that it is dense, meaning you can pack a lot of bits into a small chip, and it is inexpensive, which makes purchasing large amounts of memory affordable.

The memory cells in a DRAM chip are tiny capacitors that retain a charge to indicate a bit. The problem with DRAM is that it is dynamic—that is, its contents can be changed. With every keystroke or every mouse swipe, the contents of RAM change, and the entire contents of RAM can be wiped out by a system crash. Also, because of the design, it must be constantly refreshed; otherwise, the electrical charges in the individual memory capacitors drain and the data is lost. Refresh occurs when the system memory controller takes a tiny break and accesses all the rows of data in the memory chips. The standard refresh time is 15ms (milliseconds), which means that every 15ms, all the rows in the memory are automatically read to refresh the data.

◀◀ See the Chapter 4 section "Chipsets," **p. 192**.

Unfortunately, refreshing the memory takes processor time away from other tasks because each refresh cycle takes several CPU cycles to complete. In older systems, the refresh cycling could take up to 10% or more of the total CPU time, but with modern systems running in the multigigahertz range, refresh overhead is now on the order of a fraction of a percent or less of the total CPU time. Some systems allow you to alter the refresh timing parameters via the CMOS Setup. The time between refresh cycles is known as *tREF* and is expressed not in milliseconds, but in clock cycles (see Figure 6.1).

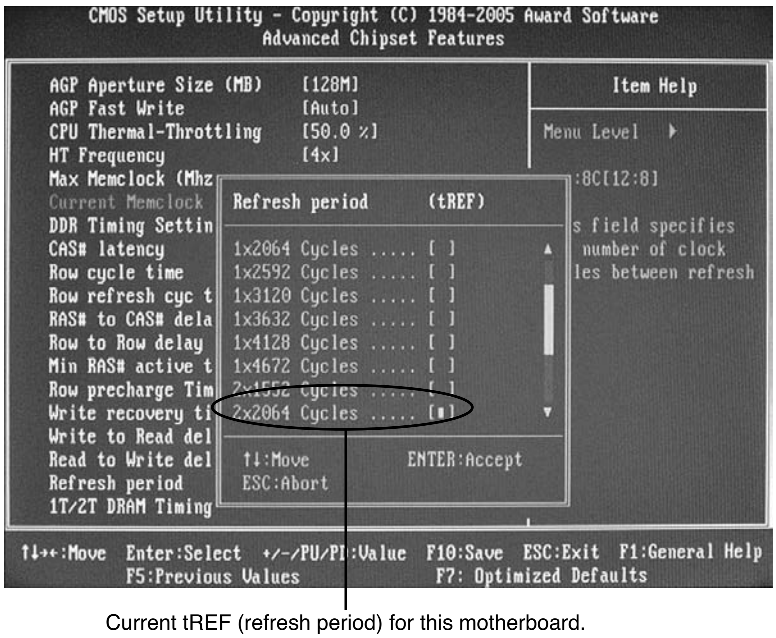


FIGURE 6.1 The refresh period dialog box and other advanced memory timings can be adjusted manually through the BIOS Setup program.

It's important to be aware that increasing the time between refresh cycles (tREF) to speed up your system can allow some of the memory cells to begin draining prematurely, which can cause random soft memory errors to appear.

A *soft error* is a data error that is not caused by a defective chip. To avoid soft errors, it is usually safer to stick with the recommended or default refresh timing. Because refresh consumes less than 1% of modern system overall bandwidth, altering the refresh rate has little effect on performance. It is almost always best to use default or automatic settings for any memory timings in the BIOS Setup. Many modern systems don't allow changes to memory timings and are permanently set to automatic settings. On an automatic setting, the motherboard reads the timing parameters out of the serial presence detect (SPD) ROM found on the memory module and sets the cycling speeds to match.

DRAMs use only one transistor and capacitor pair per bit, which makes them dense, offering more memory capacity per chip than other types of memory. Currently, DRAM chips are being prepared for production with densities up to 4Gb (512MB) per chip, which at one transistor per bit requires at least 4 billion transistors. The transistor count in memory chips is much higher than in processors because in a memory chip the transistors and capacitors are all consistently arranged in a (normally square) grid of simple repetitive structures, unlike processors, which are much more complex circuits of different structures and elements interconnected in a highly irregular fashion.

The transistor for each DRAM bit cell reads the charge state of the adjacent capacitor. If the capacitor is charged, the cell is read to contain a 1; no charge indicates a 0. The charge in the tiny capacitors is constantly draining, which is why the memory must be refreshed constantly. Even a momentary power interruption, or anything that interferes with the refresh cycles, can cause a DRAM memory cell to lose the charge and thus the data. If this happens in a running system, it can lead to blue screens, global protection faults, corrupted files, and any number of system crashes.

DRAM is used in PC systems because it is inexpensive and the chips can be densely packed, so a lot of memory capacity can fit in a small space. Unfortunately, DRAM is also relatively slow—typically much slower than the processor. For this reason, many types of DRAM architectures have been developed to improve performance. These architectures are covered later in the chapter.

Cache Memory: SRAM

Another distinctly different type of memory exists that is significantly faster than most types of DRAM. SRAM stands for static RAM, which is so named because it does not need the periodic refresh rates like DRAM. Because of the way SRAMs are designed, not only are refresh rates unnecessary, but SRAM is much faster than DRAM and much more capable of keeping pace with modern processors.

SRAM memory is available in access times of 0.25ns or less, so it can keep pace with processors running 4GHz or faster. This is because of the SRAM design, which calls for a cluster of six transistors for each bit of storage. The use of transistors but no capacitors means that refresh rates are not necessary because there are no capacitors to lose their charges over time. As long as there is power, SRAM remembers what is stored. With these attributes, why don't we use SRAM for all system memory? The answers are simple.

Compared to DRAM, SRAM is much faster but also much lower in density and much more expensive (see Table 6.1). The lower density means that SRAM chips are physically larger and store fewer bits overall. The high number of transistors and the clustered design mean that SRAM chips are both physically larger and much more expensive to produce than DRAM chips. For example, a high-density DRAM chip might store up to 4Gb (512MB) of RAM, whereas similar-sized SRAM chips can only store up to 72Mb (9MB). The high cost and physical constraints have prevented SRAM from being used as the main memory for PC systems.

Table 6.1 Comparing DRAM and SRAM

Type	Speed	Density	Cost
DRAM	Slow	High	Low
SRAM	Fast	Low	High

Even though SRAM is impractical for PC use as main memory, PC designers have found a way to use SRAM to dramatically improve PC performance. Rather than spend the money for all RAM to be SRAM memory, they design in a small amount of high-speed SRAM memory, used as cache memory, which is much more cost effective. The SRAM cache runs at speeds close to or even equal to the processor and is the memory from which the processor usually directly reads from and writes to. During read operations, the data in the high-speed cache memory is resupplied from the lower-speed main memory or DRAM in advance. To convert access time in nanoseconds to MHz, use the following formula:

1 / nanoseconds × 1000 = MHz

Likewise, to convert from MHz to nanoseconds, use the following inverse formula:

1 / MHz × 1000 = nanoseconds

Today, we have memory that runs faster than 1GHz (1 nanosecond), but up until the late 1990s, DRAM was limited to about 60ns (16MHz) in speed. Up until processors were running at speeds of 16MHz, the available DRAM could fully keep pace with the processor and motherboard, meaning that there was no need for cache. However, as soon as processors crossed the 16MHz barrier, the available DRAM could no longer keep pace, and SRAM cache began to enter PC system designs. This occurred way back in 1986 and 1987 with the debut of systems with the 386 processor running at speeds of 16MHz– 20MHz or faster. These were among the first PC systems to employ what's called *cache memory*, a high-speed buffer made up of SRAM that directly feeds the processor. Because the cache can run at the speed of the processor, it acts as a buffer between the processor and the slower DRAM in the system. The cache controller anticipates the processor memory needs and preloads the high-speed cache memory with data. Then, as the processor calls for a memory address, the data can be retrieved from the high-speed cache rather than the much lower-speed main memory.

Cache effectiveness can be expressed by a hit ratio. This is the ratio of cache hits to total memory accesses. A *hit* occurs when the data the processor needs has been preloaded into the cache from the main memory, meaning the processor can read it from the cache. A cache *miss* is when the cache controller did not anticipate the need for a specific address and the desired data was not preloaded into the cache. In that case the processor must retrieve the data from the slower main memory, instead of the faster cache. Any time the processor reads data from main memory, the processor must wait longer because the main memory cycles at a much slower rate than the processor. As an example, if the processor with integral on-die cache is running at 3.6GHz (3,600MHz) on a 1,333MHz bus, both the processor and the integral cache would be cycling at 0.28ns, whereas the main memory would most likely be cycling almost five times more slowly at 1,333MHz (0.75ns). So, every time the 3.6GHz processor reads from main memory, it would effectively slow down to only 1,333MHz. The slowdown is accomplished by having the processor execute what are called *wait states*, which are cycles in which nothing is done; the processor essentially cools its heels while waiting for the slower main memory to return the desired data. Obviously, you don't want your processors slowing down, so cache function and design become more important as system speeds increase.

To minimize the processor being forced to read data from the slow main memory, two or three stages of cache usually exist in a modern system, called Level 1 (L1), Level 2 (L2), and Level 3 (L3). The L1 cache is also called *integral* or *internal cache* because it has always been built directly into the processor as part of the processor die (the raw chip). Because of this, L1 cache always runs at the full speed of the processor core and is the fastest cache in any system. All 486 and higher processors incorporate integral L1 cache, making them significantly faster than their predecessors. L2 cache was originally called *external cache* because it was external to the processor chip when it first appeared. Originally, this meant it was installed on the motherboard, as was the case with all 386, 486, and first-generation Pentium systems. In those systems, the L2 cache runs at motherboard and CPU bus speed because it is installed on the motherboard and is connected to the CPU bus. You typically find the L2 cache physically adjacent to the processor socket in Pentium and earlier systems.

◄◄ See the Chapter 3 section "Cache Memory," **p. 58.**

In the interest of improved performance, later processor designs from Intel and AMD included the L2 cache as part of the processor. In all processors since late 1999 (and some earlier models), the L2 cache is directly incorporated as part of the processor die, just like the L1 cache. In chips with on-die L2, the cache runs at the full core speed of the processor and is much more efficient. By contrast, most processors from 1999 and earlier with integrated L2 had the L2 cache in separate chips that were external to the main processor core. The L2 cache in many of these older processors ran at only half or one-third the processor core speed. Cache speed is important, so systems having L2 cache on the motherboard were the slowest. Including L2 inside the processor made it faster, and including it directly on the processor die (rather than as chips external to the die) made it faster yet.

A third-level or L3 cache has been present in some processors since 2001. The first desktop PC processor with L3 cache was the Pentium 4 Extreme Edition, a high-end chip introduced in late 2003 with 2MB of on-die L3 cache. Although it seemed at the time that this would be a forerunner of widespread L3 cache in desktop processors, later versions of the Pentium 4 Extreme Edition (as well as its successor, the Pentium Extreme Edition) dropped the L3 cache, instead using larger L2 cache sizes to improve performance. L3 cache made a return to PC processors in 2007 with the AMD Phenom and in 2008 with the Intel Core i7, both of which have four cores on a single die. L3 is especially suited to processors with multiple cores because it provides an on-die cache that all the cores can share. Since 2009, L3 cache has become a staple in most processors with two or more cores. Figure 6.2 shows the L1/L2/L3 cache configuration as reported by CPU-Z (www.cpuid.com) for an Intel Core i5-3570K processor. For other examples, see Figures 3.2 and 3.3, p. 61-62.

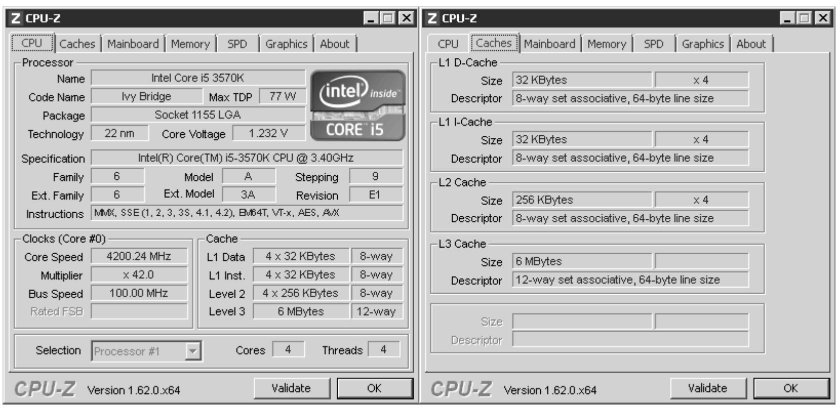


FIGURE 6.2 CPU-Z screenshots showing the CPU/Cache information for an Intel Core i5-3570K processor.

The key to understanding both cache and main memory is to see where they fit in the overall system architecture. See Chapter 4 for diagrams showing recent systems with different types of cache memory.

Memory Standards

For memory to be as inexpensive and interchangeable as possible, industry-standard specifications for both chips and modules have been developed. The Joint Electron Device Engineering Council (JEDEC) Solid State Technology Association creates most industry-standard memory chip and module designs.

JEDEC

JEDEC is the semiconductor engineering standardization body of the Electronic Industries Alliance (EIA), a trade association that represents all areas of the electronics industry. JEDEC, which was created in 1960, governs the standardization of all types of semiconductor devices, integrated circuits, and modules. JEDEC has about 300 member companies, including memory, chipset, and processor manufacturers and practically any company involved in manufacturing computer equipment using industry-standard components.

The idea behind JEDEC is simple: If one company were to create a proprietary memory technology, for example, then other companies that wanted to manufacture components compliant with that memory would have to pay license fees, assuming the company that owned it was interested in licensing at all! Parts would be more proprietary in nature, causing problems with interchangeability or sourcing reasonably priced replacements. In addition, those companies licensing the technology would have no control over future changes the owner company made.

JEDEC helps to prevent that type of scenario for items such as memory by getting all the memory manufacturers to work together creating shared industry standards covering memory chips and modules. JEDEC-approved standards for memory could then be freely shared by all the member companies, and no one single company would have control over a given standard or any of the companies producing compliant components. FPM, SDRAM, DDR, DDR2, DDR3, and DDR4 SDRAM are examples of JEDEC memory standards used in PCs, whereas EDO and RDRAM are proprietary examples. You can find out more about JEDEC standards for memory and other semiconductor technology at www.jedec.org.

Because of variations on speeds (timing), voltage, and other issues, purchasing memory matching the correct industry-standard type doesn't guarantee that it will work in a given system. Always be sure the memory you purchase works with your system or that you can get a refund or replacement if it doesn't. Even though industry standards do exist, allowing modules from many sources to fit a given system, I normally recommend that you look for memory modules the system or memory manufacturer has approved for the system. Often you can find a list of approved modules or suppliers in the system documentation or on the system or memory module manufacturer's website.

Speed and Performance

The speed and performance issues with memory are confusing to some people because of all the different ways to express the speeds of memory and processors. Memory speed was originally expressed in nanoseconds (ns), whereas the speeds of newer forms of memory are usually expressed in megahertz (MHz) and megabytes per second (MBps) instead. Processor speed was originally expressed in megahertz (MHz), whereas most current processor speeds are expressed in gigahertz (GHz). Although all these different speed units might seem confusing, it is relatively simple to translate from one to the other.

A *nanosecond* is defined as one billionth of a second—a short piece of time indeed. To put some perspective on just how small a nanosecond really is, consider that the speed of light is 186,282 miles (299,792 kilometers) per second in a vacuum. In one billionth of a second (one nanosecond), a beam of light travels a mere 11.80 inches or 29.98 centimeters—slightly less than the length of a typical ruler!

Memory speeds have often been expressed in terms of their cycle times (or how long it takes for one cycle), whereas processor speeds have almost always been expressed in terms of their cycle speeds (number of cycles per second). Cycle time and cycle speed are actually just different ways of saying the same thing; that is, you can quote chip speeds in cycles per second, or seconds per cycle, and mean the same thing.

As an analogy, you could express the speed of a vehicle using the same relative terms. In the United States vehicle speeds are normally expressed in miles per hour. If you were driving a car at 60 miles per hour (mph), it would take 1 minute per mile (mpm). At a faster speed of 120mph, it would take only 0.5mpm, and at a slower 30mph speed it would take 2.0mpm. In other words, you could give the speeds as either mph or mpm values, and they would mean exactly the same thing.

Because it is confusing to speak in these different terms for chip speeds, I thought it would be interesting to see exactly how they compare. Table 6.2 shows the relationship between commonly used clock speeds (MHz) and the nanosecond (ns) cycle times they represent.

Table 6.2 Relationship Between Megahertz (MHz) and Cycle Times in Nanoseconds (ns)

Clock Speed	Cycle Time	Clock Speed	Cycle Time	Clock Speed	Cycle Time	Clock Speed	Cycle Time
4.77MHz	210ns	250MHz	4.0ns	850MHz	1.18ns	2,700MHz	0.37ns
6MHz	167ns	266MHz	3.8ns	866MHz	1.15ns	2,800MHz	0.36ns
8MHz	125ns	300MHz	3.3ns	900MHz	1.11ns	2,900MHz	0.34ns
10MHz	100ns	333MHz	3.0ns	933MHz	1.07ns	3,000MHz	0.333ns
12MHz	83ns	350MHz	2.9ns	950MHz	1.05ns	3,100MHz	0.323ns
16MHz	63ns	366MHz	2.7ns	966MHz	1.04ns	3,200MHz	0.313ns
20MHz	50ns	400MHz	2.5ns	1,000MHz	1.00ns	3,300MHz	0.303ns
25MHz	40ns	433MHz	2.3ns	1,100MHz	0.91ns	3,400MHz	0.294ns
33MHz	30ns	450MHz	2.2ns	1,133MHz	0.88ns	3,500MHz	0.286ns
40MHz	25ns	466MHz	2.1ns	1,200MHz	0.83ns	3,600MHz	0.278ns
50MHz	20ns	500MHz	2.0ns	1,300MHz	0.77ns	3,700MHz	0.270ns
60MHz	17ns	533MHz	1.88ns	1,400MHz	0.71ns	3,800MHz	0.263ns
66MHz	15ns	550MHz	1.82ns	1,500MHz	0.67ns	3,900MHz	0.256ns
75MHz	13ns	566MHz	1.77ns	1,600MHz	0.63ns	4,000MHz	0.250ns
80MHz	13ns	600MHz	1.67ns	1,700MHz	0.59ns	4,100MHz	0.244ns
100MHz	10ns	633MHz	1.58ns	1,800MHz	0.56ns	4,200MHz	0.238ns
120MHz	8.3ns	650MHz	1.54ns	1,900MHz	0.53ns	4,300MHz	0.233ns
133MHz	7.5ns	666MHz	1.50ns	2,000MHz	0.50ns	4,400MHz	0.227ns
150MHz	6.7ns	700MHz	1.43ns	2,100MHz	0.48ns	4,500MHz	0.222ns
166MHz	6.0ns	733MHz	1.36ns	2,200MHz	0.45ns	4,600MHz	0.217ns
180MHz	5.6ns	750MHz	1.33ns	2,300MHz	0.43ns	4,700MHz	0.213ns
200MHz	5.0ns	766MHz	1.31ns	2,400MHz	0.42ns	4,800MHz	0.208ns
225MHz	4.4ns	800MHz	1.25ns	2,500MHz	0.40ns	4,900MHz	0.204ns
233MHz	4.3ns	833MHz	1.20ns	2,600MHz	0.38ns	5,000MHz	0.200ns

As you can see from Table 6.2, as clock speed increases, cycle time decreases proportionately, and vice versa.

Over the evolutionary life of the PC, main memory (what we call RAM) has had a difficult time keeping up with the processor, requiring several levels of high-speed cache memory to intercept processor requests for the slower main memory. More recently, however, systems using DDR, DDR2, DDR3, and DDR4 SDRAM have memory bus transfer rates (bandwidth) capable of equaling that of the external processor bus. When the speed of the memory bus equals the speed of the processor bus (or some even multiple thereof), main memory performance is closest to optimum for that system.

For example, using the information in Table 6.2, you can see that the 60ns DRAM memory used in the original Pentium and Pentium II PCs up until 1998 works out to be an extremely slow 16.7MHz! This slow 16.7MHz memory was installed in systems running processors up to 300MHz or faster with external processor bus speeds of up to 66MHz, resulting in a large mismatch between processor bus and main memory performance. To alleviate this performance gap, starting in 1998 the industry shifted to faster SDRAM memory, which could match the 66MHz and 100MHz processor bus speeds in use at that time. From that point forward, memory and especially memory bus performance has largely evolved in step with the processor bus, coming out with newer and faster types to match any increases in processor bus speeds.

By the year 2000, the dominant processor bus and memory speeds had increased to 100MHz and even 133MHz, called PC100 and PC133 SDRAM, respectively. Starting in early 2001, double data rate (DDR) SDRAM memory of 200MHz and 266MHz became popular. In 2002, DDR memory increased to 333MHz, and in 2003, the speeds increased further to 400MHz. In 2004, we saw the introduction of DDR2, first at 400MHz and then at 533MHz. DDR2 memory continued to match processor bus speed increases in PCs from 2005 to 2006, rising to 667MHz and 800MHz during that time. By 2007, DDR2 memory was available at speeds of up to 1066MHz. By late 2007, DDR3 came on the market at speeds of 1066MHz, with 1333MHz and 1600MHz appearing in 2008. In 2009, DDR3 memory became the most popular memory type in new systems, and faster speed grades of 1866MHz and 2133MHz were added. DDR4 was released in 2013, with a speed of 1600MHz and expected future speeds of up to 3200MHz. DDR4-based systems began to reach the market in late summer 2014. Table 6.3 lists the primary types and performance levels of PC memory.

Table 6.3 PC Memory Types and Performance Levels

Memory Type	Years Popular	Desktop Module Type	Laptop Module Type	Voltage	Max. Clock Speed	Max. Throughput Single-Channel	Max. Throughput Dual-Channel	Max. Throughput Triple-Channel
Fast Page Mode (FPM) DRAM	1987–1995	30/72-pin SIMM	72/144-pin SODIMM	5V	22MHz	177MBps	N/A	N/A
Extended Data Out (EDO) DRAM	1995–1998	72-pin SIMM	72/144-pin SODIMM	5V	33MHz	266MBps	N/A	N/A
Single Data Rate (SDR) SDRAM	1998–2002	168-pin DIMM	144-pin SODIMM	3.3V	133MHz	1,066MBps	N/A	N/A
Double Data Rate (DDR) SDRAM	2002–2005	184-pin DIMM	200-pin SODIMM	2.5V	400MTps	3,200MBps	6,400MBps	N/A
DDR2 SDRAM	2005–2009	240-pin DDR2 DIMM	200-pin SODIMM	1.8V	1,066MTps	8,533MBps	17,066MBps	N/A
DDR3 SDRAM	2009–2015	240-pin DDR3 DIMM	204-pin SODIMM	1.5V	2,133MTps	17,066MBps	34,133MBps	51,200MBps
DDR4 SDRAM	2015+	284-pin DDR4 DIMM	256-pin SODIMM	1.2V	4,266MTps	34,133MBps	68,266MBps	102,400MBps

MHz = Million cycles per second

MTps = Million transfers per second

MBps = Million bytes per second

DIMM = Dual inline memory module

SODIMM = Small outline DIMM

SIMM = Single inline memory module

Another specification to consider that is related to speed is the *CAS (column address strobe) latency*, which is often abbreviated as *CL*. This is also sometimes called *read latency*, and it's the number of clock cycles occurring between the registration of the CAS signal and the resultant output data, with lower numbers of cycles indicating faster (better) performance. If possible, choose modules with a lower CL figure because the motherboard chipset reads that specification out of the SPD (serial presence detect) ROM on the module and takes advantage of the lower latency through improved memory controller timings. Figure 6.3 shows the memory timing and SPD information as reported by CPU-Z (www.cpuid.com) for a system with DDR3-1600 SDRAM.

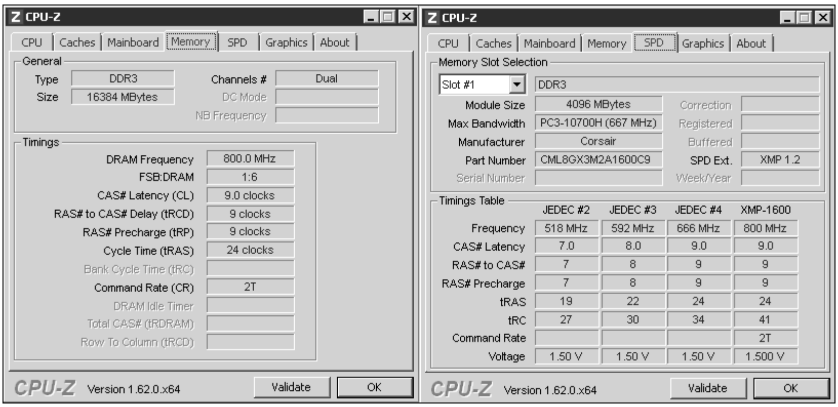


FIGURE 6.3 CPU-Z screenshots showing the Memory/SPD information for a system with DDR3-1600 SDRAM.

The following sections look at these memory types in more detail.

Fast Page Mode DRAM

Standard DRAM is accessed through a technique called *paging*. Normal memory access requires that a row and column address be selected, which takes time. Paging enables faster access to all the data within a given row of memory by keeping the row address the same and changing only the column. Memory that uses this technique is called *Page Mode* or *Fast Page Mode* memory. Other variations on Page Mode were called Static Column or Nibble Mode memory.

Paged memory is a simple scheme for improving memory performance that divides memory into pages ranging from 512 bytes to a few kilobytes long. The paging circuitry then enables memory locations in a page to be accessed with fewer wait states. If the desired memory location is outside the current page, one or more wait states are added while the system selects the new page.

To improve further on memory access speeds, systems have evolved to enable faster access to DRAM. One important change was the implementation of burst mode access in the 486 and later processors. Burst mode cycling takes advantage of the consecutive nature of most memory accesses. After setting up the row and column addresses for a given access, using burst mode, you can then access the next three adjacent addresses with no additional latency or wait states. A burst access usually is limited to four total accesses. To describe this, we often refer to the timing in the number of cycles for each access. A typical burst mode access of standard DRAM is expressed as x-y-y-y; x is the time for the first access (latency plus cycle time), and y represents the number of cycles required for each consecutive access.

Standard 60ns-rated DRAM normally runs 5-3-3-3 burst mode timing. This means the first access takes a total of five cycles (on a 66MHz system bus, this is about 75ns total, or 5 × 15ns cycles), and the consecutive cycles take three cycles each (3 × 15ns = 45ns). As you can see, the actual system timing is somewhat less than the memory is technically rated for. Without the bursting technique, memory access would be 5-5-5-5 because the full latency is necessary for each memory transfer. The 45ns cycle time during burst transfers equals about a 22.2MHz effective clock rate; on a system with a 64-bit (8-byte) wide memory bus, this would result in a maximum throughput of 177MBps (22.2MHz × 8 bytes = 177MBps).

DRAM memory that supports paging and this bursting technique is called *Fast Page Mode (FPM)* memory. This term refers to the ability to access data on the same memory page faster than data on other memory pages.

Most 386, 486, and Pentium systems from 1987 through 1995 used FPM memory, which came in either 30-pin or 72-pin SIMM form.

Another technique for speeding up FPM memory is called *interleaving*. In this design, two separate banks of memory are used together, alternating access from one to the other as even and odd bytes. While one is being accessed, the other is being precharged, when the row and column addresses are being selected. Then, by the time the first bank in the pair is finished returning data, the second bank in the pair is finished with the latency part of the cycle and is now ready to return data. While the second bank is returning data, the first bank is being precharged, selecting the row and column address of the next access. This overlapping of accesses in two banks reduces the effect of the latency or precharge cycles and allows for faster overall data retrieval. The only problem is that to use interleaving, you must install identical pairs of banks together, doubling the number of modules required.

Extended Data Out RAM

In 1995, a newer type of DRAM called extended data out (EDO) RAM became available for Pentium systems. EDO, a modified form of FPM memory, is sometimes referred to as Hyper Page mode. EDO was invented and patented by Micron Technology, although Micron licensed production to many other memory manufacturers.

EDO memory consists of specially manufactured chips that allow a timing overlap between successive accesses. The name *extended data out* refers specifically to the fact that, unlike FPM, the data output drivers on the chip are not turned off when the memory controller removes the column address to begin the next cycle. This enables the next cycle to overlap the previous one, saving approximately 10ns per cycle.

The effect of EDO is that cycle times are improved by enabling the memory controller to begin a new column address instruction while it is reading data at the current address. This is almost identical to what was achieved in older systems by interleaving banks of memory, but unlike interleaving, with EDO you didn't need to install two identical banks of memory in the system at a time.

EDO RAM allows for burst mode cycling of 5-2-2-2, compared to the 5-3-3-3 of standard fast page mode memory. To do four memory transfers, then, EDO would require 11 total system cycles, compared to 14 total cycles for FPM. This is a 22% improvement in overall cycling time. The resulting two-cycle (30ns) cycle time during burst transfers equals a 33.3MHz effective clock rate, compared to 45ns/22MHz for FPM. On a system with a 64-bit (8-byte) wide memory bus, this would result in a maximum throughput of 266MBps (33.3MHz × 8 bytes = 266MBps). Due to the processor cache, EDO typically increased overall system benchmark speed by only 5% or less. Even though the overall system improvement was small, the important thing about EDO was that it used the same basic DRAM chip design as FPM, meaning that there was practically no additional cost over FPM. In fact, in its heyday EDO cost less than FPM yet offered higher performance.

EDO RAM generally came in 72-pin SIMM form. Figure 6.4 (later in this chapter) shows the physical characteristics of these SIMMs.

To actually use EDO memory, your motherboard chipset had to support it. Most motherboard chipsets introduced on the market from 1995 (Intel 430FX) through 1997 (Intel 430TX) offered support for EDO, making EDO the most popular form of memory in PCs from 1995 through 1998. Because EDO memory chips cost the same to manufacture as standard chips, combined with Intel's support of EDO in motherboard chipsets, the PC market jumped on the EDO bandwagon full force.

◀◀ See the Chapter 4 sections "Fifth-Generation (P5 Pentium Class) Chipsets," **p. 202**, and "Sixth-Generation (P6 Pentium Pro/II/III Class) Chipsets," **p. 203**.

EDO RAM was used in systems with CPU bus speeds of up to 66MHz, which fit perfectly with the PC market up through 1998. However, starting in 1998, with the advent of 100MHz and faster system bus speeds, the market for EDO rapidly declined, and faster SDRAM architecture became the standard.

One variation of EDO that never caught on was called *burst EDO (BEDO)*. BEDO added burst capabilities for even speedier data transfers than standard EDO. Unfortunately, the technology was owned by Micron and not a free industry standard, so only one chipset (Intel 440FX Natoma) ever supported it. BEDO was quickly overshadowed by industry-standard SDRAM, which came into favor among PC system chipset and system designers over proprietary designs. As such, BEDO never really saw the light of production, and to my knowledge no systems ever used it.

SDRAM

SDRAM is short for synchronous DRAM, a JEDEC standard for a type of DRAM that runs in synchronization with the memory bus. SDRAM delivers information in very high-speed bursts using a high-speed clocked interface. SDRAM removes most of the latency involved in asynchronous DRAM because the signals are already in synchronization with the motherboard clock.

As with any newly introduced type of memory on the market, motherboard chipset support is required before it can be usable in systems. Starting in 1996 with the 430VX and 430TX, most of Intel's chipsets began to support industry-standard SDRAM, and in 1998 the introduction of the 440BX chipset caused SDRAM to eclipse EDO as the most popular type on the market.

SDRAM performance is dramatically improved over that of FPM or EDO RAM. However, because SDRAM is still a type of DRAM, the initial latency is the same, but burst mode cycle times are much faster than with FPM or EDO. SDRAM timing for a burst access would be 5-1-1-1, meaning that four memory reads would complete in only eight system bus cycles, compared to 11 cycles for EDO and 14 cycles for FPM. This makes SDRAM almost 20% faster than EDO.

Besides being capable of working in fewer cycles, SDRAM is capable of supporting up to 133MHz (7.5ns) system bus cycling. Most PC systems sold from 1998 through 2002 included SDRAM memory.

SDRAM is sold in DIMM form and is normally rated by clock speed (MHz) rather than cycling time (ns), which was confusing during the initial change from FPM and EDO DRAM. Figure 6.5 (later in this chapter) shows the physical characteristics of DIMMs.

To meet the stringent timing demands of its chipsets, Intel created specifications for SDRAM called PC66 and PC100. For example, you would think 10ns would be considered the proper rating for 100MHz operation, but the PC100 specification promoted by Intel called for faster 8ns memory to ensure all timing parameters could be met with sufficient margin for error.

In May 1999, JEDEC created a specification called PC133. It achieved this 33MHz speed increase by taking the PC100 specification and tightening up the timing and capacitance parameters. The faster PC133 quickly caught on for any systems running a 133MHz processor bus. The original chips used in PC133 modules were rated for exactly 7.5ns or 133MHz; later chips were rated at 7.0ns, which is technically 143MHz. These faster chips were still used on PC133 modules, but they allowed for improvements in column address strobe latency (abbreviated as CAS or CL), which somewhat improves overall memory cycling time.

SDRAM normally came in 168-pin DIMMs, running at several speeds. Table 6.4 shows the standard single data rate SDRAM module speeds and resulting throughputs.

Table 6.4 JEDEC Standard SDRAM Module (168-Pin DIMM) Speeds and Transfer Rates

Module Type	Chip Type	Clock Speed	Cycles per Clock	Bus Speed	Bus Width	Transfer Rate
PC66	15ns 10ns	66MHz	1	66MTps	8 bytes	533MBps
PC100	8ns	100MHz	1	100MTps	8 bytes	800MBps
PC133	7.5ns 7ns	133MHz	1	133MTps	8 bytes	1,066MBps

MHz = Million cycles per second
MTps = Million transfers per second
MBps = Million bytes per second
ns = Nanoseconds (billionths of a second)

▶▶ See "Memory Modules," **p. 375**.

Some module manufacturers sold modules they claimed were "PC150" or "PC166," even though those speeds did not exist as official JEDEC or Intel standards, and no chipsets or processors officially supported those speeds. These modules actually used hand-picked 133MHz-rated chips that could run overclocked at 150MHz or 166MHz speeds. In essence, PC150 or PC166 memory was PC133 memory that was tested to run at overclocked speeds not supported by the original chip manufacturer. This overclockable memory was sold at a premium to enthusiasts who wanted to overclock their motherboard chipsets, thereby increasing the speed of the processor and memory bus.

Caution

In general, PC133 memory is considered to be backward compatible with PC100 memory. However, some chipsets or motherboards had more specific requirements for specific types of 100MHz or 133MHz chips and module designs. If you need to upgrade an older system that requires PC100 memory, you should not purchase PC133 memory unless the memory is specifically identified by the memory vendor as being compatible with the system. You can use the online memory-configuration tools provided by most major memory vendors to ensure that you get the right memory for your system.

Typically, you find SDRAM modules rated CL 2 or CL 3.

DDR SDRAM

DDR SDRAM memory is a JEDEC standard that is an evolutionary upgrade in which data transfers twice as quickly as standard SDRAM. Instead of doubling the actual clock rate, DDR memory achieves the doubling in performance by transferring twice per transfer cycle: once at the leading (falling) edge and once at the trailing (rising) edge of the cycle (see Figure 6.4). This effectively doubles the transfer rate, even though the same overall clock and timing signals are used. To eliminate confusion with DDR, regular SDRAM is often called *single data rate (SDR)*.

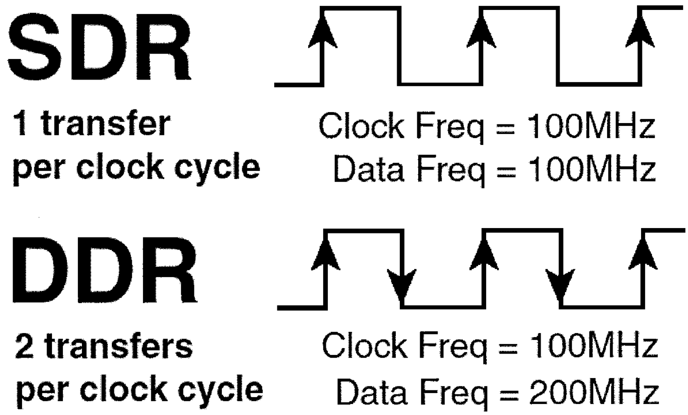


FIGURE 6.4 SDR versus DDR cycling.

DDR SDRAM first came to market in the year 2000 and was initially used on high-end graphics cards because there were no motherboard chipsets to support it at the time. DDR finally became popular in 2002 with the advent of mainstream supporting motherboards and chipsets. From 2002 through 2005, DDR was the most popular type of memory in mainstream PCs. DDR SDRAM uses a DIMM module design with 184 pins. Figure 6.6 (later in this chapter) shows the 184-pin DDR DIMM.

DDR DIMMs come in a variety of speed or throughput ratings and normally run on 2.5 volts. Table 6.5 compares the types of industry-standard DDR SDRAM modules. As you can see, the raw chips are designated by their speed in megatransfers per second, whereas the modules are designated by their approximate throughput in megabytes per second.

Table 6.5 JEDEC Standard DDR Module (184-Pin DIMM) Speeds and Transfer Rates

Module Type	Chip Type	Base Clock Speed	Cycle Time	Cycles per Clock	Bus Speed	Bus Width	Module Transfer Rate	Dual-Channel Transfer Rate
PC1600	DDR200	100MHz	10.0ns	2	200MTps	8 bytes	1,600MBps	3,200MBps
PC2100	DDR266	133MHz	7.5ns	2	266MTps	8 bytes	2,133MBps	4,266MBps
PC2700	DDR333	166MHz	6.0ns	2	333MTps	8 bytes	2,667MBps	5,333MBps
PC3200	DDR400	200MHz	5.0ns	2	400MTps	8 bytes	3,200MBps	6,400MBps

DDR = Double data rate
MHz = Million cycles per second
MTps = Million transfers per second
MBps = Million bytes per second
ns = Nanoseconds (billionths of a second)

The major memory chip and module manufacturers normally produce parts that conform to the official JEDEC standard speed ratings. However, to support overclocking, several memory module manufacturers purchase unmarked and untested chips from the memory chip manufacturers and then independently test and sort them by how fast they run. These are then packaged into modules with unofficial designations and performance figures that exceed the standard ratings. Table 6.6 shows the popular unofficial speed ratings I've seen on the market. Note that because the speeds of these modules are beyond the standard default motherboard and chipset speeds, you won't see an advantage to using them unless you are overclocking your system to match.

Table 6.6 Overclocked (Non-JEDEC) DDR Module (184-Pin DIMM) Speeds and Transfer Rates

Module Standard	Chip Type	Clock Speed (MHz)	Cycles per Clock	Bus Speed (MTps)	Bus Width (Bytes)	Transfer Rate (MBps)	Dual-Channel Transfer Rate (MBps)
PC3500	DDR433	216	2	433	8	3,466	6,933
PC3700	DDR466	233	2	466	8	3,733	7,466
PC4000	DDR500	250	2	500	8	4,000	8,000
PC4200	DDR533	266	2	533	8	4,266	8,533
PC4400	DDR550	275	2	550	8	4,400	8,800
PC4800	DDR600	300	2	600	8	4,800	9,600

DDR = Double data rate

MHz = Million cycles per second

MTps = Million transfers per second

MBps = Million bytes per second

ns = Nanoseconds (billionths of a second)

Most chipsets that support DDR also support dual-channel operation—a technique in which two matching DIMMs are installed to function as a single bank, with double the bandwidth of a single module. For example, if a chipset supports standard PC3200 modules, the bandwidth for a single module would be 3,200MBps. However, in dual-channel mode, the total bandwidth would double to 6,400MBps. Dual-channel operation optimizes PC design by ensuring that the CPU bus and memory bus both run at the same speeds (meaning throughput, not MHz) so that data can move synchronously between the buses without delays.

The cycle time in nanoseconds (billionths of a second) matches the base clock speed, but DDR modules transfer twice per cycle, so the bus speed frequency is always equal to double the clock frequency. The throughput or bandwidth is simply the bus frequency times the width, which gives the rate at which data can be read from or written to the module.

Typically, you can find DDR modules rated CL 2, CL 2.5, or CL 3.

With DDR, it is generally okay to install a module that is faster than the system requires, but you should not install a slower module than the motherboard requires. For example, you can usually install PC2700 memory even if the system requires only PC2100 or even PC1600, but if the system requires PC2700, you should not install the slower PC2100 or PC1600 modules.

DDR2 SDRAM

DDR2 is a faster version of DDR memory. It achieves higher throughput by using differential pairs of signal wires to allow faster signaling without noise and interference problems. DDR2 is still double data rate just as with DDR, but the modified signaling method enables you to achieve higher clock speeds with more immunity to noise and crosstalk between the signals. The additional signals required for differential pairs add to the pin count—DDR2 DIMMs have 240 pins, which is more than the 184 pins of DDR. The original DDR specification officially topped out at 400MHz (although faster unofficial overclocked modules were produced), whereas DDR2 starts at 400MHz and goes up to an official maximum of 1,066MHz.

JEDEC began working on the DDR2 specification in April 1998 and published the standard in September 2003. DDR2 chip and module production actually began in mid-2003 (mainly samples and prototypes), and the first chipsets, motherboards, and systems supporting DDR2 appeared for Intel processor–based systems in mid-2004. At that time, variations of DDR2 such as G-DDR2 (Graphics DDR2) began appearing in graphics cards as well. Mainstream motherboard chipset support for DDR2 on Intel processor–based systems appeared in 2005. Notable for its lack of DDR2 support through 2005 was AMD, whose Athlon 64 and Opteron processor families included integrated DDR memory controllers. AMD processor–based systems first supported DDR2 in mid-2006, with the release of socket AM2 motherboards and processors to match. (AMD's Socket F, otherwise known as 1207 FX, also supports DDR2 memory.)

Note that AMD was almost two years behind Intel in the transition from DDR to DDR2. This is because AMD included the memory controller in its Athlon 64 and all subsequent processors, rather than incorporating the memory controller in the chipset North Bridge, as with the more traditional Intel designs. Although there are advantages to integrating the memory controller in the CPU, one disadvantage is the inability to quickly adopt new memory architectures because doing so requires that both the processor and processor socket be redesigned. However, with the release of the Core i-series processors in 2008, Intel also moved the memory controller from the chipset into the processor, thus putting Intel and AMD in the same situation in terms of memory architecture.

In addition to providing greater speeds and bandwidth, DDR2 has other advantages. It uses lower voltage than conventional DDR (1.8V versus 2.5V), so power consumption and heat generation are reduced. Because of the greater number of pins required on DDR2 chips, the chips typically use fine-pitch ball grid array (FBGA) packaging rather than the thin small outline package (TSOP) chip packaging used by most DDR and conventional SDRAM chips. FPGA chips connect to the substrate (meaning the memory module in most cases) via tightly spaced solder balls on the base of the chip.

Table 6.7 shows the various official JEDEC-approved DDR2 module types and bandwidth specifications.

Table 6.7 JEDEC Standard DDR2 Module (240-Pin DIMM) Speeds and Transfer Rates

Module Type	Chip Type	Base Clock Speed	Cycle Time	Cycles per Clock	Bus Speed	Bus Width	Module Transfer Rate	Dual-Channel Transfer Rate
PC2-3200	DDR2-400	200MHz	5.00ns	2	400MTps	8 bytes	3,200MBps	6,400MBps
PC2-4200	DDR2-533	266MHz	3.75ns	2	533MTps	8 bytes	4,266MBps	8,533MBps
PC2-5300	DDR2-667	333MHz	3.00ns	2	667MTps	8 bytes	5,333MBps	10,667MBps
PC2-6400	DDR2-800	400MHz	2.50ns	2	800MTps	8 bytes	6,400MBps	12,800MBps
PC2-8500	DDR2-1066	533MHz	1.88ns	2	1,066MTps	8 bytes	8,533MBps	17,066MBps

DDR = Double data rate

MHz = Million cycles per second

MTps = Million transfers per second

MBps = Million bytes per second

ns = Nanoseconds (billionths of a second)

The fastest official JEDEC-approved standard is DDR2-1066, which is composed of chips that run at an effective speed of 1,066MHz (really megatransfers per second), resulting in modules designated PC2-8500 having a bandwidth of 8,533MBps. However, just as with DDR, many of the module manufacturers produce even faster modules designed for overclocked systems. These are sold as modules with unofficial designations and performance figures that exceed the standard ratings. Table 6.8 shows the popular unofficial speed ratings I've seen on the market. Note that because the speeds of these modules are beyond the standard default motherboard and chipset speeds, you won't see an advantage to using these unless you are overclocking your system to match.

Table 6.8 Overclocked (Non-JEDEC) DDR2 Module (240-Pin DIMM) Speeds and Transfer Rates

Module Standard	Chip Type	Clock Speed (MHz)	Cycles per Clock	Bus Speed (MTps)	Bus Width (Bytes)	Transfer Rate (MBps)	Dual Channel-Transfer Rate (MBps)
PC2-6000	DDR2-750	375	2	750	8	6,000	12,000
PC2-7200	DDR2-900	450	2	900	8	7,200	14,400
PC2-8000	DDR2-1000	500	2	1000	8	8,000	16,000
PC2-8800	DDR2-1100	550	2	1100	8	8,800	17,600
PC2-8888	DDR2-1111	556	2	1111	8	8,888	17,777
PC2-9136	DDR2-1142	571	2	1142	8	9,136	18,272
PC2-9200	DDR2-1150	575	2	1150	8	9,200	18,400
PC2-9600	DDR2-1200	600	2	1200	8	9,600	19,200
PC2-10000	DDR2-1250	625	2	1250	8	10,000	20,000

DDR = Double data rate

MHz = Million cycles per second

MTps = Million transfers per second

MBps = Million bytes per second

ns = Nanoseconds (billionths of a second)

Typically, you can find DDR2 modules rated between CL 3 and CL 6.

DDR3 SDRAM

DDR3 enables higher levels of performance along with lower power consumption and higher reliability than DDR2. JEDEC began working on the DDR3 specification in June 2002, and the first DDR3 memory modules and supporting chipsets (versions of the Intel 3x series) were released for Intel-based systems in mid-2007. Due to initial high cost and limited support, DDR3 didn't start to become popular until late 2008 when Intel released the Core i7 processor, which included an integrated tri-channel DDR3 memory controller. In early 2009, popularity increased when AMD released Socket AM3 versions of the Phenom II, the first from AMD to support DDR3. In 2009, with full support from both Intel and AMD, DDR3 finally began to achieve price parity with DDR2, causing DDR3 to begin to eclipse DDR2 in sales.

DDR3 modules use advanced signaling techniques, including self-driver calibration and data synchronization, along with an optional onboard thermal sensor. DDR3 memory runs on only 1.5V, which is nearly 20% less than the 1.8V that DDR2 memory uses. The lower voltage combined with higher efficiency reduces overall power consumption by up to 30% compared to DDR2.

The 240-pin DDR3 modules are similar in pin count, size, and shape to the DDR2 modules; however, the DDR3 modules are incompatible with the DDR2 circuits and are designed with different keying to make them physically noninterchangeable.

DDR3 modules are available in speeds of 800MHz (effective) and higher. Just as with DDR and DDR2, the true clock speed is half the effective rate, which is technically expressed in million transfers per second (MTps). Table 6.9 shows the JEDEC-approved DDR3 module types and bandwidth specifications.

Table 6.9 JEDEC Standard DDR3 Module (240-Pin DIMM) Speeds and Transfer Rates

Module Type	Chip Type	Base Clock Speed	Cycle Time	Cycles per Clock	Bus Speed	Bus Width	Module Transfer Rate	Dual-Channel Transfer Rate	Tri-Channel Transfer Rate
PC3-6400	DDR3-800	400MHz	2.50ns	2	800MTps	8 bytes	6,400MBps	12,800MBps	19,200MBps
PC3-8500	DDR3-1066	533MHz	1.88ns	2	1,066MTps	8 bytes	8,533MBps	17,066MBps	25,600MBps
PC3-10600	DDR3-1333	667MHz	1.50ns	2	1,333MTps	8 bytes	10,667MBps	21,333MBps	32,000MBps
PC3-12800	DDR3-1600	800MHz	1.25ns	2	1,600MTps	8 bytes	12,800MBps	25,600MBps	38,400MBps
PC3-14900	DDR3-1866	933MHz	1.07ns	2	1,866MTps	8 bytes	14,933MBps	29,866MBps	44,800MBps
PC3-17000	DDR3-2133	1066MHz	0.94ns	2	2,133MTps	8 bytes	17,066MBps	34,133MBps	51,200MBps

DDR = Double data rate

MHz = Million cycles per second

MTps = Million transfers per second

MBps = Million bytes per second

ns = Nanoseconds (billionths of a second)

The fastest official JEDEC-approved standard is DDR3-2133, which is composed of chips that run at an effective speed of 2,133MHz (really megatransfers per second), resulting in modules designated PC3-17000 and having a bandwidth of 17,066MBps. However, just as with DDR and DDR2, many manufacturers produce nonstandard modules designed for overclocked systems. These are sold as modules with unofficial designations, clock speeds, and performance figures that exceed the standard ratings.

Table 6.10 shows the popular unofficial DDR3 speed ratings I've seen on the market. Note that because the speeds of these modules don't conform to the standard default motherboard and chipset speeds, you won't see an advantage to using them unless you are overclocking your system and your motherboard supports the corresponding overclocked processor and memory settings that these modules require. In addition, because these modules use standard-speed chips that are running overclocked, they almost always require custom voltage settings that are higher than the 1.5V that standard DDR3 memory uses. For system stability, I generally don't recommend using overclocked (higher voltage) memory, instead preferring to use only that which runs on the DDR3 standard 1.5V.

Table 6.10 Overclocked (Non-JEDEC) DDR3 Module (240-Pin DIMM) Speeds and Transfer Rates

Module Standard	Chip Type	Clock Speed (MHz)	Cycles per Clock	Bus Speed (MTps)	Bus Width (Bytes)	Transfer Rate (MBps)	Dual-Channel Transfer Rate (MBps)	Tri-Channel Transfer (MBps)
PC3-11000	DDR3-1375	688	2	1375	8	11,000	22,000	33,000
PC3-13000	DDR3-1625	813	2	1625	8	13,000	26,000	39,000
PC3-14400	DDR3-1800	900	2	1800	8	14,400	28,800	43,200
PC3-14900	DDR3-1866	933	2	1866	8	14,933	29,866	44,800
PC3-15000	DDR3-1866	933	2	1866	8	14,933	29,866	44,800
PC3-16000	DDR3-2000	1000	2	2000	8	16,000	32,000	48,000

DDR = Double data rate
MHz = Million cycles per second
MTps = Million transfers per second
MBps = Million bytes per second
ns = Nanoseconds (billionths of a second)

Typically, you can find DDR3 modules rated CL 5 through CL10.

DDR4 SDRAM

DDR4 is the most recent JEDEC memory standard. It enables higher levels of performance along with lower power consumption and higher reliability than DDR3 does. JEDEC began working on the DDR4 specification in 2005, with the final specification published in September 2012. Samsung produced the first prototype DDR4 modules in late 2010 and released the first sample 16GB DDR4 module in July 2012. The first motherboards supporting DDR4 memory were released in August 2014, using Intel's X99 chipset.

DDR4 modules use a Pseudo Open Drain (POD) interface (previously used in high-performance graphic DRAM) and run on a lower 1.2V voltage (compared to 1.5V for DDR3). This enables DDR4 modules to consume about 40% less power overall than previous DDR3 modules, thus saving energy while also producing less heat. DDR4 also supports write Cyclic Redundancy Check (CRC) to improve system reliability.

288-pin DDR4 modules are 1mm longer and 1mm taller than 240-pin DDR3/DDR2 modules. This was accomplished by making the individual pins only 0.85mm wide, versus the 1mm wide pins used on previous modules. DDR4 modules also feature a slight curvature about halfway between each edge and the center notch, making the outside pins shorter than the pins nearer the center notch for easier installation. Because of the different size and signaling used, DDR4 modules are both physically and electrically incompatible with previous memory module and socket designs (see Figure 6.5).

DDR4 modules are available in speeds of 1,600MHz (effective) and higher, with speeds of up to 3,200MHz (effective) expected in the future. Just as with DDR, DDR2, and DDR3 the true clock speed is half the effective rate, which is technically expressed in million transfers per second (MTps). Table 6.11 shows the official JEDEC-approved DDR4 module types and bandwidth specifications.

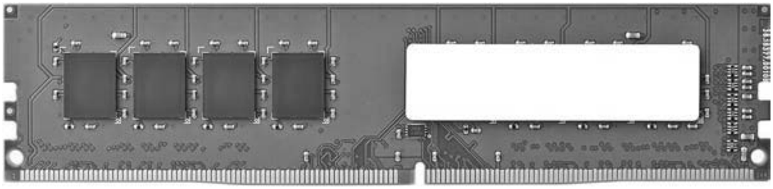


FIGURE 6.5 A typical 8GB DDR4 module without heat spreaders.

Table 6.11 JEDEC Standard DDR4 Module (260-Pin DIMM) Speeds and Transfer Rates

Module Type	Chip Type	Base Clock Speed	Cycle Time	Cycles per Clock	Bus Speed	Bus Width	Module Transfer Rate	Dual-Channel Transfer Rate
PC4-12800	DDR4-1600	800MHz	1.25ns	2	1,600MTps	8 bytes	12,800MBps	25,600MBps
PC4-14900	DDR4-1866	933MHz	1.07ns	2	1,866MTps	8 bytes	14,933MBps	29,866MBps
PC4-17000	DDR4-2133	1066MHz	0.94ns	2	2,133MTps	8 bytes	17,066MBps	34,133MBps
PC4-19200	DDR4-2400	1,200MHz	0.83ns	2	2,400MTps	8 bytes	19,200MBps	38,400MBps
PC4-21300	DDR4-2666	1,333MHz	0.75ns	2	2,666MTps	8 bytes	21,333MBps	42,666MBps
PC4-25600	DDR4-3200	1,600MHz	0.63ns	2	3,200MTps	8 bytes	25,600MBps	51,200MBps

DDR = Double data rate
MHz = Million cycles per second
MTps = Million transfers per second
MBps = Million bytes per second
ns = Nanoseconds (billionths of a second)

The topology of DDR4 is technically not a bus as was used in the DDR3 and earlier memory standards. DDR4 uses a point-to-point connection instead, where each channel in the memory controller is connected to a single module.

Typically, you can find DDR4 modules rated CL12 through CL16.

RDRAM

Rambus DRAM (RDRAM) was a proprietary (non-JEDEC) memory technology found mainly in certain Intel-based Pentium III and 4 systems from 2000 through 2002. Very few of these systems are still in use today.

For more information about RDRAM and RIMM modules, see Chapter 6, "Memory," in *Upgrading and Repairing PCs*, 19th Edition.

Memory Modules

Originally, PCs had memory installed via individual chips. They are often referred to as *dual inline package (DIP)* chips because of their physical designs. The original IBM XT and AT systems had 36 sockets on the motherboard for these individual chips, and more sockets could often be found on memory cards plugged into the bus slots. I remember spending hours populating boards with these chips, which was a tedious job.

Besides being a time-consuming and labor-intensive way to deal with memory, DIP chips had one notorious problem—they crept out of their sockets over time as the system went through thermal cycles. Every day, when you powered the system on and off, the system heated and cooled and the chips gradually walked their way out of the sockets—a phenomenon called *chip creep*. Eventually, good contact was lost and memory errors resulted. Fortunately, reseating all the chips back in their sockets usually rectified the problem, but that method was labor intensive if you had many systems to support.

The alternative to this at the time was to have the memory soldered into either the motherboard or an expansion card. This prevented the chips from creeping and made the connections more permanent, but it caused another problem. If a chip did go bad, you had to attempt desoldering the old one and resoldering a new one or resort to scrapping the motherboard or memory card on which the chip was installed. This was expensive and made memory troubleshooting difficult.

A chip was needed that was both soldered and removable, which was made possible by using memory modules instead of individual chips. Early modules had one row of electrical contacts and were called single inline memory modules (SIMMs), whereas later modules had two rows and were called dual inline memory modules (DIMMs) or Rambus inline memory modules (RIMMs). These small boards plug into special connectors on a motherboard or memory card. The individual memory chips are soldered to the module, so removing and replacing them is impossible. Instead, you must replace the entire module if any part of it fails. The module is treated as though it were one large memory chip.

Several types of SIMMs, DIMMs, and RIMMs have been commonly used in desktop systems. The various types are often described by their pin count, memory row width, or memory type.

SIMMs, for example, are available in two main physical types—30-pin (8 bits plus an option for 1 additional parity bit) and 72-pin (32 bits plus an option for 4 additional parity bits)—with various capacities and other specifications. The 30-pin SIMMs are physically smaller than the 72-pin versions, and either version can have chips on one or both sides. SIMMs were widely used from the late 1980s to the late 1990s but have become obsolete.

DIMMs are available in five main types. SDR (single data rate) DIMMs have 168 pins, one notch on either side, and two notches along the contact area. DDR DIMMs, on the other hand, have 184 pins, two notches on each side, and only one offset notch along the contact area. DDR2 and DDR3 DIMMs have 240 pins, two notches on each side, and one near the center of the contact area. DDR4 DIMMs have 288 pins, two notches on each side (the notches are more squared off than with previous DIMM designs), and one near the center of the contact area. All DIMMs are either 64 bits (non-ECC/parity) or 72 bits (data plus parity or error-correcting code [ECC]) wide. The main physical difference between SIMMs and DIMMs is that DIMMs have different signal pins on each side of the module, resulting in two rows of electrical contacts. That is why they are called dual inline memory modules, and why with only 1 inch of additional length, they have many more pins than a SIMM.

Note

There is confusion among users and even in the industry regarding the terms *single-sided* and *double-sided* with respect to memory modules. In truth, the single- or double-sided designation actually has nothing to do with whether chips are physically located on one or both sides of the module, and it has nothing to do with whether the module is a SIMM or DIMM (meaning whether the connection pins are single- or double-inline). Instead, the terms single-sided and double-sided indicate whether the module has one or two internal banks (called *ranks*) of memory chips installed.

A dual-rank DIMM module has two complete 64-bit wide banks of chips logically stacked so that the module is twice as deep (has twice as many 64-bit rows). In most (but not all) cases, this requires chips to be on both sides of the module; therefore, the term ***double-sided*** often indicates that a module has two ranks, even though the term is technically incorrect.

Single-rank modules (incorrectly referred to as single-sided) can also have chips physically mounted on both sides of the module, and dual-rank modules can have chips physically mounted on only one side. I recommend using the terms ***single rank*** or ***dual rank*** instead, because they are much more accurate and easily understood.

Figures 6.6 through 6.13 show a typical 30-pin (8-bit) SIMM, 72-pin (32-bit) SIMM, 168-pin SDRAM DIMM, 184-pin DDR SDRAM (64-bit) DIMM, 240-pin DDR2 DIMM, 240-pin DDR3 DIMM, 288-pin DDR4 DIMM, and 184-pin RIMM, respectively. The pins are numbered from left to right and are connected to both sides of the module on the SIMMs. The pins on the DIMM are different on each side, but on a SIMM, each side is the same as the other and the connections carry through. Note that all dimensions are in both inches and millimeters (in parentheses), and modules are generally available in ECC versions with 1 extra ECC (or parity) bit for every 8 data bits (multiples of 9 in data width) or versions that do not include ECC support (multiples of 8 in data width).

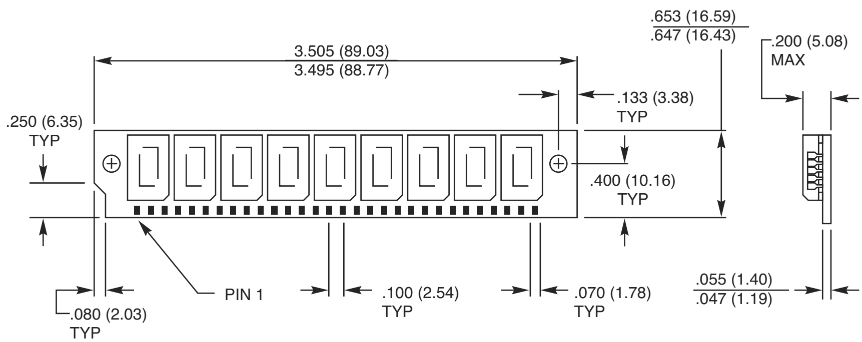


FIGURE 6.6 A typical 30-pin SIMM.

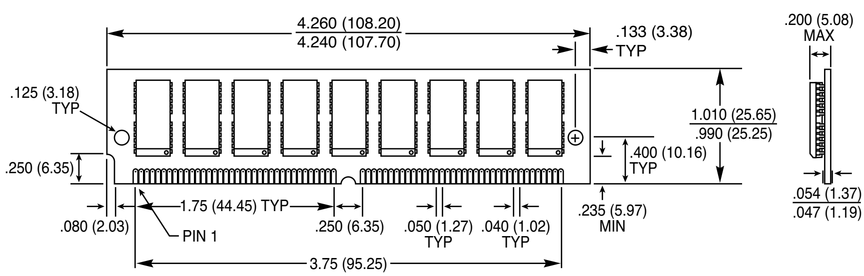


FIGURE 6.7 A typical 72-pin SIMM.

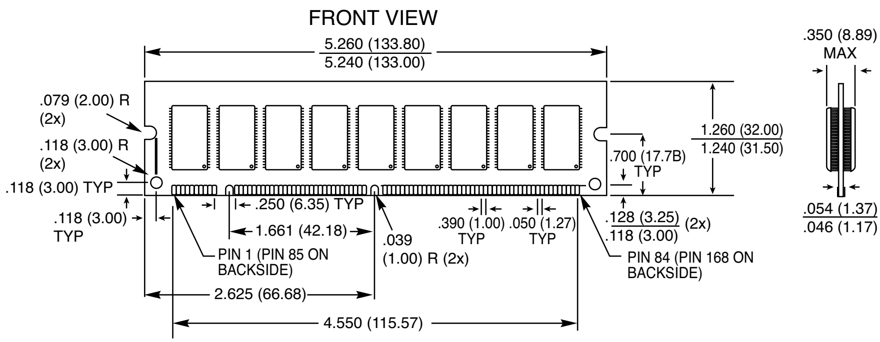


FIGURE 6.8 A typical 168-pin SDRAM DIMM.

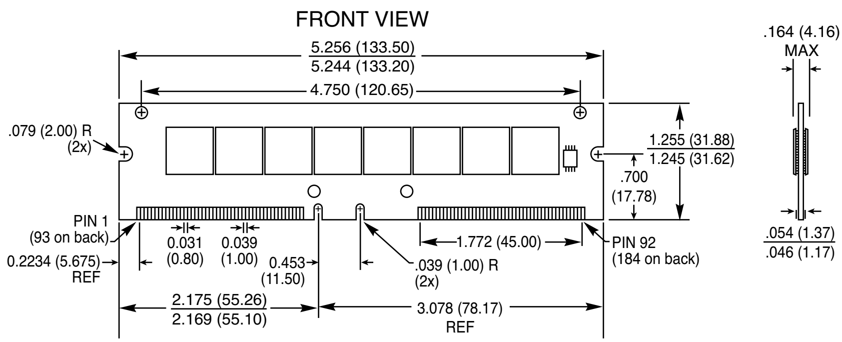


FIGURE 6.9 A typical 184-pin DDR DIMM.

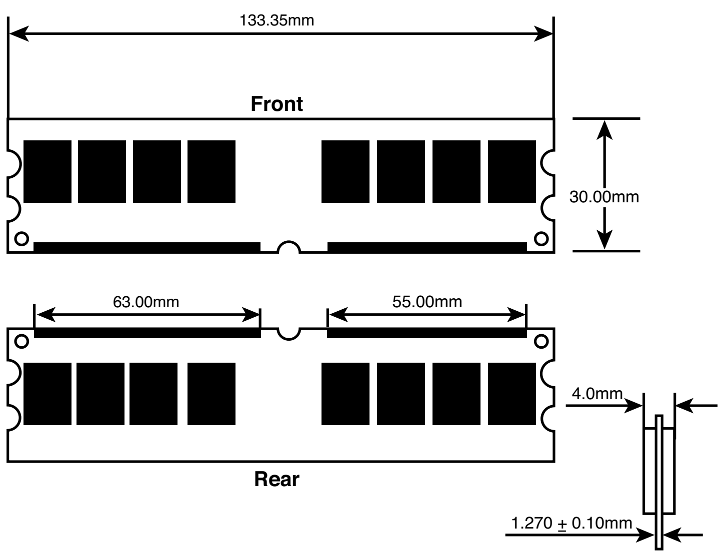


FIGURE 6.10 A typical 240-pin DDR2 DIMM.

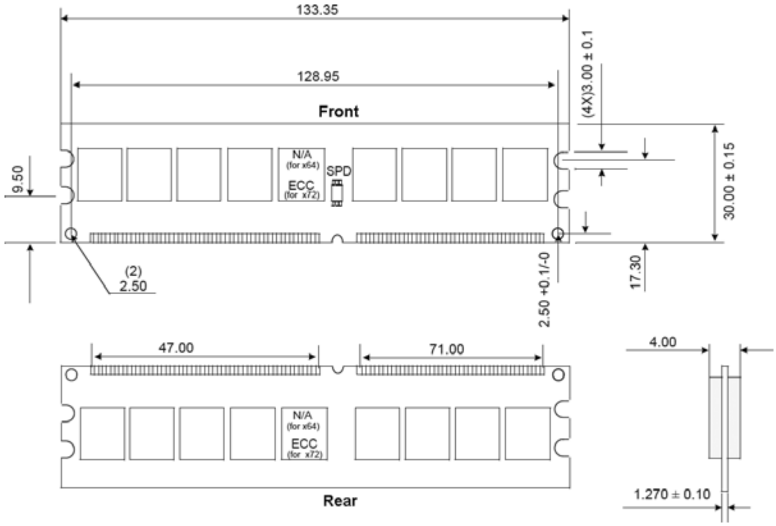


FIGURE 6.11 A typical 240-pin DDR3 DIMM.

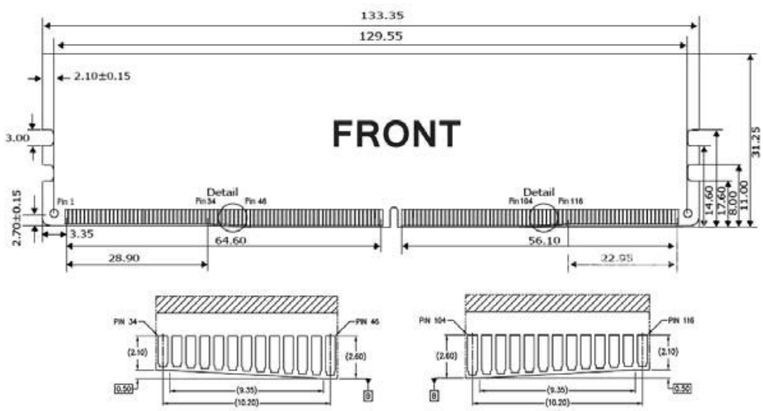


FIGURE 6.12 A typical 288-pin DDR4 DIMM.

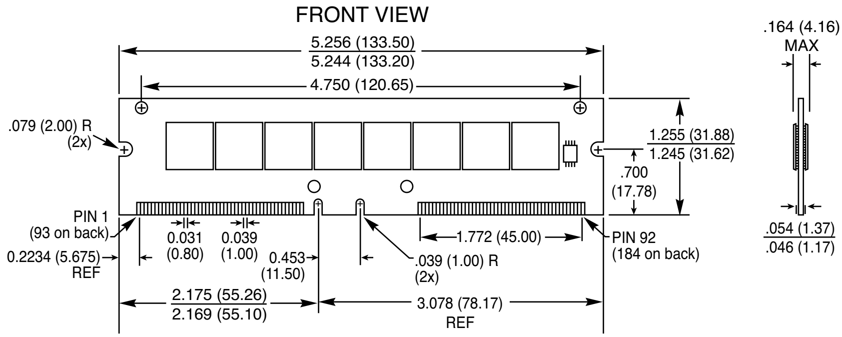


FIGURE 6.13 A typical 184-pin RIMM (RIMM modules use RDRAM).

All these memory modules are fairly compact considering the amount of memory they hold and are available in several capacities and speeds. Table 6.12 lists the various capacities available for SIMMs, DIMMs, and RIMMs.

Table 6.12 SIMM, DIMM, and RIMM Capacities

Capacity	Standard Depth×Width	Parity/ECC Depth×Width
30-Pin SIMM		
256KB	256K×8	256K×9
1MB	1M×8	1M×9
4MB	4M×8	4M×9
16MB	16M×8	16M×9
72-Pin SIMM		
1MB	256K×32	256K×36
2MB	512K×32	512K×36
4MB	1M×32	1M×36
8MB	2M×32	2M×36
16MB	4M×32	4M×36
32MB	8M×32	8M×36
64MB	16M×32	16M×36
128MB	32M×32	32M×36
168/184-Pin DIMM/DDR DIMM		
8MB	1M×64	1M×72
16MB	2M×64	2M×72
32MB	4M×64	4M×72
64MB	8M×64	8M×72

Capacity	Standard Depth×Width	Parity/ECC Depth×Width
128MB	16M×64	16M×72
256MB	32M×64	32M×72
512MB	64M×64	64M×72
1,024MB	128M×64	128M×72
2,048MB	256M×64	256M×72
240-Pin DDR2/DDR3 DIMM		
256MB	32M×64	32M×72
512MB	64M×64	64M×72
1,024MB	128M×64	128M×72
2,048MB	256M×64	256M×72
4,096MB	512M×64	512M×72
8,192MB	1,024M×64	1,024M×72
*288-Pin DDR4 DIMM**		
4,096MB	512M×64	512M×72
8,192MB	1,024M×64	1,024M×72
184-Pin RIMM		
64MB	32M×16	32M×18
128MB	64M×16	64M×18
256MB	128M×16	128M×18
512MB	256M×16	256M×18
1,024MB	512M×16	512M×18

**Higher capacities are available for servers*

Memory modules of each type and capacity are available in various speed ratings. Consult your motherboard documentation for the correct memory speed and type for your system. If a system requires a specific speed memory module, you can almost always substitute faster speeds if the one specified is not available. Generally, no problems occur in mixing module speeds, as long as you use modules equal to or faster than what the system requires. Because there's little price difference between the various speed versions, I often buy faster modules than are necessary for a particular application, especially if they are the same price as slower modules. This might make them more usable in a future system that could require the faster speed.

Because SDRAM and newer modules have an onboard SPD ROM that reports their speed and timing parameters to the system, most systems run the memory controller and memory bus at the speed matching the slowest module installed.

Note

A bank is the smallest amount of memory needed to form a single row of memory addressable by the processor. It is the minimum amount of physical memory that the processor reads or writes at one time and usually corresponds to the data bus width of the processor. If a processor has a 64-bit data bus, a bank of memory also is 64 bits wide. If the memory runs dual- or tri-channel, a virtual bank is formed that is two or three times the absolute data bus width of the processor.

You can't always replace a module with a higher-capacity unit and expect it to work. Systems might have specific design limitations for the maximum capacity of module they can take. A larger-capacity module works only if the motherboard is designed to accept it in the first place. Consult your system documentation to determine the correct capacity and speed to use. With some systems, a BIOS update might enable the use of higher-capacity and/or faster modules than the system was originally designed to use. Check with the system vendor to see whether a BIOS update is available.

Registered Modules

SDRAM through DDR4 modules are available in unbuffered and registered versions. Most PC motherboards are designed to use unbuffered modules, which enable the memory controller signals to pass directly to the memory chips on the module with no interference. This is not only the cheapest design, but also the fastest and most efficient. The only drawback is that the motherboard designer must place limits on how many modules (meaning module sockets) can be installed on the board, and it could limit how many chips can be on a module. So-called double-sided modules that really have multiple banks of chips onboard might be restricted on some systems in certain combinations.

Systems designed to accept extremely large amounts of RAM (such as servers) often require registered modules. A registered module uses an architecture that has register chips on the module that act as an interface between the actual RAM chips and the chipset. The registers temporarily hold data passing to and from the memory chips and enable many more RAM chips to be driven or otherwise placed on the module than the chipset could normally support. This enables motherboard designs that can support many modules and enables each module to have a larger number of chips. In general, registered modules are required by server or workstation motherboards designed to support more than four sockets. One anomaly is the initial version of the AMD Athlon 64 FX processor, which also uses registered memory because its Socket 940 design was based on the AMD Opteron workstation and server processor. Subsequent Socket 939, AM2, AM2+, AM3, and AM3+ versions of the Athlon FX no longer require registered memory.

To provide the space needed for the buffer chips, a registered DIMM is often taller than a standard DIMM. Figure 6.14 compares a typical registered DIMM to a typical unbuffered DIMM.

Tip

If you are installing registered DIMMs in a slimline case, clearance between the top of the DIMM and the case might be a problem. Some vendors sell low-profile registered DIMMs that are about the same height as unbuffered DIMMs. Use this type of DIMM if your system does not have enough head room for standard registered DIMMs. Some vendors sell only this type of DIMM for particular systems.

The important thing to note is that you can use only the type of module your motherboard (or chipset) is designed to support. For most, that is standard unbuffered modules or, in some cases, registered modules.

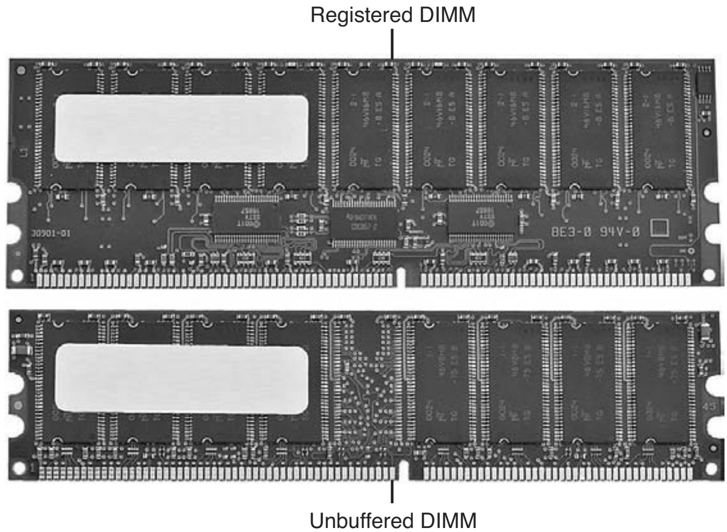


FIGURE 6.14 A typical registered DIMM is taller than a typical unbuffered DIMM to provide room for buffer chips.

SDR DIMM Details

SDR DIMMs use a completely different type of presence detect than SIMMs, called *serial presence detect (SPD)*. It consists of a small EEPROM or flash memory chip on the DIMM that contains specially formatted data indicating the DIMM's features. This serial data can be read via the serial data pins on the DIMM, and it enables the motherboard to autoconfigure to the exact type of DIMM installed.

DIMMs can come in several varieties, including unbuffered and buffered as well as 3.3V and 5V. Buffered DIMMs have additional buffer chips on them to interface to the motherboard. Unfortunately, these buffer chips slow down the DIMM and are not effective at higher speeds. For this reason, PC systems (those that do not use registered DIMMs) use only unbuffered 3.5V DIMMs. The voltage is simple—DIMM designs for PCs are almost universally 3.3V. If you install a 5V DIMM in a 3.3V socket, it would be damaged, but keying in the socket and on the DIMM prevents that.

Apple and other non-PC systems can use the buffered 5V versions. Fortunately, the key notches along the connector edge of a DIMM are spaced differently for buffered/unbuffered and 3.3V/5V DIMMs, as shown in Figure 6.15. This prevents inserting a DIMM of the wrong type into a given socket.

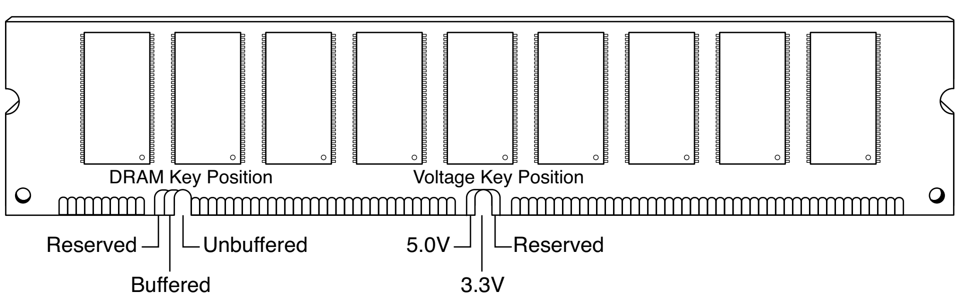


FIGURE 6.15 The 168-pin DRAM DIMM notch key definitions.

DDR DIMM Details

The 184-pin DDR DIMMs use a single key notch to indicate voltage, as shown in Figure 6.16.

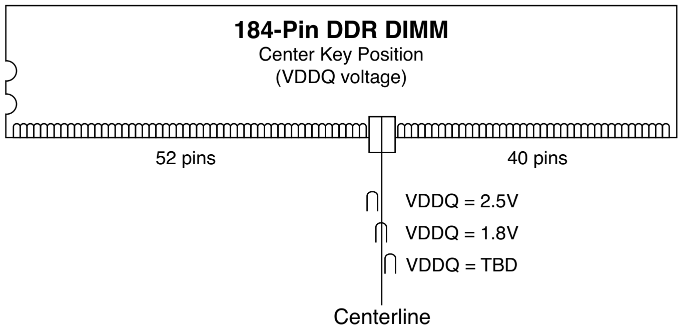


FIGURE 6.16 The 184-pin DDR SDRAM DIMM keying.

DDR DIMMs also use two notches on each side to enable compatibility with both low- and high-profile latched sockets. Note that the key position is offset with respect to the center of the DIMM to prevent inserting it backward in the socket. The key notch is positioned to the left, centered, or to the right of the area between pins 52 and 53. This indicates the I/O voltage for the DDR DIMM and prevents installing the wrong type into a socket that might damage the DIMM.

DDR2 DIMM Details

The 240-pin DDR2 DIMMs use two notches on each side to enable compatibility with both low- and high-profile latched sockets. The connector key is offset with respect to the center of the DIMM to prevent inserting it backward in the socket. The key notch is positioned in the center of the area between pins 64 and 65 on the front (184/185 on the back), and there is no voltage keying because all DDR2 DIMMs run on 1.8V.

DDR3 DIMM Details

The 240-pin DDR3 DIMMs use two notches on each side to enable compatibility with both low- and high-profile latched sockets. The connector key is offset with respect to the center of the DIMM to prevent inserting it backward in the socket. The key notch is positioned in the center of the area between pins 48 and 49 on the front (168/169 on the back), and there is no voltage keying because all DDR3 DIMMs run on 1.5V.

DDR4 DIMM Details

The 288-pin DDR4 DIMMs use two notches on each side to enable compatibility with both low- and high-profile latched sockets. The notches use a more squared-off shape than with previous designs. The connector key is offset with respect to the center of the DIMM to prevent inserting it backward in the socket. The key notch is positioned in the center of the area between pins 77 and 78 on the front, and there is no voltage keying because all DDR4 DIMMs run on 1.2V.

Determining a Memory Module's Size and Features

Most memory modules are labeled with a sticker indicating the module's type, speed rating, and manufacturer. If you are attempting to determine whether existing memory can be used in a new computer, or if you need to replace memory in an existing computer, this information can be essential. Figure 6.17 compares the markings on a 512MB DDR2 module from Crucial Technology and a 2GB DDR2 module from Kingston Technology.

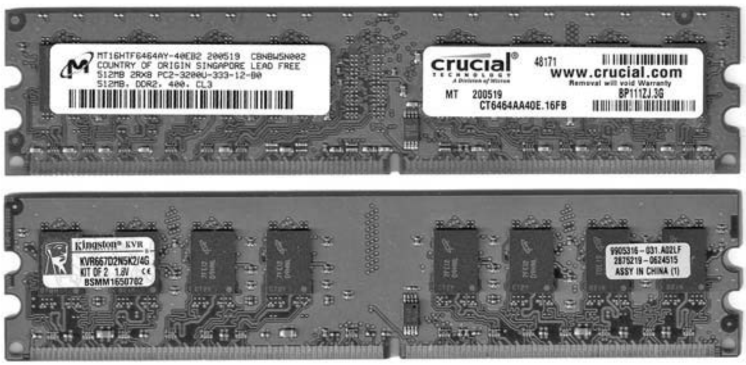


FIGURE 6.17 Markings on a 512MB DDR2 module (top) from Crucial Technology compared to markings on a 2GB (bottom) DDR2 memory module from Kingston Technology.

However, if you have memory modules that are not labeled, you can still determine the module type, speed, and capacity if the memory chips on the module are clearly labeled. For example, assume you have a memory module with chips labeled as follows:

MT46V64M8TG-75

By using an Internet search engine such as Google and entering the number from one of the memory chips, you can usually find the data sheet for the memory chips. Consider the following example: Say you have a registered memory module and want to look up the part number for the memory chips (usually eight or more chips) rather than the buffer chips on the module (usually from one to three, depending on the module design). In this example, the part number turns out to be a Micron memory chip that decodes like this:

MT = Micron Technologies (the memory chip maker)

46 = DDR SDRAM

V = 2.5V DC

64M8 = 8 million rows × 8 (equals 64) × 8 banks (often written as 64 Meg × 8)

TG = 66-pin TSOP chip package

–75 = 7.5ns @ CL2 latency (DDR 266)

The full datasheet for this example is located at http://download.micron.com/pdf/datasheets/dram/ddr/512MBDDRx4x8x16.pdf.

From this information, you can determine that the module has the following characteristics:

- The module runs at DDR266 speeds using standard 2.5V DC voltage.
- The module has a latency of CL2, so you can use it on any system that requires CL2 or slower latency (such as CL2.5 or CL3).
- Each chip has a capacity of 512Mb (64 × 8 = 512).
- Each chip contains 8 bits. Because it takes 8 bits to make 1 byte, you can calculate the capacity of the module by grouping the memory chips on the module into groups of eight. If each chip contains 512Mb, a group of eight means that the module has a size of 512MB (512Mb × 8 = 512MB). A dual-bank module has two groups of eight chips for a capacity of 1GB (512Mb × 8 = 1024MB, or 1GB).

If the module has nine instead of eight memory chips (or 18 instead of 16), the additional chips are used for parity checking and support ECC error correction on servers with this feature.

To determine the size of the module in megabytes or gigabytes and to determine whether the module supports ECC, count the memory chips on the module and compare them to Table 6.13. Note that the size of each memory chip in megabits is the same as the size in megabytes if the memory chips use an 8-bit design.

Table 6.13 Module Capacity Using 512Mb (64Mbit × 8) Chips

Number of Chips	Number of Bits in Each Bank	Module Size	Supports ECC?	Single or Dual Bank
8	64	512MB	No	Single
9	72	512MB	Yes	Single
16	64	1GB	No	Dual
18	72	1GB	Yes	Dual

The additional chip that each group of eight chips uses provides parity checking, which the ECC function employs on most server motherboards to correct single-bit errors.

A registered module contains 9 or 18 memory chips for ECC plus additional memory buffer chips. These chips are usually smaller in size and located near the center of the module, as shown previously in Figure 6.10.

Note

Some modules use 16-bit wide memory chips. In such cases, only 4 chips are needed for single-bank memory (5 with parity/ECC support), and 8 are needed for double-bank memory (10 with parity/ECC support). These memory chips use a design listed as capacity times 16, like this: 256Mb × 16.

You can also see this information if you look up the manufacturer and the memory type in a search engine. For example, a web search for "Micron Unbuffered DIMM Design" locates a table showing various DIMM organization, SDRAM density, and other information for listed modules.

As you can see, with a little detective work, you can determine the size, speed, and type of a memory module—even if the module isn't marked—as long as the markings on the memory chips themselves are legible.

Tip

If you are unable to decipher a chip part number, you can use a program, such as CPU-Z (www.cpuid.com) or HWiNFO (www.hwinfo.com) to identify your memory module, as well as many other facts about your computer, including chipset, processor, empty memory sockets, and much more.

Memory Banks

Memory chips (DIPs, SIMMs, SIPPs, and DIMMs) are organized in banks on motherboards and memory cards. You should know the memory bank layout and position on the motherboard and memory cards.

You need to know the bank layout when adding memory to the system. In addition, memory diagnostics report error locations by byte and bit addresses, and you must use these numbers to locate which bank in your system contains the problem.

The banks usually correspond to the data bus capacity of the system's microprocessor. Table 6.14 shows the widths of individual banks based on the type of PC.

Table 6.14 Memory Bank Widths on Various Systems

Processor	Data Bus	Memory Bank Width	Memory Bank Width (Parity/ECC)	30-Pin SIMMs per Bank	72-Pin SIMMs per Bank	DIMMs per Bank
8088	8-bit	8 bits	9 bits	1	—	—
8086	16-bit	16 bits	18 bits	2	—	—
286	16-bit	16 bits	18 bits	2	—	—
386SX, SL, SLC	16-bit	16 bits	18 bits	2	—	—
486SLC, SLC2	16-bit	16 bits	18 bits	2	—	—
386DX	32-bit	32 bits	36 bits	4	1	—
486SX, DX, DX2, DX4, 5x86	32-bit	32 bits	36 bits	4	1	—
x86 and x86-64 running single-channel mode	64-bit	64 bits	72 bits	—	—	1
x86 and x86-64 running dual-channel mode	64-bit	128 bits	144 bits	—	—	2
x86 and x86-64 running tri-channel mode	64-bit	192 bits	216 bits	—	—	3

DIMMs are ideal for Pentium (and higher) systems because the 64-bit width of the DIMM exactly matches the 64-bit width of the Pentium processor data bus. Therefore, each DIMM represents an individual bank, and these can be added or removed one at a time. Many recent systems have been designed to use matched pairs or triples of memory modules for faster performance. So-called “dual-channel” and “tri-channel” designs treat two or three matched modules as a single bank of memory.

The physical orientation and numbering of the SIMMs or DIMMs used on a motherboard are arbitrary and determined by the board’s designers, so documentation covering your system or card comes in handy. You can determine the layout of a motherboard or an adapter card through testing, but that takes time and might be difficult, particularly after you have a problem with a system.

Caution

If your system supports dual- or tri-channel memory, be sure you use the correct memory sockets to enable multichannel operation. Check the documentation to ensure that you use the correct sockets. Most multichannel systems still run in single-channel mode if the memory is not installed in a way that permits full multichannel operation, but performance is lower than if the memory were installed properly. Some systems provide dual-channel support if an odd number of modules is installed, as long as the total capacity of two modules installed in one channel equals the size of the single module in the other channel and all modules are the same speed and latency. Again, check your documentation for details.

Memory Module Speed

When you replace a failed memory module or install a new module as an upgrade, you typically must install a module of the same type and speed as the others in the system. You can substitute a module with a different (faster) speed, but only if the replacement module’s speed is equal to or faster than that of the other modules in the system.

Some people have had problems when “mixing” modules of different speeds. With the variety of motherboards, chipsets, and memory types, few ironclad rules exist. When in doubt as to which speed module to install in your system, consult the motherboard documentation for more information.

Substituting faster memory of the same type doesn’t result in improved performance if the system still operates the memory at the same speed. Systems that use DIMMs or RIMMs can read the speed and timing features of the module from a special SPD ROM installed on the module and then set chipset (memory controller) timing accordingly. In these systems, you might see an increase in performance by installing faster modules, to the limit of what the chipset supports.

To place more emphasis on timing and reliability, some Intel and JEDEC standards governing memory types require certain levels of performance. These standards certify that memory modules perform within Intel’s timing and performance guidelines.

The same common symptoms result when the system memory has failed or is simply not fast enough for the system’s timing. The usual symptoms are frequent parity check errors or a system that does not operate. The POST might report errors, too. If you’re unsure of which chips to buy for your system, contact the system manufacturer or a reputable chip supplier.

▶▶ See “Parity Checking,” **p. 390**.

Parity and ECC

Part of the nature of memory is that it inevitably fails. These failures are usually classified as two basic types: hard fails and soft errors.

The best understood are hard fails, in which the chip is working and then, because of some flaw, physical damage, or other event, becomes damaged and experiences a permanent failure. Fixing this type of failure normally requires replacing some part of the memory hardware, such as the chip, SIMM, or DIMM. Hard error rates are known as HERs.

The other, more insidious type of failure is the soft error, which is a nonpermanent failure that might never recur or could occur only at infrequent intervals. Soft error rates are known as SERs.

In the late 1970s, Intel made a discovery about soft errors that shook the memory industry. It found that alpha particles were causing an unacceptably high rate of soft errors or single event upsets (SEUs) in the 16KB DRAMs that were available at the time. Because alpha particles are low-energy particles that can be stopped by something as thin and light as a sheet of paper, it became clear that for alpha particles to cause a DRAM soft error, they would have to be coming from within the semiconductor material. Testing showed trace elements of thorium and uranium in the plastic and ceramic chip packaging materials used at the time. This discovery forced all the memory manufacturers to evaluate their manufacturing processes to produce materials free from contamination.

Today, memory manufacturers have all but totally eliminated the alpha-particle source of soft errors, and more recent discoveries prove that alpha particles are now only a small fraction of the cause of DRAM soft errors.

As it turns out, the biggest cause of soft errors today is cosmic rays. IBM researchers began investigating the potential of terrestrial cosmic rays in causing soft errors similar to alpha particles. The difference is that cosmic rays are high-energy particles and can't be stopped by sheets of paper or other more powerful types of shielding. The leader in this line of investigation was Dr. J.F. Ziegler of the IBM Watson Research Center in Yorktown Heights, New York. He has produced landmark research into understanding cosmic rays and their influence on soft errors in memory. One interesting set of experiments found that cosmic ray–induced soft errors were eliminated when the DRAMs were moved to an underground vault shielded by more than 50 feet of rock.

Cosmic ray–induced errors are even more of a problem in SRAMs than DRAMS because the amount of charge required to flip a bit in an SRAM cell is less than is required to flip a DRAM cell capacitor. Cosmic rays are also more of a problem for higher-density memory. As chip density increases, it becomes easier for a stray particle to flip a bit. It has been predicted by some that the soft error rate of a 64MB DRAM is double that of a 16MB chip, and a 256MB DRAM has a rate four times higher. As memory sizes continue to increase, it's likely that soft error rates will also increase.

Unfortunately, the PC industry has largely failed to recognize this cause of memory errors. Electrostatic discharge, power surges, and unstable software can much more easily explain away the random and intermittent nature of a soft error, especially right after a new release of an operating system (OS) or major application.

Although cosmic rays and other radiation events are perhaps the biggest cause of soft errors, soft errors can also be caused by the following:

- **Power glitches or noise on the line**—This can be caused by a defective power supply in the system or by defective power at the outlet.
- **Incorrect type or speed rating**—The memory must be the correct type for the chipset and match the system access speed.
- **RF (radio frequency) interference**—Caused by radio transmitters in close proximity to the system, which can generate electrical signals in system wiring and circuits. Keep in mind that the increased use of wireless networks, keyboards, and mouse devices can lead to a greater risk of RF interference.

- **Static discharges**—These discharges cause momentary power spikes, which alter data.
- **Timing glitches**—Data doesn't arrive at the proper place at the proper time, causing errors. Often caused by improper settings in the BIOS Setup, by memory that is rated slower than the system requires, or by overclocked processors and other system components.
- **Heat buildup**—High-speed memory modules run hotter than older modules. RDRAM RIMM modules were the first memory to include integrated heat spreaders, and many high-performance DDR, DDR2, DDR3, and DDR4 memory modules now include heat spreaders to help fight heat buildup.

Most of these problems don't cause chips to permanently fail (although bad power or static can damage chips permanently), but they can cause momentary problems with data.

How can you deal with these errors? The best way to deal with this problem is to increase the system's fault tolerance. This means implementing ways of detecting and possibly correcting errors in PC systems. Three basic levels and techniques are used for fault tolerance in modern PCs:

- Nonparity
- Parity
- ECC

Nonparity systems have no fault tolerance. The only reason they are used is because they have the lowest inherent cost. No additional memory is necessary, as is the case with parity or ECC techniques. Because a parity-type data byte has 9 bits versus 8 for nonparity, memory cost is approximately 12.5% higher. Also, the nonparity memory controller is simplified because it does not need the logic gates to calculate parity or ECC check bits. Portable systems that place a premium on minimizing power might benefit from the reduction in memory power resulting from fewer DRAM chips. Finally, the memory system data bus is narrower, which reduces the number of data buffers. The statistical probability of memory failures in a modern office desktop computer is now estimated at about one error every few months. Errors will be more or less frequent depending on how much memory you have.

This error rate might be tolerable for low-end systems that are not used for mission-critical applications. In this case, the extreme market sensitivity to price probably can't justify the extra cost of parity or ECC memory, and such errors then must be tolerated.

Parity Checking

One standard IBM set for the industry is that the memory chips in a bank of nine each handle 1 bit of data: 8 bits per character plus 1 extra bit called the *parity bit*. The parity bit enables memory-control circuitry to keep tabs on the other 8 bits—a built-in cross-check for the integrity of each byte in the system.

Originally, all PC systems used parity-checked memory to ensure accuracy. Starting in 1994, most vendors began shipping systems without parity checking or any other means of detecting or correcting errors on-the-fly. These systems used cheaper nonparity memory modules, which saved about 10%–15% on memory costs for a system.

Parity memory results in increased initial system cost, primarily because of the additional memory bits involved. Parity can't correct system errors, but because parity can detect errors, it can make the user aware of memory errors when they happen.

Since then, Intel, AMD, and other manufacturers have put support for ECC memory primarily in server chipsets and processors. Chipsets and processors for standard desktop or laptop systems typically lack support for either parity or ECC.

How Parity Checking Works

IBM originally established the odd parity standard for error checking. The following explanation might help you understand what is meant by odd parity. As the 8 individual bits in a byte are stored in memory, a parity generator/checker, which is either part of the CPU or located in a special chip on the motherboard, evaluates the data bits by adding up the number of 1s in the byte. If an even number of 1s is found, the parity generator/checker creates a 1 and stores it as the ninth bit (parity bit) in the parity memory chip. That makes the sum for all 9 bits (including the parity bit) an odd number. If the original sum of the 8 data bits is an odd number, the parity bit created would be a 0, keeping the sum for all 9 bits an odd number. The basic rule is that the value of the parity bit is always chosen so that the sum of all 9 bits (8 data bits plus 1 parity bit) is stored as an odd number. If the system used even parity, the example would be the same except the parity bit would be created to ensure an even sum. It doesn't matter whether even or odd parity is used; the system uses one or the other, and it is completely transparent to the memory chips involved. Remember that the 8 data bits in a byte are numbered 0 1 2 3 4 5 6 7. The following examples might make it easier to understand:

```
Data bit number:      0 1 2 3 4 5 6 7     Parity bit

Data bit value:       1 0 1 1 0 0 1 1     0
```

In this example, because the total number of data bits with a value of 1 is an odd number (5), the parity bit must have a value of 0 to ensure an odd sum for all 9 bits.

Here is another example:

```
Data bit number:      0 1 2 3 4 5 6 7     Parity bit

Data bit value:       1 1 1 1 0 0 1 1     1
```

In this example, because the total number of data bits with a value of 1 is an even number (6), the parity bit must have a value of 1 to create an odd sum for all 9 bits.

When the system reads memory back from storage, it checks the parity information. If a (9-bit) byte has an even number of bits, that byte must have an error. The system can't tell which bit has changed or whether only a single bit has changed. If 3 bits changed, for example, the byte still flags a parity-check error; if 2 bits changed, however, the bad byte could pass unnoticed. Because multiple bit errors (in a single byte) are rare, this scheme gives you a reasonable and inexpensive ongoing indication that memory is good or bad.

ECC

ECC goes a big step beyond simple parity-error detection. Instead of just detecting an error, ECC allows a single bit error to be corrected, which means the system can continue without interruption and without corrupting data. ECC, as implemented in most PCs, can only detect, not correct, double-bit errors. Because studies have indicated that approximately 98% of memory errors are the single-bit variety, the most commonly used type of ECC is one in which the attendant memory controller detects and corrects single-bit errors in an accessed data word. (Double-bit errors can be detected but not corrected.) This type of ECC is known as *single-bit error-correction double-bit error detection (SEC-DED)* and requires an additional 7 check bits over 32 bits in a 4-byte system and an additional 8 check bits over 64 bits in an 8-byte system. If the system uses SIMMs, two 36-bit (parity) SIMMs are added for each bank (for a total of 72 bits), and ECC is done at the bank level. If the system uses DIMMs, a single parity/ECC 72-bit DIMM is used as a bank and provides the additional bits. RIMMs are installed in singles or pairs, depending on the chipset and motherboard. They must be 18-bit versions if parity/ECC is desired.

ECC entails the memory controller calculating the check bits on a memory-write operation, performing a compare between the read and calculated check bits on a read operation, and, if necessary, correcting bad bits. The additional ECC logic in the memory controller is not very significant in this age of inexpensive, high-performance VLSI logic, but ECC actually affects memory performance on writes. This is because the operation must be timed to wait for the calculation of check bits and, when the system waits for corrected data, reads. On a partial-word write, the entire word must first be read, the affected byte(s) rewritten, and then new check bits calculated. This turns partial-word write operations into slower read-modify writes. Fortunately, this performance hit is small, on the order of a few percent at maximum, so the trade-off for increased reliability is a good one.

Most memory errors are of a single-bit nature, which ECC can correct. Incorporating this fault-tolerant technique provides high system reliability and attendant availability. An ECC-based system is a good choice for servers, workstations, or mission-critical applications in which the cost of a potential memory error outweighs the additional memory and system cost to correct it, along with ensuring that it does not detract from system reliability.

Unfortunately, most standard desktop and laptop PC processors, motherboards (chipsets), and memory modules don't support ECC. If you want a system that supports ECC, make sure all the components involved support ECC. This usually means purchasing more expensive processors, motherboards, and RAM designed for server or high-end workstation applications.

RAM Upgrades

Adding memory to a system is one of the most useful upgrades you can perform and also one of the least expensive—especially when you consider the increased performance of Windows, Linux, and their applications when you give them access to more memory. In some cases, doubling the memory can practically double the speed of a computer. But it doesn't always pay to go overboard because adding memory you don't really need will cost money and power, and you will gain little or nothing in speed. The best philosophy to take when adding RAM to a computer is that "more is better, up to a point."

The maximum physical memory capacity of a system is dictated by several factors. The first is the amount addressable by the processor itself, which is based on the number of physical address lines in the chip. The original PC processors (8086/8088) had 20 address lines, which resulted in those chips being able to recognize up to 1MB (2^{20} bytes) of RAM. The 286/386SX increased memory addressing capability to 24 lines, making them capable of addressing 16MB (2^{24} bytes). Modern x86 processors have 32– 36 address lines, resulting in from 4GB to 64GB of addressable RAM. Modern x86-64 (64-bit) processors have 40 address lines, resulting in a maximum of 1TB (1 terabyte) of supported physical RAM.

◄◄ See the Chapter 3 section "Processor Specifications," **p. 40**.

The operating mode of the processor may place further limits on memory addressability. For example, when the processor is operating in backward-compatible real mode, only 1MB of memory is supported.

◄◄ See the Chapter 3 section "Processor Modes," **p. 49**.

Note that even though modern 64-bit processors can technically address up to 1TB, modern motherboards or chipsets generally limit the maximum amount of RAM to 8GB, 16GB, 32GB, or 64GB. The type of software also has an effect. The 32-bit versions of Windows (XP and newer) limit memory support to 4GB (with only about 3.5GB usable by programs), whereas the 64-bit versions limit support to 8GB, 16GB, or 192GB, depending on the edition.

◄◄ See the Chapter 4 section "Chipsets," **p. 192**, for the memory limits on motherboard chipsets.

If you run Windows XP, you should specify a *minimum* of 256MB, and preferably 512MB–1GB or more depending on the applications you intend to run. If you run 32-bit Windows 7/8.1/10 or Vista, the *absolute minimum* should be 512MB according to Microsoft, but I recommend a minimum of 1GB, with 2GB–3GB preferred. 64-bit versions of Windows 7 and Windows 8/8.1/10 have a 2GB minimum, but perform better with 4GB or more of memory.

Beyond having the minimum to run the OS you choose, the way you use your system, especially the applications you run, can be the major determining factor as to just how much memory is best. For example, if you are a power user with four or more displays simultaneously connected to your system, each with multiple open applications, or you run memory-intensive applications such as photo- and video-editing programs, or if you use a virtual machine manager (VMM) like Virtual PC or VMware to run multiple OSs simultaneously (each of those with open applications), you might want as much memory as you can possibly install. Many older laptops won't accept as much memory as you might want (or need) to install, so if you upgrade an older system that uses an obsolete (and expensive) type of memory, the best tip might be to consider moving up to a newer system that can accept more memory of a mainstream type that is less expensive.

When purchasing a new system, try to get it with all the memory you need right away. Some motherboards are more limited in the number of memory sockets they contain, and some of those will already be filled when the system is delivered. This means you might need to remove some of the existing memory to add more, which makes future upgrades more expensive. The only caveat here is that I often find that I can purchase memory much more inexpensively from third-party vendors than from the system manufacturer. When purchasing a new system, check on how much the manufacturer charges for the amount of memory you want, as opposed to taking the system with the default minimum and immediately adding the desired memory yourself, purchased from a third-party memory vendor.

The following sections discuss adding memory, including selecting memory chips, installing memory chips, and testing the installation.

Upgrade Options and Strategies

Adding memory can be an inexpensive solution; the cost of mainstream memory is extremely low relative to other system components, and adding more memory can give your computer's performance a big boost.

How do you add memory to your PC? You have two options, listed in order of convenience and cost:

- Adding memory in vacant slots on your motherboard
- Replacing your current motherboard's memory with higher-capacity memory

If you decide to upgrade to a more powerful computer system or motherboard, you usually can't salvage the memory from your previous system. Most of the time it is best to plan on equipping a new board with the optimum type of memory that it supports.

Be sure to carefully weigh your future needs for computing speed and a multitasking OS against the amount of money you spend to upgrade current equipment.

How can you tell if you have enough memory or not? The best way is to run your most demanding applications (all that would be open at the same time and with your largest datasets) and then check the memory usage using the Windows Task Manager (taskmgr.exe). With Task Manager running, click the Performance tab to see the amount of Physical Memory being used versus the total available. Under Windows 7/8.1/10 and Vista the memory usage is shown both as a percentage of the total as well as an amount.

Figure 6.18 shows the Task Manager running on a Windows 7 system reporting 18% or 2.85GB of memory being used on a system with 16GB installed. Figure 6.19 shows the Task Manager running on a Windows XP system reporting 1.63GB being used (under the misnomer "PF Usage") on a system with about 3.5GB of RAM available (4GB minus about 0.5GB reserved by the hardware), which is about 47% of the total.

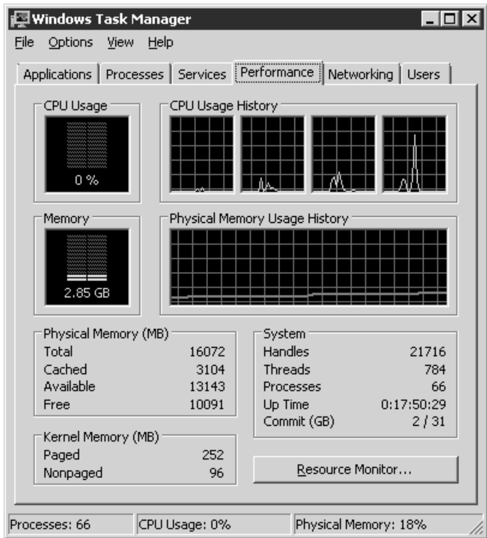


FIGURE 6.18 Windows 7 Task Manager showing 18% or 2.85GB Physical Memory use on a system with 16GB installed.

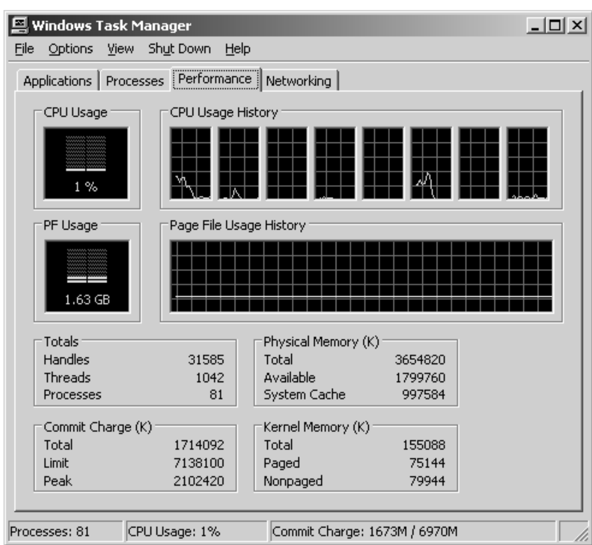


FIGURE 6.19 Windows XP Task Manager showing 47% or 1.63GB Physical Memory use on a system with about 3.5GB available (out of 4GB installed).

If you see the amount of physical memory being used is higher than around 80% of the total then you might consider adding more memory. In both of these examples the amount of memory installed is more than enough to run the applications that were open at the time; in fact, the system with 16GB installed would have been fine with 8GB or less.

Before you add RAM to a system (or replace defective RAM chips), you must determine the memory modules required for your system. Your system documentation has this information.

If you need to replace a defective memory module or add more memory to your system, you have several methods for determining the correct module for your system:

- Inspect the modules installed in your system. Each module has markings that indicate its capacity and speed. RAM capacity and speed were discussed in detail earlier in this chapter. You can write down the markings on the memory module and use them to determine the type of memory you need. Check with a local store or an online memory vendor for help.
- Look up your system using the online memory-configuration utility provided by your preferred memory vendor. Originally, these configuration utilities were primarily for users of name-brand systems. However, most vendors have now added major motherboard brands and models to their databases. Therefore, if you know your system or motherboard brand and model, you can find the memory that is recommended.
- Download and run analysis software that the memory module maker or a third party provides. CPU-Z and similar programs use the SPD chip on each module to determine this information.
- Consult your system documentation. I list this option last for a reason. If you have installed BIOS upgrades, you might be able to use larger and faster memory than your documentation lists as supported by your system. You should check the latest tech notes and documentation available online for your system and check the BIOS version installed in your system to determine which memory-related features it has. A BIOS upgrade might enable your system to use faster memory.

Adding the wrong modules to a system can make it as unreliable as leaving a defective module installed and trying to use the system in that condition.

Purchasing Memory

When purchasing memory, you need to consider certain issues. Some are related to the manufacturing and distribution of memory, whereas others depend on the type of memory you are purchasing.

Suppliers

Many companies sell memory, but only a few companies actually make memory. Additionally, only a few companies make memory chips, but many more companies make memory modules. Most of the companies that make the actual RAM chips also make modules containing their own chips. Other companies, however, strictly make modules; these companies purchase memory chips from several chip makers and then produce modules with these chips. Finally, some companies don't make either the chips or the modules. Instead, they purchase modules made by other companies and relabel them.

I refer to memory modules made by the chip manufacturers as *first-party modules*, whereas those made by module (but not chip) manufacturers I call *second-party modules*. Finally, those that are simply relabeled first- or second-party modules under a different name are called *third-party modules*. I always prefer to purchase first- or second-party modules if I can because they are better documented. In essence, they have a better pedigree, and their quality is generally more assured. Not to mention that purchasing from the first or second party eliminates one or more middlemen in the distribution process as well.

First-party manufacturers (where the same company makes the chips and the modules) include Micron (www.crucial.com), Samsung, Hynix, and others. Second-party companies that make the modules (but not the chips) include Kingston, Viking, PNY, Simple Tech, Smart, and Mushkin. At the third-party level, you are not purchasing from a manufacturer but from a reseller or remarketer instead.

Most of the large manufacturers don't sell small quantities of memory to individuals, but some have set up factory outlet stores where individuals can purchase as little as a single module. One of the largest memory manufacturers in the world, Micron, sells direct to the consumer at www.crucial.com. Because you are buying direct, the pricing at these outlets is often competitive with second- and third-party suppliers.

Considerations in Purchasing DIMMs

When you are purchasing DIMMs, here are the main things to consider:

- Do you need DDR2, DDR3, or DDR4 versions?
- Do you need ECC or non-ECC?
- Do you need standard (unbuffered) or registered versions?
- Do you need a specific voltage?
- What speed grade do you need?
- Do you need a specific CAS latency?

Currently, DIMMs come in DDR2, DDR3, and DDR4 versions. They are not interchangeable because they use completely different signaling and have different notches to prevent a mismatch. High-reliability systems such as servers can use ECC versions, although most desktop systems use the less-expensive non-ECC types. Most systems use standard unbuffered DIMMs, but file server or workstation motherboards designed to support large amounts of memory might require registered DIMMs (which also include ECC support). Registered DIMMs contain their own memory registers, enabling the module to hold more memory than a standard DIMM. DIMMs come in a variety of speeds, with the rule that you can always substitute a faster one for a slower one, but not vice versa.

Some memory modules are designed to run on non-standard voltages, which may be useful when overclocking them. Unfortunately this can also cause problems for systems where stock (nonoverclocked) memory settings are used. Standard voltages for DDR, DDR2, DDR3, and DDR4 modules are 2.5V, 1.8V, 1.5V, and 1.2V, respectively. If you buy a DDR3 module rated at a higher voltage (1.6V or higher), it might not perform reliably when run on the standard 1.5V setting. I've seen systems with constant lockup and crashing problems due to improperly configured memory like this. My recommendation is to purchase only memory rated to run on the standard voltage for that type, which is 1.5V in the case of DDR3 and 1.2V in the case of DDR4.

Another speed-related issue, as discussed earlier in this chapter, is the column address strobe latency. Sometimes this specification is abbreviated CAS or CL and is expressed in a number of cycles, with lower numbers indicating higher speeds (fewer cycles). The lower CAS latency shaves a cycle off a burst mode read, which marginally improves memory performance. Single data rate DIMMs are available in CL3 or CL2 versions. DDR DIMMs are available in CL2.5 or CL2 versions. DDR2 DIMMs are available in CL 3, 4, or 5. DDR3 DIMMs are available in CL 5, 6, 7, 8, and 9. DDR4 modules are available in CL 12, 14, 15, and 16. With all memory types, the lowest CL number is the fastest (and usually the most expensive) memory type. You can mix DIMMs with different CAS latency ratings, but the system usually defaults to cycling at the slower speeds of the lowest common denominator.

Considerations in Purchasing Obsolete Memory

Many people are surprised to find that obsolete memory types cost much more than those that current systems use. This is because of simple supply and demand; what is least popular generally costs the most. This can make adding memory to older systems cost prohibitive.

For example, in February 2015, a 1GB DDR PC2700 module cost about $30 at Crucial.com. A 1GB DDR2 PC2-5300 or PC2-6400 module was about $23, and a 1GB DDR3 PC3-12800 module was about $16. As you can see from this comparison, it could cost almost twice as much to buy the same amount of RAM for a system that uses the old DDR memory as current DDR3 memory (which is much faster and also available at higher capacities).

Tip

Instead of buying "new" obsolete memory for older systems, check with computer repair shops, eBay, Craigslist, websites specializing in surplus memory, or other users who might have a collection of old parts.

High-reliability systems might want or need ECC versions, which have extra ECC bits. As with other memory types, you can mix ECC and non-ECC types, but systems can't use the ECC capability.

Replacing Modules with Higher-Capacity Versions

If all the memory module slots on your motherboard are occupied, your best option is to remove an existing bank of memory and replace it with higher-capacity modules.

However, just because higher-capacity modules are available to plug into your motherboard, don't automatically assume the higher-capacity memory will work. Your system's chipset, BIOS, and OS set limits on the capacity of the memory you can use. Check your system or motherboard documentation to see which size modules work with it before purchasing the new RAM. You should make sure you have the latest BIOS for your motherboard when installing new memory.

If your system supports dual- or triple-channel memory, you must use modules in matched pairs or triples (depending on which type your system supports) and install them in the correct location on the motherboard. Consult your motherboard manual for details.

Installing Memory Modules

When you install or remove memory, you are most likely to encounter the following problems:

- Electrostatic discharge
- Improperly seated modules
- Incorrect memory configuration settings in the BIOS Setup

To prevent electrostatic discharge (ESD) when you install sensitive memory chips or boards, you shouldn't wear synthetic-fiber clothing or leather-soled shoes because these promote the generation of static charges. Remove any static charge you are carrying by touching the system chassis before you begin, or better yet, wear a good commercial grounding strap on your wrist. You can order one from any electronics parts store. A grounding strap consists of a conductive wristband grounded at the other end through a 1-meg ohm resistor by a wire clipped to the system chassis. Be sure the system you are working on is unplugged.

Caution

Be sure to use a properly designed commercial grounding strap; do not make one yourself. Commercial units have a 1-meg ohm resistor that serves as protection if you accidentally touch live power. The resistor ensures that you do not become the path of least resistance to the ground and therefore become electrocuted. An improperly designed strap can cause the power to conduct through you to the ground, possibly killing you.

Follow this procedure to install memory on a typical desktop PC:

1. Shut down the system and unplug it. As an alternative to unplugging it, you can turn off the power supply using the on/off switch on the rear of some power supplies. Wait about 10 seconds for any remaining current to drain from the motherboard.
2. Open the system. See the system or case instructions for details.
3. Connect a static guard wrist strap to your wrist and then to a metal portion of the system chassis, such as the frame. Make sure the metal plate on the inside of the wrist strap is tight against the skin of your wrist.
4. Some motherboards feature an LED that glows as long as the motherboard is receiving power. Wait until the LED dims before removing or installing memory.
5. Move obstructions inside the case, such as cables or wires, out of the way of the memory modules and empty sockets. If you must remove a cable or wire, note or take a picture of its location and orientation so you can replace it later.
6. If you need to remove an existing module, flip down the ejector tab at each end of the module and lift the module straight up out of the socket. Note the keying on the module.
7. Note the specific locations needed if you are inserting modules to operate in dual-channel mode. The sockets for dual-channel memory might use a different-colored plastic to distinguish them from other sockets, but ultimately you should consult the documentation for your motherboard or system to determine the proper orientation.
8. To insert a module into a socket, ensure that the ejector tabs are flipped down on the socket you plan to use. DIMMs are keyed by one or more notches along the bottom connector edges that are offset from the center so they can be inserted in only one direction, as shown in Figure 6.20.
9. Push down on the module until the ejector tabs lock into place in the notch on the side of the module. It’s important that you not force the module into the socket. If the module does not slip easily into the slot and then snap into place, it is probably not oriented or aligned correctly. Forcing the module could break it or the socket. Refer to Figure 6.20 for details.
10. Replace any cables or wires you disconnected.
11. Close the system, reconnect the power cable, and turn on the PC.

After installing the memory and putting the system back together, you might have to run the BIOS Setup and resave with the new amount of memory being reported. Most systems automatically detect the new amount of memory and reconfigure the BIOS Setup settings for you. Most systems today also don’t require setting any jumpers or switches on the motherboard to configure them for your new memory.

After configuring your system to work properly with the additional memory, you might want to run a memory-diagnostics program to ensure that the new memory works properly.

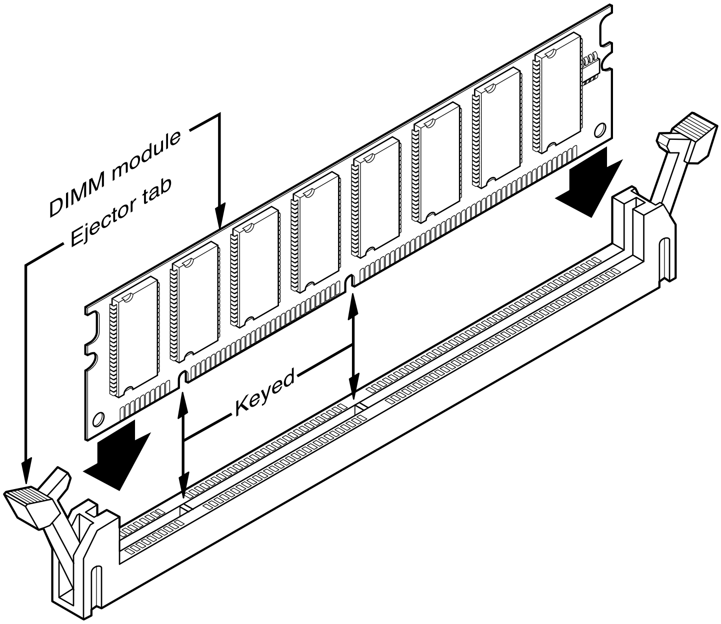


Figure 6.20 DIMM key(s) should match the protrusion(s) in the DIMM sockets.

Troubleshooting Memory

Memory problems can be difficult to troubleshoot. For one thing, computer memory is still mysterious to people because it is a kind of "virtual" thing that can be hard to grasp. The other difficulty is that memory problems can be intermittent and often look like problems with other areas of the system, even software. This section shows simple troubleshooting steps you can perform if you suspect you are having a memory problem.

To troubleshoot memory, you first need some memory-diagnostics testing programs. You already have several and might not know it. Every motherboard BIOS has a memory diagnostic in the POST that runs when you first turn on the system. In most cases, you also receive a memory diagnostic on a utility disk that came with your system. Many commercial diagnostics programs are on the market, and almost all of them include memory tests.

When the POST runs, it not only tests memory, but it also counts it. The count is compared to the amount counted the last time BIOS Setup was run; if it is different, an error message is issued. As the POST runs, it writes a pattern of data to all the memory locations in the system and reads that pattern back to verify the memory works. If any failure is detected, you see or hear a message. Audio messages (beeping) are used for critical or "fatal" errors that occur in areas important for the system's operation. If the system can access enough memory to at least allow video to function, you see error messages instead of hearing beep codes.

See Chapter 19, "PC Diagnostics, Testing, and Maintenance," for detailed listings of the BIOS beep and other error codes, which are specific to the type of BIOS you have.

If your system makes it through the POST with no memory error indications, there might not be a hardware memory problem, or the POST might not be able to detect the problem. Intermittent memory errors are often not detected during the POST, and other subtle hardware defects can be hard for the POST to catch. The POST is designed to run quickly, so the testing is not nearly as thorough as it could be. That is why you often have to boot from standalone diagnostic media (normally an optical disc or a bootable flash drive) and run a true hardware diagnostic to do more extensive memory testing. You can run these types of tests continuously and leave them running for days if necessary to hunt down an elusive intermittent defect.

Fortunately, several excellent memory test programs are available for free download. Here are some I recommend:

- **Microsoft Windows Memory Diagnostic**—included with Vista and later
- **Memtest86**—www.memtest86.com
- **Ultimate Boot CD**—www.ultimatebootcd.com

Not only are these free, but they also are available in a bootable format, which means you don't have to install software on the system you are testing. The bootable format is actually required in a way because Windows and other OSs prevent the direct access to memory and other hardware required for testing. These programs use algorithms that write different types of patterns to all the memory in the system, testing every bit to ensure it reads and writes properly. They also turn off the processor cache to ensure direct testing of the modules and not the cache. Some, such as Windows Memory Diagnostic, even indicate the module that is failing should you encounter an error. Note that a version of the Windows Memory Diagnostic is also included with Windows 7/8.1/10 and Vista. It can be found as part of the Administrative Tools (mdsched.exe), as well as on the bootable install DVDs under the Repair option.

The Ultimate Boot CD includes Memtest86 and several other memory diagnostic programs. Both the Ultimate Boot CD and the Windows Vista and later install DVDs (containing the Windows Memory Diagnostic) can also be installed on a bootable USB flash drive, which makes them much more convenient to use. To create a bootable flash drive with the Ultimate Boot CD you would use the ubcd2usb command as described in the "Customizing UBCD" page on the www.ultimatebootcd.com website. To create a bootable flash drive version of a Windows 7/8.1/10 or Vista installation DVD you can use the Windows 7 USB/DVD download tool provided by Microsoft (http://tinyurl.com/4qfdm4x). Note that although the tool has "Windows 7" in the name, it works on Windows 8.1/10 and Vista discs as well. If you want to test the memory on a system that already has Windows 7/8.1/10 or Vista installed, merely run the mdsched.exe command or open the Control Panel, Administrative Tools and select the Windows Memory Diagnostic, which enables you to restart the system and run the test immediately or set the system to run the test automatically on the next restart.

One problem with software-based memory diagnostics is that they do only pass/fail type testing; that is, all they can do is write patterns to memory and read them back. They can't determine how close the memory is to failing—only whether it worked. For the highest level of testing, the best thing to have is a dedicated memory test machine, usually called a *module tester*. These devices enable you to insert a module and test it thoroughly at a variety of speeds, voltages, and timings to let you know for certain whether the memory is good or bad. Versions of these testers are available to handle all types of memory modules. I have defective modules, for example, that work in some systems (slower ones) but not others. What I mean is that the same memory test program fails the module in one machine but passes it in another. In the module tester, it is always identified as bad right down to the individual bit, and it even tells me the actual speed of the device, not just its rating. Companies that offer memory module testers include Tanisys (www.tanisys.com), CST (www.simmtester.com), and Innoventions (www.memorytest.com). They can be expensive, but for a high volume system builder or repair shop, using one of these module testers can save time and money in the long run.

After your OS is running, memory errors can still occur, typically identified by error messages you might receive. Here are the most common:

- **Parity errors**—The parity-checking circuitry on the motherboard has detected a change in memory since the data was originally stored. (See the "How Parity Checking Works" section earlier in this chapter.)
- **General or global protection faults**—A general-purpose error indicating that a program has been corrupted in memory, usually resulting in immediate termination of the application. This can also be caused by buggy or faulty programs.
- **Fatal exception errors**—Error codes returned by a program when an illegal instruction has been encountered, invalid data or code has been accessed, or the privilege level of an operation is invalid.
- **Divide error**—A general-purpose error indicating that a division by 0 was attempted or the result of an operation does not fit in the destination register.

If you are encountering these errors, they could be caused by defective or improperly configured memory, but they can also be caused by software bugs (especially drivers), bad power supplies, static discharges, close proximity radio transmitters, timing problems, and more.

If you suspect the problems are caused by memory, there are ways to test the memory to determine whether that is the problem. Most of this testing involves running one or more memory test programs.

Another problem with software-based diagnostics is running memory tests with the system caches enabled. This effectively invalidates memory testing because most systems have what is called a *write-back cache*. This means that data written to main memory is first written to the cache. Because a memory test program first writes data and then immediately reads it back, the data is read back from the cache, not the main memory. It makes the memory test program run quickly, but all you tested was the cache. The bottom line is that if you test memory with the cache enabled, you aren't really writing to the modules, but only to the cache. Before you run any memory test programs, be sure your processor/memory caches are disabled. Many older systems have options in the BIOS Setup to turn off the caches. Current software-based memory test software such as the Windows Memory Diagnostic and Memtest86 can turn off the caches on newer systems.

The following steps enable you to effectively test and troubleshoot your system RAM. Figure 6.21 provides a boiled-down procedure to help you step through the process quickly.

First, let's cover the memory-testing and troubleshooting procedures:

1. Power up the system and observe the POST. If the POST completes with no errors, basic memory functionality has been tested. If errors are encountered, go to the defect isolation procedures.
2. Restart the system and then enter your BIOS (or CMOS) Setup. In most systems, this is done by pressing the Del, F1, or F2 key during the POST but before the boot process begins (see your system or motherboard documentation for details). Once in BIOS Setup, verify that the memory count is equal to the amount that has been installed. If the count does not match what has been installed, go to the defect isolation procedures.
3. Find the BIOS Setup options for cache and then set all cache options to disabled if your system supports this option. Figure 6.22 shows a typical Advanced BIOS Features menu with the cache options highlighted. Save the settings and reboot to bootable media containing the memory diagnostics program.

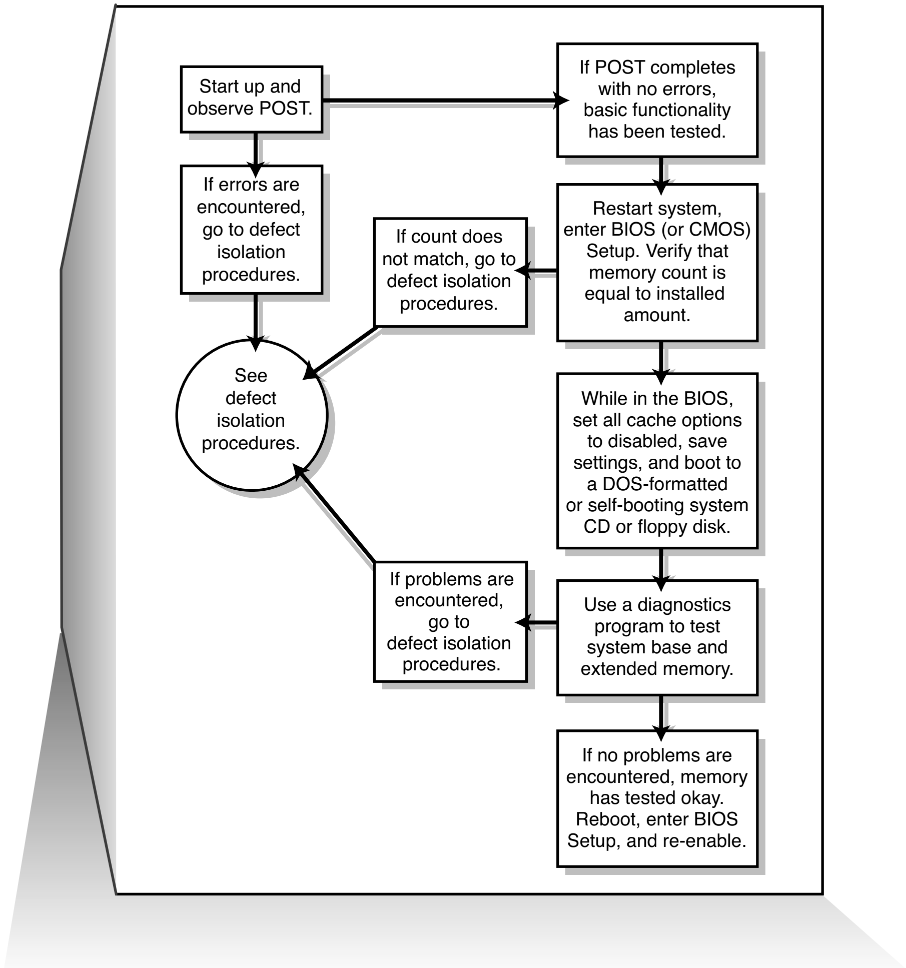


Figure 6.21 Testing and troubleshooting memory.

Tip

Most systems do not permit you to disable cache in BIOS setup. In such cases, I recommend using Windows Memory Diagnostic and use its advanced options to disable cache before testing memory.

4. Follow the instructions that came with your diagnostic program to have it test the system base and extended memory. Most programs have a mode that enables them to loop the test—that is, to run it continuously, which is great for finding intermittent problems. If the program encounters a memory error, proceed to the defect isolation procedures.
5. If no errors are encountered in the POST or in the more comprehensive memory diagnostic, your memory has tested okay in hardware. Be sure at this point to reboot the system, enter the

BIOS Setup, and re-enable the cache. The system will run very slowly until the cache is turned back on.

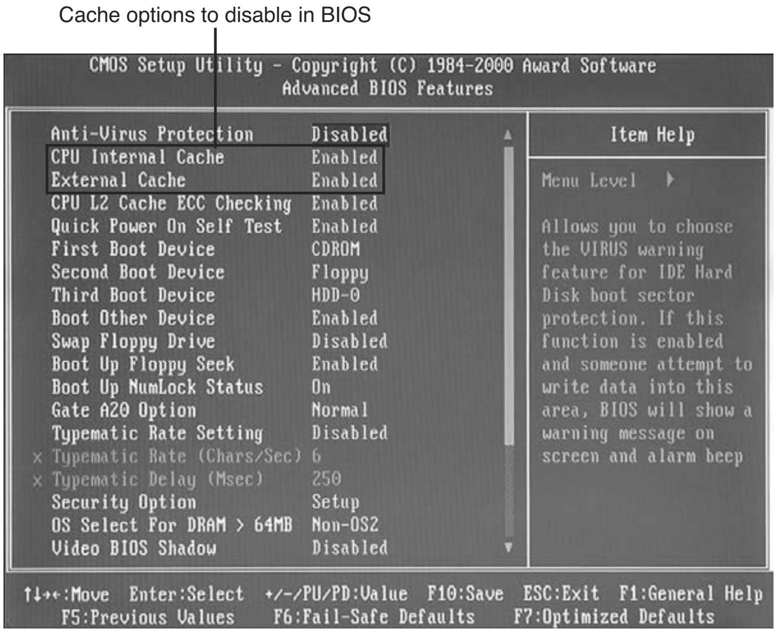


FIGURE 6.22 For accurate results, before testing memory the CPU Internal (L1) and External (L2 and L3) caches should be disabled in the system BIOS Setup on systems having this option.

6. If you are having memory problems yet the memory still tests okay, you might have a problem undetectable by simple pass/fail testing, or your problems could be caused by software or one of many other defects or problems in your system. You might want to bring the memory to a module tester for a more accurate analysis. Some larger PC repair shops have such a tester. I would also check the software (especially drivers, which might need updating), power supply, and system environment for problems such as static, radio transmitters, and so forth.

Memory Defect Isolation Procedures

To use these steps, I am assuming you have identified an actual memory problem that the POST or disk-based memory diagnostics have reported. If this is the case, see the following steps and Figure 6.23 for instructions to identify or isolate which module is causing the problem.

1. Restart the system and enter the BIOS Setup. Under a menu usually called Advanced or Chipset Setup might be memory timing parameters. Select BIOS or Setup defaults, which are usually the slowest settings. If the memory timings have been manually set, as shown in Figure 6.21, reset the memory configuration to By SPD.
2. Save the settings, reboot, and retest using the testing and troubleshooting procedures listed earlier. If the problem has been solved, improper BIOS settings were the problem. If the problem remains, you likely do have defective memory, so continue to the next step.

3. Open the system for physical access to the modules on the motherboard. Identify the bank arrangement in the system. Using the manual or the legend silk-screened on the motherboard, identify which modules correspond to which banks. Remember that if you are testing a multi-channel system, you must be sure you remove all the modules in the same channel.

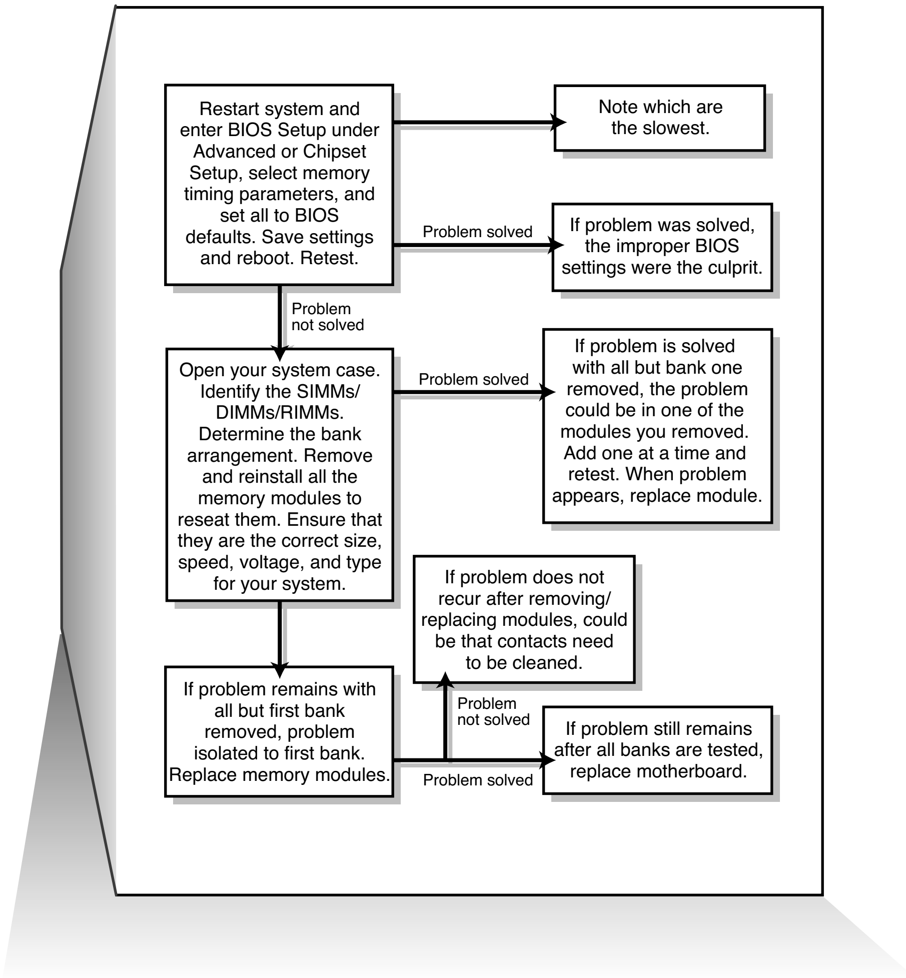


FIGURE 6.23 Follow these steps if you are still encountering memory errors after completing the steps in Figure 6.21.

4. Remove all the memory except the first bank, and then retest using the troubleshooting and testing procedures listed earlier (see Figure 6.24). If the problem remains with all but the first bank removed, the problem has been isolated to the first bank, which you must replace.
5. Replace the memory in the first bank (preferably with known good spare modules, but you can also swap in others that you have removed) and then retest. If the problem still remains after

testing all the memory banks (and finding them all to be working properly), it is likely the motherboard is bad (probably one of the memory sockets). Replace the motherboard and retest.

6. At this point, the first (or previous) bank has tested to be good, so the problem must be in the remaining modules that have been temporarily removed. Install the next bank of memory and retest. If the problem resurfaces now, the memory in that bank is defective. Continue testing each bank until you find the defective module.

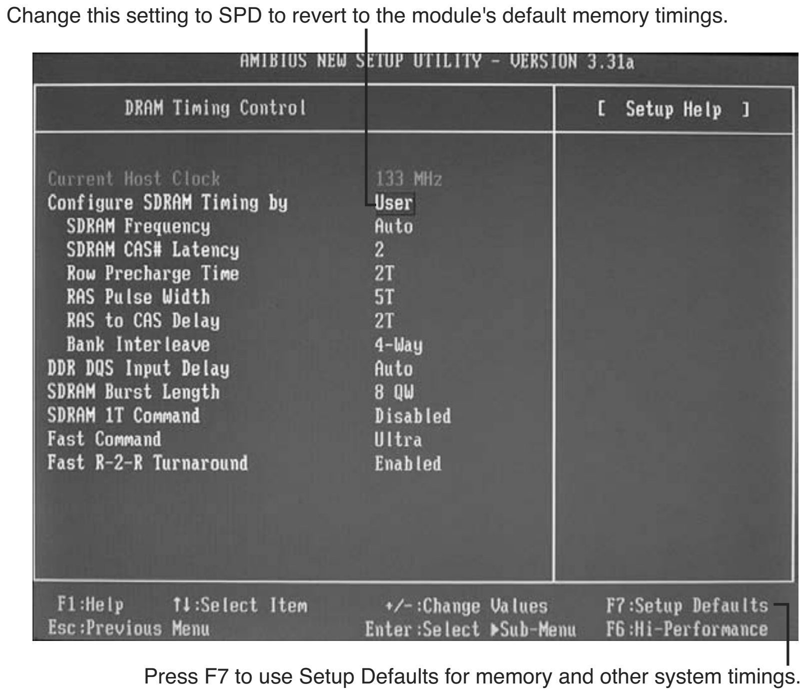


FIGURE 6.24 This system is using user-defined memory timings, which could cause the memory to be unstable.

7. Repeat the preceding step until all remaining banks of memory are installed and have been tested. If the problem has not resurfaced after you have removed and reinstalled all the memory, the problem was likely intermittent or caused by poor conduction on the memory contacts. Often simply removing and replacing memory can resolve problems because of the self-cleaning action between the module and the socket during removal and reinstallation.

The System Logical Memory Layout

The original PC had a total of 1MB of addressable memory, and the top 384KB of that was reserved for use by the system. Placing this reserved space at the top (between 640KB and 1,024KB, instead of at the bottom, between 0KB and 640KB) led to what is often called the *conventional memory barrier*. The constant pressures on system and peripheral manufacturers to maintain compatibility by never breaking from the original memory scheme of the first PC has resulted in a system memory structure that is (to put it kindly) a mess. More than three decades after the first PC was introduced, even the newest systems are limited in many important ways by the memory map of the first PCs.

The original PC used an Intel 8088 processor that could run only 16-bit instructions or code, which ran in what was called the *real mode* of the processor. These early processors had only enough address lines to access up to 1MB of memory, and the last 384KB of that was reserved for use by the video card as video RAM, other adapters (for on-card ROM BIOS or RAM buffers), and finally the motherboard ROM BIOS.

The 286 processor brought more address lines, enough to allow up to 16MB of RAM to be used, and a new mode called protected mode that you had to be in to use. One area of confusion was that RAM was now noncontiguous; that is, the OS could use the first 640KB and the last 15MB, but not the 384KB of system reserved area that sat in between.

When Intel released the first 32-bit processor in 1985 (the 386DX), the memory architecture of the system changed dramatically. There were now enough address lines for the processor to use 4GB of memory, but this was accessible only in 32-bit protected mode, in which only 32-bit instructions or code could run. Unfortunately, it took 10 years for the industry to transition from 16-bit to 32-bit OSs and applications. From a software instruction perspective, all the 32-bit processors since the 386 are really just faster versions of the same.

When AMD released the first x86-64 processor in 2003 (Intel followed suit in 2004), the 64-bit era was born. In addition to 16-bit and 32-bit modes, these chips have a 64-bit mode (commonly referred to as x64 or x86-64). 64-bit processors have three distinctly different modes, with unique memory architectures in each. For backward compatibility, 64-bit processors can run in 64-bit, 32-bit, or 16-bit modes, and 32-bit processors can run in 32-bit or 16-bit modes, each with different memory limitations. For example, a 64-bit processor running in 32-bit mode can only address 4GB of RAM, and a 64-bit or 32-bit processor running in 16-bit mode can only address 1MB of RAM. All Intel-compatible PC processors begin operation in 16-bit real mode when they are powered on. When a 32-bit or 64-bit OS loads, it is that OS code that instructs the processor to switch into 32-bit or 64-bit protected mode.

When a 32-bit OS such as Windows is loaded, the processor is switched into 32-bit protected mode early in the loading sequence. Then, 32-bit drivers for all the hardware can be loaded, and the rest of the OS can load. In 32-bit protected mode, the OSs and applications can access all the memory in the system up to 4GB. Similarly, on a 64-bit OS, the system switches into 64-bit protected mode early in the boot process and loads 64-bit drivers, followed by the remainder of the OS.

The 32-bit editions of Windows support 4GB of physical memory (RAM). What many don't realize is that the PC system hardware uses some or all of the fourth gigabyte for the BIOS, motherboard resources, memory mapped I/O, PCI configuration space, device memory (graphics aperture), VGA memory, and so on. This means that if you install 4GB (or more) RAM, none of it past 4GB will be seen at all, and most or all of the fourth gigabyte (that is, the RAM between 3GB and 4GB) will be disabled because it is already occupied by other system hardware. This is called the *3GB limit*, which is analogous to the 640K memory limit we had on 16-bit systems in the 1980s. The 16-bit addressing supported 1MB, but the upper 384K was already in use by the system hardware (BIOS, video, adapter ROM, and so on).

Figure 6.25 shows the memory map for a system using an Intel G45 chipset, which supports a maximum of 16GB of RAM. For a 32-bit OS, the line labeled "Top of usable DRAM (32-bit OS)" is at 4,096MB. Note that the PCI memory range, FLASH, APIC (Advanced Programmable Interrupt Controller), and Reserved areas take up a total of 770MB of the memory below 4GB. You can also see the 384K (0.375MB) of memory below 1MB that is used by the system. This means that if you are running a 32-bit OS, even if you have 4GB of RAM installed, the amount usable by the OS would be 4,096MB – 770MB – 0.375MB, which is 3,325.625MB (or about 3.24GB, rounded down).

Can any of that unused memory between 3GB and 4GB be reclaimed? For those running a 32-bit OS, the answer is no. However, if you are running a 64-bit OS on a system that supports memory remapping (primarily a function of the motherboard chipset and BIOS), the answer is yes. Most modern motherboard chipsets have a feature that can remap the otherwise disabled RAM in the fourth gigabyte to the fifth (or higher) gigabyte, where it will be both visible to and usable by a 64-bit OS. Note, however, that if the motherboard doesn't support remapping, even when a 64-bit OS is being run, the memory will be lost.

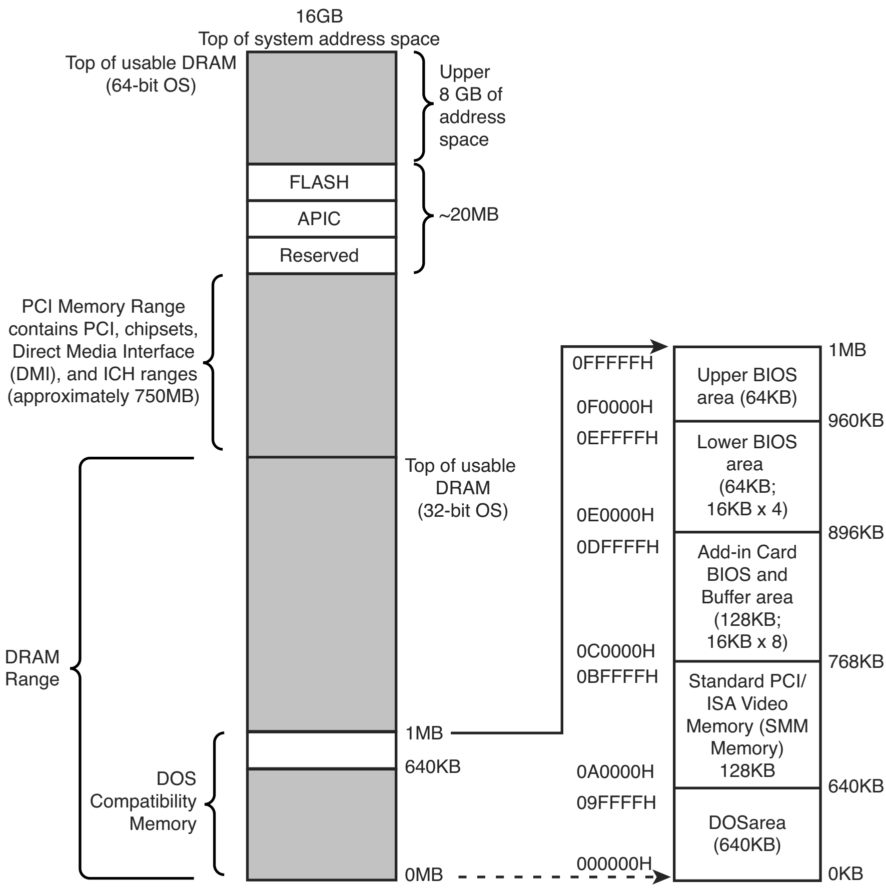


FIGURE 6.25 Memory map for a system using an Intel G45 chipset.

Note that the 3GB limit is not as strictly defined as it was with the 640K limit. This means that if you do install 4GB, you might get to use as much as 3.5GB of it, or possibly as little as 2.5GB or less. It depends largely on the types of buses in the system as well as the type and number of video cards installed. With a single low-end video card, you may have access to 3.5GB. However, on a newer system with two or more PCIe x16 slots, and especially with two or more high-end PCI Express video cards installed, you may drop the usable limit to something close to 2GB.

When running 32-bit editions of Windows, I used to recommend installing a maximum of 3GB RAM, because most if not all of the fourth GB is unusable. However, on systems that support dual-channel memory, it is usually cheaper to install two 2GB modules to get 4GB than it is to install two 1GB modules and two 512MB to get 3GB. On desktop systems that support dual-channel memory, you would not want to install three 1GB modules because in that case not all the memory would run in dual-channel mode.

Index

Symbols

C

D

E

F

J

K

L

O

P

Q

R

S

T

U

V

W

X

Y-Z