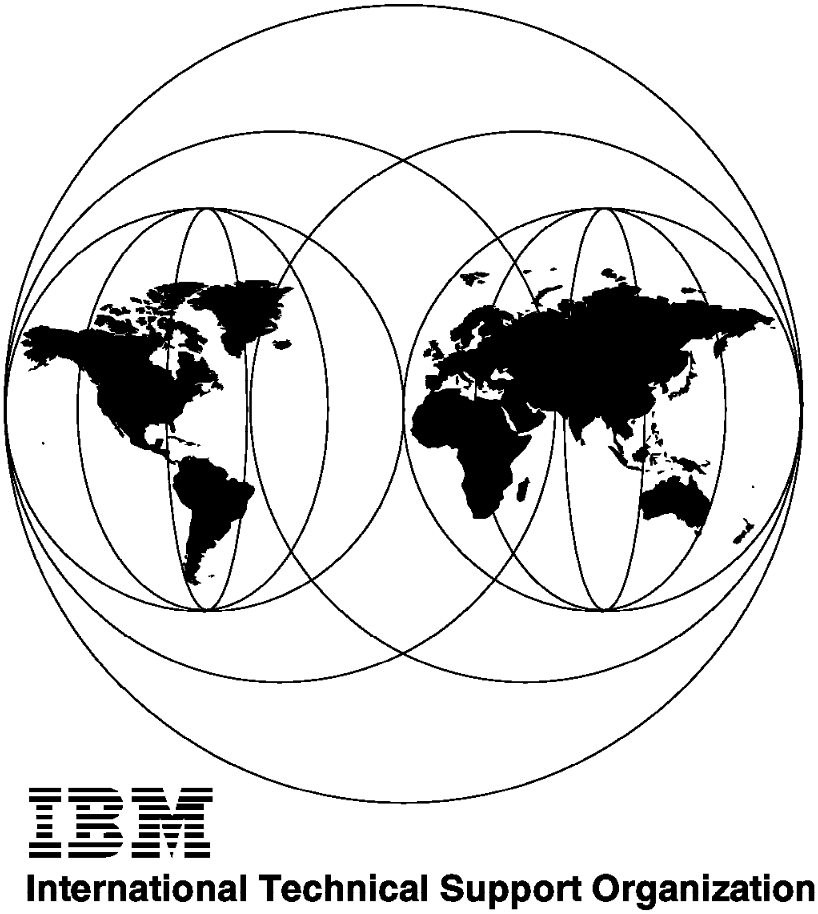
SG24-4760-01

IBM PC Server Selection Guide

December 1996



IBM

International Technical Support Organization
Raleigh Center

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix C, "Special Notices" on page 207.

Second Edition (December 1996)

This edition applies to IBM PC Servers for use with IBM and OEM operating systems.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. HZ8 Building 678
P.O. Box 12195
Research Triangle Park, NC 27709-2195

Contents

Preface

This redbook is a reference guide for selecting IBM PC Server hardware. Its primary purpose is to describe the technology that is used to produce high-capacity machines suitable for today′s business environment. The complete IBM PC Server line is described showing diagrams of the internal layout of the systems and details of the technical specifications of each component. Each model in the server range is summarized in tabular form for easy reference. Various options such as external storage enclosures, SCSI controllers and SCSI devices are also described.

This book is intended for IBM customers, dealers, and other technical professionals who need to understand IBM′s PC Server product line.

Basic knowledge of PCs and PC Servers is assumed.

How This Redbook Is Organized

This redbook contains 224 pages. It is organized into three parts as follows:

- Part 1, "Server Requirements"

  This part leads the reader through a discussion of how to determine what form their server solution should take to best satisfy the business requirements.

- Part 2, "Server Technology"

  This part describes the technologies used in the current class of IBM PC servers and what benefits can be gained by using them. Technologies covered are:

  – Processors
  – Buses
  – Memory and cache
  – Storage
  – Network
  – Security
  – Systems management

- Part 3, "Server Products"

  The part describes all the PC servers currently available from IBM and the options that can be used to enhance the server′s function. Topics in this part are:

  – PC Server 310
  – PC Server 320
  – PC Server 325
  – PC Server 330
  – PC Server 520
  – PC Server 704
  – PC Server 720
  – Server enclosures
  – Storage controllers
  – SCSI disks
  – IBM PC Server options

The appendixes at the end of this redbook describe the TechConnect program and its benefits, as well as possible sources of information, software and updates applicable to IBM PC servers.

The Team That Wrote This Redbook

This redbook was produced by a team of specialists from around the world working at the Systems Management and Networking ITSO Center, Raleigh.

Tim Kearby is an Advisory ITSO Specialist for PC Servers at the Systems Management and Networking ITSO Center, Raleigh. He conducts residencies and writes redbooks on IBM PC Servers and network operating systems. Tim has held various positions in his IBM career including assignments in product development, systems engineering, and consulting. He holds a Bachelors of Science degree in Electrical Engineering from Purdue University.

David Watts is a PC hardware marketing specialist working for IBM in Australia. He has been working with PCs for the past 14 years and in the last two years, he has been a server specialist for the IBM PC Company. He has a Bachelors of Engineering degree from the University of Queensland and has worked for IBM Australia for eight years since graduating.

Jonathan (Jonno) Wright is a Country Specialist in IBM New Zealand. He supports IBM PC Servers and corporate desktops for customers throughout New Zealand. He has worked for IBM for three years and has held various positions in customer engineering.

Werner Orgas is a PC Server Marketing Specialist in IBM Germany. He has worked for IBM for nearly 20 years and in his career, he has held various positions in marketing and systems engineering including MVS and VM operating systems support and VM systems programming. For the past two years, Werner has specialized on Intel-based PC Servers.

This redbook contains information from previous redbooks and several other IBM documents. Thanks to the following authors for their contributions:

- Dr. Pat Bowlds
- Dave Gover
- Frank Morassutti
- Dan Neal
- Barry Nusbaum
- Gregory Pfister
- Angelo Rimoldi
- Jean-Paul Simoen
- Ken Stark
- Chuck Tribolet
- Wuilbert Martinez Zamora

Thanks also to the following people for the invaluable advice and guidance provided in the production of this redbook:

- Mike Anderson, Product Adminstrator, PC Server Brand Management, Raleigh
- Fran Collins, Senior Systems Engineer, Personal Systems Competency Center, Dallas
- John Dinwiddie, Chief Engineer, High-End Server Development, Raleigh

- Art Drummond, OS/2 Specialist, IBM Germany
- Steve Edwards, Worldwide ServerGuide Product Manager, Raleigh
- Dave Laubscher, Senior Systems Engineer, PC Server Competency Center, Raleigh
- Mikael Pettersson, PC Server Specialist, IBM Sweden
- Marc Shelley, Advisory Systems Engineer, PC Server Competency Center, Raleigh
- Sean Whalen, Product Administrator, PC Server Brand Management, Raleigh
- Gail Wojton, Linda Robinson, Brian Elkins, and all in the Edit/Graphics Team, ITSO Raleigh

Comments Welcome

We want our redbooks to be as helpful as possible. Should you have any comments about this or other redbooks, please send us a note at the following address:

redbook@vnet.ibm.com

Your comments are important to us!

Part 1. Server Requirements

Chapter 1. How to Select a PC Server

There are three main steps in determining which server solution is the best for your (or your customer′s) environment:

1. Establish the business requirements
2. Determine the technology to meet the requirements
3. Select the products that best use this technology

The three parts of this book cover these three steps. Part 1, "Server Requirements" on page 1 discusses the first step, establishing the business requirements. It also discusses the second step, determining the technology to meet these requirements.

Part 2, "Server Technology" on page 11 discusses the major technologies that are currently implemented in the IBM PC Server product line. Part 3, "Server Products" on page 99 describes the products in the IBM PC Server line that are available in terms of their use of this technology and their applicability in meeting customers′ business requirements.

1.1 Establishing the Business Requirements

The choice of a server starts with establishing the business-related requirements. For example, what problems is the server going to solve? What tasks are going to be done better, or faster, or more reliably in this department, location, or enterprise because a PC server has now been installed? It is important to clearly understand these requirements before proceeding to the next step.

The process of establishing these requirements generally consists of gathering requirements in the following areas:

- Functional requirements
- Performance requirements
- Reliability/availability requirements
- Systems management issues
- Cost constraints

Gathering requirements in each of these areas will lead to a set of overall business requirements that your PC Server solution will need to meet.

1.1.1 Functional Requirements

The first step in the process of selecting and implementing PC Servers, whether one server or two thousand, is to understand clearly what the server is supposed to do. PC Servers primarily take on one or more *roles* which are somewhat standardized in today′s LAN environment. Table 1 on page 4 lists these roles and gives a brief explanation of each.

Role	Description
File Server	Provides access to files that can be used by individual users or shared by several users as a single point of reference. These files can be stored on fixed disks, CD-ROMs, or optical devices. Used also to serve application programs that are required by multiple users, especially when the code/data storage requirements for such applications are large. The applications are stored on the server and downloaded to the client at run time. Facilities to control multiple, simultaneous access must be included.
Print Server	Allows access to shared printers either directly attached to the server or redirected through the server to a LAN attached printer. The print spools for these printers are normally stored and managed by the print server.
Application Server	Distributes the processing workload of applications between the client and the server. One form this can take is that of a transaction server. This is similar to the transaction processing traditionally provided by mainframes. The server either carries out the processing locally, or acts as an intermediary to check and forward transactions on to a larger host system.
Database Server	Provides access to shared data, typically through a relational database management system. Processing can also be distributed between the client and the server via stored procedure calls from the client which are executed on the server. This technique also reduces the amount of LAN traffic between client and server. The database may reside on the server, or there may be links to databases on other systems. These links are managed independently of the user, who need not be aware of the specific location or means of access used.
Communications Server	A communication server enables a LAN client to access resources outside the LAN environment. The client machine does not need any special communications hardware except for the LAN adapter. All access to the outside is through the communications server. Examples include the gateway function that allows the client to access host resources, a modem pool that allows access to outside dial-up services, and a fax service that provides both in-bound and out-bound facsimile transmission through a shared fax card in the server.
Remote Access Server	Allows access to LAN resources from the outside world typically through dial-up connections. This service is needed for people who need access to the LAN resources from remote locations. Examples include mobile workers, telecommuters, and business partners.
E-Mail Server	Manages electronic mail communications and provides local mail box facilities. Manages mail user definitions and access for all users of the local mail system and provides the e-mail gateway function for connecting local mail users to other mail systems, either inter or intra-enterprise.
Collaboration Server	Combines the functions of e-mail, file serving, and database to provide applications that facilitate the automation of business processes, especially those that are communications-centric and/or require document distribution and/or work flow.
Software Distribution	Manages distribution packages for automated, remote installation of operating systems, applications, and maintenance releases. This type of server can work along with a host system which manages the package creation and scheduling or can work independently.
Multimedia Server	Provides shared access to audio and video data from a central repository. The data can be stored on traditional data storage devices such as hard disks, CD-ROMs, or optical devices, or can be provided by standard A/V equipment such as VCRs and laser disk players through specialized interface adapters in the server.

Table 1 (Page 1 of 2). Different Server Roles in Today's LAN

Role	Description
Table 1 (Page 2 of 2). Different Server Roles in Today's LAN	
Security Server	A centralized authentication resource on the LAN that holds access control databases and tables for validity checking. It may also include specialized encryption adapters.
RIPL Server	Supports media-less workstations by providing initial program load (IPL) images for initializing the workstations on the LAN.
Application Development Server	This type of server supports the application developer with a set of programming tools and usually a library management system for developed source code and executables. This role is a subset of the file server role with one exception: The server may have additional burdens placed upon it if it will be performing remote compiles.

1.1.2 Performance Requirements

The PC Server you select must offer sufficient performance and capacity to provide the user with the level of service that is appropriate to the business, even during periods of peak activity. A good way to find the performance requirements of a PC Server is to examine further what role the machine will play. By looking at the intended role, you can determine which of the server's major subsystems will be stressed the most. Depending on the role, the server may become stressed in the processor, the memory, the I/O bus, the disk storage, or the communications adapter(s). Each of these can become the system bottleneck which must be eliminated or avoided if higher performance is to be achieved.

Table 2 on page 6 is meant to give you an overview of the different performance requirements as a function of the different roles that a PC Server can take on.

Where the server plays multiple roles, you will need to combine the requirements according to the mix of roles to be performed. Also keep in mind that when performing multiple roles, the requirements may change as the workload changes throughout a normal working day. For example, at initial startup there may be a heavy demand on the CPU as the server is performing logon authentication and security functions. Later, the load may shift to the disk subsystem as users are running applications that pull data from disk.

Note: Please see Part 2, "Server Technology" on page 11 of this book for a complete discussion on the technologies, such as RAID and SMP, which are mentioned in the table.

Table 2 (Page 1 of 2). Performance Requirements for PC Server Roles	
Role	**Description**
File Server	For best performance, typically needs a lot of memory for file caching and a fast disk subsystem. The processor is not usually the bottleneck except for when it must handle simultaneous multiple logon requests. RAID is beneficial, especially in the case of serving application packages because these are normally stored sequentially on disk. RAID offers less of an advantage for random I/O requests.
Print Server	Typically, performance is limited more by the speed of the attached printers. The client usually formats the print job and then spools it to the print server. (Depending on the number of users supported and the size and frequency of the print jobs, the print server may need substantial disk space to handle the spooling.) The print server then transmits the job to the printer. If a bottleneck occurs inside the print server, it will usually be the LAN adapter or the disk subsystem.
Application Server	This type of server will need a balanced design of all the major subsystems. Obviously, the CPU is important since a portion of the application processing is performed at the server. People often select symmetric multiprocessing (SMP) platforms to serve in this role. As the data is usually centrally managed at the server, the disk subsystem is also important. Since multiple clients will be accessing the data, the throughput of the LAN subsystem will need to be high to prevent bottlenecks.
Database Server	Like the application server, the CPU is important, especially if the server is executing stored procedures. The disk subsystem is also critical. If the data requests are more at random, then RAID level 5 would not be the best choice. RAID level 0 offers the best performance, but you would need to make certain that adequate procedures are in place to protect the data.
Communications Server	The processor and the I/O subsystem are important. In terms of performance, you can think of the communications server as a funnel through which all data must pass to get to its destination. The bandwidth of the LAN adapter(s) must be adequate to handle the load.
Remote Access Server	If you have ever dialed into a server and downloaded a large file, you understand how frustrating it can be to access LAN resources at dial-up speeds. Obviously, you need fast modems to make this service feasible. However, one factor that is often overlooked is the characteristics of the serial ports on your machine. These need the ability to work with the faster serial interface speeds, which normally implies a newer Universal Synchronous/Asynchronous Receiver/Transmitter (USART or UART for short). An example of one such chip is the National Semiconductor 16550A used on IBM PC Servers. Serial ports that support Direct Memory Access (DMA) are also important as the number of users is increased.
E-Mail Server	Since e-mail systems are normally based on store-and-forward techniques, the interactive performance requirements are not as great. Performance issues associated with this role are usually more closely related to capacity. As the number of users on a particular server grows, all subsystems can become stressed.
Collaboration Server	A server performing this role will need substantial resources in both CPU and storage subsystems. This server has more of an interactive workload than does the e-mail server and hence, needs more performance capability. Because these servers usually replicate their databases to other servers on the network, the LAN subsystem should also not be overlooked.
Software Distribution	As it relates to performance, this role can be thought of as a file server. If large, batch distributions are being performed, then additional memory for file caching can be beneficial.

Role	Description
Multimedia Server	As this role usually requires a large amount of data to be transferred, the I/O subsystem is critical. This includes the bandwidth of the I/O bus also. If the digital content is coming off a hard disk, then the disk subsystem is also very important. RAID level 0 is often employed as it offers the best performance and since the data is usually read-only, it can be restored easily from a backup source if necessary.
Security Server	This role places the largest load on the CPU. As the number of users increases, the CPU can become a bottleneck. SMP can work well.
RIPL Server	Like the multi-media server, the disk and LAN subsystems are the most critical here. RAID level 0 can be effective, especially if you are in an environment where the clients are all doing an IPL at the same time such as first thing in the morning. If all clients are using the same IPL image, then extra memory in the server is justified in order to cache the image (depending on the network operating system) or load it onto a virtual disk.
Application Development Server	This role will have the same basic performance requirements as the file server. If your programmers will be compiling code on the server, then the CPU can become a bottleneck. SMP machines can work well here, and are justified, especially when comparing the cost of providing all developer workstations with the equivalent CPU horsepower.

Table 2 (Page 2 of 2). Performance Requirements for PC Server Roles

You should always keep in mind that performance can also be limited by other factors that are not directly related to the machine itself. One example might be the LAN transport system. If you have a requirement to serve multimedia content to 150 users, then a 4 Mbps token-ring segment would probably be inappropriate.

Another example might be system software that is not a good match for the function to be performed. For instance, if you attempt to run 16-bit code on a Pentium Pro processor, be prepared for poor performance.

1.1.3 Reliability/Availability Issues

Most companies today are very concerned about the system availability or *uptime* of the LAN. Certainly, availability is a function of product reliability (though not the sole variable in the equation). If the system never fails, or fails less frequently, then you will be on the right track to providing high systems availability. Also, when a component is unsuccessful, what fault-tolerant features does the system employ to ensure that this failure does not cause lengthy disruptions in service to the end users.

It is worthwhile to stop and define a few terms:

Availability: the percentage of time that the system is up and running and *available* for end users to do productive work. It is only calculated for the hours for which the system is supposed to be available. For example, if your business requires the system to be up from 7:00 a.m. to 11:00 p.m. Monday through Friday, then downtime during the hours of 2:00 a.m. until 4:00 a.m. for system maintenance is not counted against the availability number. As a point of reference, system availability in a mainframe environment where businesses have traditionally hosted mission critical applications is measured in the 99.0 to 99.5 percent range.

High Availability: a target that implies that the system will be available a higher percentage of time than it would otherwise be if no special system features or operational procedures were employed. High availability percentages are more often in the area of 99.95 percent. You can only get to this level by eliminating or masking unplanned outages during scheduled periods of operations. Failure avoidance, fault tolerance, and fast restart techniques are usually employed in order to do this.

Failure Avoidance: the use of highly reliable hardware and software components in a system.

Fault Tolerance: the ability of a system to continue to deliver an acceptable level of service in the event of a component failure. Obviously, in this case, the proper system features and/or operational procedures have to be in place to keep critical system resources up even if a piece of hardware fails.

The most common method of providing fault tolerance is to provide redundancy of critical resources either in the same machine or somewhere else on the network so that the backup can be made available in the event of a failing primary resource.

Fast Restart: the ability to quickly detect and recover from a failure in such a way that disruption in service is minimized.

1.1.3.1 Cost Impact

To be sure, there is a cost associated with increasing systems availability and it grows more quickly as your targets increase. The cost of moving from high availability to continuous availability can be prohibitive for many organizations. Fortunately, most companies don′t require continuous operation in order to achieve their business objectives.

1.1.3.2 Determining Availability Requirements

The best place to start is to decide what level of availability is required, and then to look at technologies and begin to design strategies that will allow for this requirement to be satisfied. Two key questions that should be answered first are:

- What percentage of the time must my systems be available?
- How quickly must I recover from an unplanned outage?

Every business will have different answers to these questions, but generally, it is the unplanned outages of several hours or days that are the most costly to a business. Unfortunately, these have been historically quite common in the LAN environment.

1.1.3.3 Technology Considerations

There are many technologies implemented in a modern PC Server that can aid in achieving higher availability. Most are concerned with eliminating single points of failure in a system.

RAID and hot-pluggable disk drives, SMP, ECC memory, improved system cooling, UPS, and redundant power supplies can all be employed to achieve high availability or fault tolerance if needed. Please see Part 2, “Server Technology” on page 11 of this book for a more complete discussion of these technologies.

1.1.4 Systems Management Issues

The realities of implementing and supporting the system must be kept clearly in mind during the selection process. This will encompass many areas, but traditionally, IBM has classified the general issues of systems management into the following categories:

- Change Management
- Operations Management
- Problem Management
- Performance Management
- Configuration Management
- Business Management

As it relates to the selection of PC Servers, much of the discussion will focus on your strategy regarding centralization versus decentralization of server function.

With a centralized strategy, multiple server roles can be carried out on the same piece of hardware. Previous departmental servers that may have provided service for 10 to 50 users can be consolidated into a fewer number of larger machines that might provide service for 1000 users or more.

With the emergence of a new class of PC *super server* on the market, the centralized approach becomes more feasible in large organizations where capacity limitations would have ruled it out just a short time ago. Some traditional mainframe shops have already decided to consolidate and have implemented an almost glass-house type of environment where their PC Servers are maintained on raised floors and a professional operations staff perform routine management tasks.

In contrast, the distributed approach utilizes a larger number of usually smaller machines that provide a more limited role in the network. Taken to the extreme, each machine would have one specific function for which it is responsible. The *LAN* then becomes this collection of distributed functions comprised of multiple systems placed strategically on the network.

The debate of this issue is ongoing. There are advantages and disadvantages to either approach. Each also have significant implications regarding systems management. If you decide to consolidate your PC Server function, your systems management issues will be quite different than if you decided to distribute server function across the network.

Ultimately, you will have to decide which approach is best for your organization. Some of the factors that might dictate choosing one approach over the other include:

- Transport topology

  The LAN topology may dictate a distributed approach, especially if your enterprise is spread across multiple, geographically-dispersed establishments. In this case, you may have performance requirements that would not allow server access over wide-area links, or the cost of such links that provide sufficient bandwidth may be prohibitive.

  If your enterprise is confined to one establishment or to a metropolitan area, then a consolidated approach may be feasible and even desirable depending upon some of the other considerations below.

- Availability requirements

With respect to systems availability, there are definitely trade-offs to make regarding the decision of whether to centralize the server function. If you decide to consolidate, then you will have less components that can fail and your overall *reliability* (see 1.1.3, "Reliability/Availability Issues" on page 7) will be higher. However, if all or much of the server function is consolidated on one machine, then you will have less flexibility in designing recovery procedures in the event of a component failure.

- Human resource constraints

  Simply stated, the manpower needed to manage systems increases with the number of systems that you have to manage. This includes the time necessary for initial installation and configuration, backup, applying necessary hardware and software maintenance, troubleshooting, and account administration. The latter is especially true in environments where IDs have to be duplicated on multiple machines. Disaster recovery procedures are also usually more complicated with a distributed approach. If you do not have ample personnel available for these tasks, then a consolidated approach may be more desirable.

- Anticipated growth

  If you anticipate a lot of growth and you are looking at a consolidated approach, then you will have to buy enough capacity to meet both the initial and anticipated requirements, or, the machine will have to be upgraded in order to satisfy the future need. You will have more flexibility in this regard with the distributed approach. As additional capacity is needed, systems can be added to the network and/or function can be reallocated to existing boxes.

  It should be noted that most shops underestimate the growth requirements of the LAN function.

1.1.5 Cost Constraints

Obviously, cost is always a factor in selecting new equipment. However, you should consider more than just the initial cost of the system. You should also examine how flexible the system must be in order to adapt to changing requirements. This is key in today′s LAN environment where change is occurring at an ever increasing rate. Some potential questions to ask are:

- Is the system being purchased to fulfill a very specialized role?
- Can it be upgraded for greater capacity or to incorporate new technology?
- Can I find third party adjunct products that can be incorporated into the system?
- What resources will be needed to manage the system?

1.2 The Selection Process

After you have collected requirements in each of the preceding areas and rolled them up into a set of overall system requirements, it is time to begin evaluating hardware solutions. The remainder of this book provides the details that you will need in order to evaluate the IBM line of PC Servers. Part II discusses in detail the technology used in implementing the product line. Part 3, "Server Products" on page 99 presents the current line and gives the specifications for each specific model in the line.

Part 2. Server Technology

Chapter 2. Introduction

In today's environment, it is often hardware and software technology that makes or breaks the usefulness of a computer system. This is especially so in a server. It is important that a server use technology to maximize the business advantage that it offers for two, somewhat similar reasons.

The first reason revolves around getting the most out of technology. Every single component in the server you are about to purchase to run your business should work as best it can, without having to slow down or wait for other components (or at worst, wait the least amount of time within technological and financial boundaries). It is important that every sub-system in the server be tuned and balanced. If the CPU offers high performance, the memory access, disk access and network access must be on-par. Otherwise you are simply not getting the most out of your acquisition.

The second reason is all about longevity of technology. Everyone knows that when you buy technology (be it a car, computer or whatever), the moment it leaves the shop floor there is something new that is better than the old one. The secret to buying technology and getting value for money is to ensure that what you buy will go the distance, that is, it has enough expansion capability and upgradability to be *future proof*. If your business is likely to double in the next three years and you want to buy a machine that will be useful to your business all through that time, then you must choose a machine that can expand to meet your growth requirements.

This part of the book, Part 2, describes what technology is available to you so that you can go into a purchase decision with your eyes open. If you understand the technology, then it is easier to make the most appropriate decision to meet your business's requirements.

Chapter 3. Processor and Bus Technology

This chapter describes the technology used in the processors and I/O buses of PC servers.

Note: In this book, MB refers to megabyte and Mb refers to megabit.

3.1 Processors

The microprocessor is the central processing unit (CPU) of the server. It is the place where most of the control and computing functions occur and where all operating system and application program instructions are executed. Most information passes through it, whether it is a keyboard stroke, data from a magnetic disk unit, or information from a communication network.

The processor needs data and instructions for each processing operation that it performs. Data and instructions are loaded from memory into data-storage locations, known as *registers*, in the processor. These registers are also used to store the data that results from each processing operation, until the data is transferred to memory.

The microprocessor is packaged as an integrated circuit that contains one or more arithmetic logic units (ALUs), a floating point unit, on-board cache, registers for holding instructions and data, and control circuitry.

Note: The ALUs and the floating point unit are often collectively referred to as *execution units*.

3.1.1.1 Clock Rate

A fundamental characteristic of all microprocessors is the rate at which they perform operations. This is called the clock rate and it is measured in millions of clock cycles per second or megahertz (MHz). The maximum clock rate of a microprocessor is determined by how fast the internal logic of the chip can be switched. As silicon fabrication processes are improved, the integrated devices on the chip become smaller and can be switched faster. Thus, the clock speed can be increased.

For example, the Intel Pentium Pro processor in the IBM PC Server 704 operates at clock speeds as high as 200 MHz. This processor is based on a fabrication process where transistors on the chip have a channel width of 0.35 microns. 1 micron is one millionth of a meter. As a comparison, a human hair is between 50 (blonde) and 100 microns (black) in diameter and the wavelength of visible light is 0.4 to 0.7 microns. The process used to produce such a device is called a *0.35 micron BiCMOS process*.

Early Pentium processors were based on a 0.8 micron process and could only be clocked at up to 66 MHz.

The clock rate of the external components may be different from the rate at which the processor is clocked internally. This is a common and efficient systems-design technique when faster external logic components are not available or are prohibitively expensive. Table 3 on page 16 shows the internal and external processor speeds of the Intel processors currently used in IBM servers.

Table 3. Processors Used in IBM Servers			
Processor	**Internal Speed**	**External Speed**	**Where Used**
Pentium P5	60 MHz	60 MHz	withdrawn model
Pentium P54C	75 MHz	50 MHz	withdrawn model
Pentium P54C	90 MHz	60 MHz	withdrawn model
Pentium P54C	100 MHz	66 MHz	withdrawn model
Pentium P54C	133 MHz	66 MHz	PC Server 310, 320, 520
Pentium P54C	166 MHz	66 MHz	PC Server 310, 320, 520, 720
Pentium P54C	200 MHz	66 MHz	PC Server 720
Pentium Pro P6	166 MHz	66 MHz	PC Server 704
Pentium Pro P6	180 MHz	60 MHz	PC Server 325
Pentium Pro P6	200 MHz	66 MHz	PC Server 325, 330, 704
Note: P54C is the Intel code name for the 0.6 micron Pentium processor. P6 is the Intel code name for the 0.6 and 0.35 micron Pentium Pro processors. Refer to Part 3, "Server Products" on page 99 for information on IBM servers.			

It should be noted that while the internal and external clock speeds of the processor are important, they are not the only factors to consider when looking at server performance. Components such as the network interface card, disk controller and main memory also must be considered and, in fact, are often the performance bottleneck in PC servers.

For example, upgrading the processor from a Pentium 90 MHz to a Pentium 133 MHz will achieve little if the disk controller is the bottleneck in your server. Server configurations cannot be compared by looking at the processor speed alone.

3.1.1.2 External Interfaces

The processor data interface, or data bus, is the data connection between the processor and external logic components. The Pentium family of processors (including the Pentium Pro) uses a 64-bit data bus, which means that they are capable of reading in 8 bytes of data in one memory cycle from the processor main memory. The older Intel 486 processor family had a data bus of only 32 bits, which limits its memory cycles to 4 bytes of data per cycle.

The width of the processor address interface, or address bus, determines the amount of physical memory the processor can address. A processor with a 24-bit address bus, such as the i286 class of processors, can address a maximum of 16 MB of physical memory. Starting with the i386 class of processors, the address bus was increased to 32 bits, which correlates to 4 GB of addressability.

3.1.2 Intel Pentium

The Pentium is the successor to the Intel 486 family of processors. It has a 32-bit address bus and 64-bit data bus. It has internal split data and instruction caches of 8 KB each. It employs new architectural features and enhancements to the 486 architecture that results in performance which is 3 to 5 times faster (5 to 10 times the floating point intensive applications) than a 33 MHz i486DX and 2.5 times when faster compared to the 66 MHz i486DX2 CPU. The more important features of this class of processor are discussed below.

3.1.2.1 Superscalar Architecture

The Pentium incorporates a superscalar architecture, built around two instruction pipelines, each capable of performing independently. These pipelines allow the Pentium processor to execute two integer instructions in a single clock cycle, nearly doubling the chip′s performance relative to an i486 chip operating at the same speed.

The Pentium processor′s pipelines are similar to the single pipeline of the i486 CPU, but they have been optimized to provide increased performance. Like the i486 CPU pipeline, the pipelines in the Pentium processor execute integer instructions in five stages: Pre-fetch, Instruction Decode, Address Generate, Execute, and Write Back. Instructions are moved through the pipeline such that each stage is working on different instructions simultaneously. As an instruction is passed from the pre-fetch to the decode stage, the pre-fetch is already retrieving the next instruction into the pipeline. The instructions are moved in a like manner through all the stages until they are *retired* in the final stage.

In many instances the Pentium processor can issue two instructions at once, one to each of the pipelines, in a process known as *instruction pairing*. Each pipeline has its own ALU (arithmetic logic unit), address generation circuitry and interface to the data cache.

3.1.2.2 On-Chip Caches

While the i486 incorporated a single 8 KB cache, the Pentium features two 8 KB caches, one for instructions and one for data. These caches act as temporary storage for instructions and data obtained from slower, main memory or off-chip level 2 cache.

The use of separate caches for instructions and data provides increased performance and faster throughput compared to the i486 microprocessor. For example, with a single cache, conflicts can occur between instructions being pre-fetched and data being accessed. Providing separate caches for instructions and data precludes such conflicts and allows both to take place simultaneously.

The Pentium caches are two-way set-associative caches organized with 32-byte lines. This allows the cache circuitry to search only two 32-byte lines rather than the entire cache. The use of 32-byte lines (up from 16-byte lines on the i486DX) is a good match for the Pentium processor′s bus width (64 bits) and burst length (four chunks).

When the circuitry needs to store instructions in a cache which is already filled, it discards the least recently used (LRU) information, according to an algorithm implemented in hardware, and replaces it with the new information.

When data is removed from the data cache, it has to be written back into main memory. Two techniques are available for writing cache data to memory: *write-back* and *write-through*. Write-back caching provides better performance than simpler write-through caching, in which the processor writes data to the cache and main memory at the same time. The instruction cache is a write-through cache and the data cache is a write-back design.

The data cache has two interfaces, one to each of the pipelines, which allows it to provide data for two separate operations in a single clock cycle.

To ensure that the data in the cache and in main memory are consistent with one another (especially of concern with multiprocessor systems), the data cache implements a cache consistency protocol known as *MESI* (Modified, Exclusive, Shared, Invalid). This protocol defines four states, which are assigned to each line of the cache based on actions performed on that line by a CPU. By obeying the rules of the protocol during memory read/writes, the Pentium processor maintains cache consistency and circumvents problems that might be caused by multiple processors using the same data. Refer to 4.2.2.1, "MESI Protocol" on page 39 for more information.

The Pentium processor also increases performance by using a small cache known as the Branch Target Buffer (BTB) to provide dynamic branch prediction. When an instruction leads to a branch, the BTB remembers the instruction and the address of the branch taken. The BTB uses this information to predict which way the instruction will branch the next time it is used, thereby saving time that would otherwise be required to retrieve the desired branch target. When the BTB makes a correct prediction, the branch is executed without delay, which enhances performance.

The combination of instruction pairing and dynamic branch prediction can speed operations considerably. For example, a single iteration of the classic Sieve of Eratosthenes benchmark requires six clock cycles to execute on the i486. The same code executes in only two clock cycles on the Pentium.

3.1.2.3 Floating Point Unit

The floating point unit (FPU) in the Pentium has been completely redesigned over the FPU in the i486. It incorporates an eight-stage pipeline, which can execute one floating point operation every clock cycle (it can execute two floating point operations per clock when the second instruction is an Exchange). The first four stages of the FPU pipeline are the same as that of the integer pipelines. The final four stages consist of a two-stage floating point execute, rounding and writing of the result to the register file, and error reporting. The FPU incorporates new algorithms that increase the speed of common operations (such as ADD, MUL, and LOAD) by a factor of three.

3.1.2.4 Data Integrity

Pentium chips implement parity checking at the external interfaces and also on the on-chip memory structures of cache, buffers, and microcode ROM.

For situations where data integrity is especially crucial, the Pentium supports Functional Redundancy Checking (FRC). FRC requires the use of two Pentiums, one acting as the master and the other as the checker. The two chips run in tandem, and the checker compares its output with that of the master Pentium to ensure that errors have not occurred. The use of FRC results in an error detection rate that is greater than 99 percent.

Pentium chips also execute a built-in self test upon power-on and reset which tests 70 percent of the circuitry.

3.1.3 Intel Pentium Pro

The Pentium Pro, code named the P6, is the latest of Intel′s mainstream processors. While offering backward compatibility to previous x86 processors, it provides a number of enhancements that allow the processor to deliver performance 40 to 60 percent faster than a Pentium processor of the same clock speed:

- Dynamic Execution

- More Integration
- Higher clock speeds

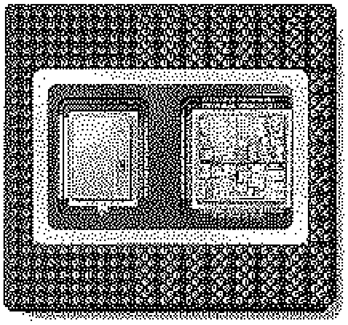


Figure 1. The Intel Pentium Pro. On the right is the Pentium Pro die with 5.5 million transistors. On the left is the on-chip 256 KB of level two cache with 15.5 million transistors (31 million for the 512 KB version).

3.1.3.1 Dynamic Execution

Like previous generations of Intel processors, the Pentium Pro uses a pipelined architecture. The Pentium Pro, however, increases the number of stages in each pipeline to 14. Each stage has been reduced in complexity, which enables it to be executed faster. This allows the chip to be clocked at a faster rate.

The Pentium Pro designers also added two more execution units for a total of 5 and another pipeline for a total of three. This classifies the chip as a three-issue superscalar architecture. However, what is most notable is the manner in which the separate pipelines are utilized to maximize performance.

Dynamic Execution is the term Intel uses to describe this optimization. It is actually a combination of three different techniques: multiple branch prediction, data flow analysis, and speculative execution.

Multiple branch prediction allows the processor to anticipate jumps in the instruction flow and to predict where the next instructions are located in memory.

Data-flow analysis is used to determine the most likely sequence of instructions. Up to 30 instructions are read in advance to determine the best approach to execution. Data based on previous branching is also used to determine the most likely sequence.

Speculative execution or out-of-order execution means that the instructions in the pipeline are examined and then re-ordered such that they are executed in the most efficient sequence. This is not necessarily the order in which they were coded. It is speculative since they are executed with the chance that a branch in the program may cause that instruction not to be needed.

Once all the instructions in the actual sequence are performed using out-of-order execution, the final result is known and the remaining false branches are discarded.

With wide (three-issue) and deep (14-stage) pipelines, there are instructions that can cause relatively long delays. Examples of these are I/O instructions and segment loads. These are unavoidable and require the entire pipeline system to drain (that is, to allow all instructions executed out of order to complete) prior to executing the instruction in-order.

Another example is the use of instructions that attempt reads from partially written registers. The prime example of this 16-bit code. If a partial write of a register is near or following a full read of the same register, then it is likely that both instructions will be in the pipeline(s) at the same time. In this instance, the processor will have to wait until the partial write is completely through the pipeline until the full read can occur.

It is for this reason that 16-bit code can perform very poorly on the Pentium Pro (often *worse* than on the Pentium). 16-bit code contains many partial register references as well as other stalling instructions such as segment loads.

3.1.3.2 Packaging

Like the Pentium, the Pentium Pro has 8 KB of cache for instructions and 8 KB of cache for data. This level one cache is the first point outside of the main storage processor (yet still on-chip) from where instructions and data are fetched. The Pentium Pro also has 256 KB or 512 KB of level 2 cache on-chip. In the PC Server 325 and 330,the Pentium Pro 200 MHz with 256 KB cache is used, while the PC Server 704 is using the 200 MHz with 512 KB cache. The biggest advantage of an integrated level 2 cache is that it operates at the internal speed of the processor: 150 MHz and higher.

The external interfaces of the P6 (the address and data buses) are isolated from the internal CPU bus via an integrated on-chip bus interface unit. These allow the processor to issue requests for data or instructions from external memory and not have to wait for the response. This has an additional advantage in that it can tolerate (relatively) large delays, especially important in multiprocessor designs.

The external speed of the processor has been set at 60 MHz (or 66 MHz) at 64-bits wide to keep the cost of the support logic down. However, by moving the level 2 cache on-chip, only external memory and I/O are accessed at the slower external speed.

As shown in Figure 2 on page 21, the processor has built-in logic for the support of up to four-way SMP, including an advanced programmable interrupt controller (APIC) and implementation of the MESI protocol for sharing data between multiple caches. (Refer to 3.2.2, “Symmetric Multiprocessing” on page 22 and 4.2.2.1, “MESI Protocol” on page 39 for more information.)

It is this implementation that is in the PC Server 704. Since all the electronics are on-chip, an upgrade from two-way to three-way is achieved by simply adding the third Pentium Pro processor.

Table 4 (Page 1 of 2). Pentium Pro Summary

Processor Speed	**150 MHz**	**166 MHz**	**180 MHz**	**200 MHz**	**200 MHz**
Bus Speed	60 MHz	66 MHz	60 MHz	66 MHz	66 MHz
L2 Cache	256 KB	512 KB	256 KB	256 KB	512 KB
Manufacturing Process	0.6 micron	0.35 micron	0.35 micron	0.35 micron	0.35 micron
CPU Voltage	3.1V	3.3V	3.3V	3.3V	3.3V

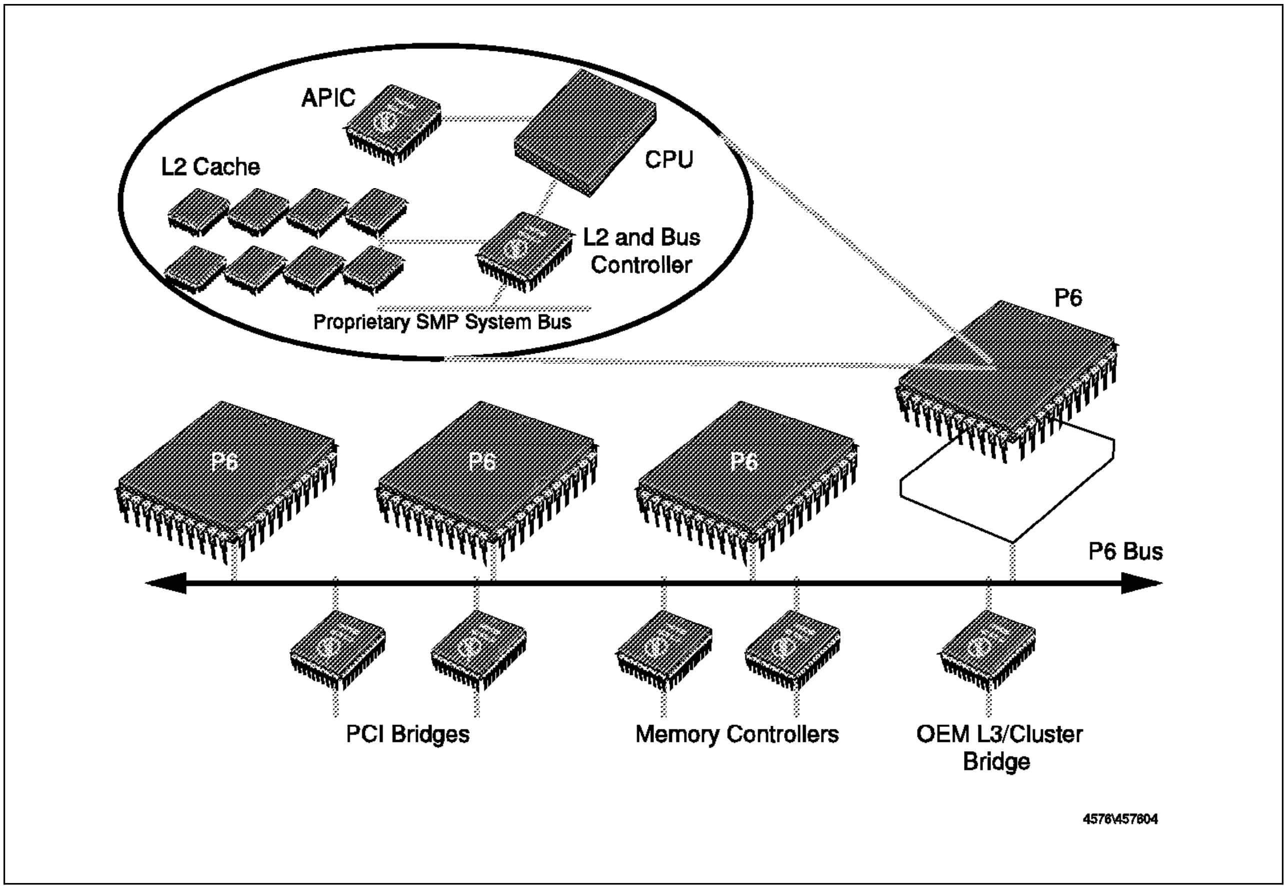


Figure 2. The Pentium Pro in a Four-Way SMP Configuration

Table 4 (Page 2 of 2). Pentium Pro Summary					
Processor Speed	**150 MHz**	**166 MHz**	**180 MHz**	**200 MHz**	**200 MHz**
Maximum Power Usage	29.2W	35.0W	31.7W	35.0W	37.9W
SPECint95 Rating	N/A	7.11	7.28	8.2	8.58
SPECfp95 Rating	5.42	6.21	5.59	6.21	6.48

3.2 Multiprocessing

Multiprocessing uses two or more processors in a system to increase throughput. Multiprocessing yields high performance for CPU-intensive applications such as database and client/server applications.

There are two types of multiprocessing:

1. Asymmetric multiprocessing
2. Symmetric multiprocessing

3.2.1 Asymmetric Multiprocessing

In asymmetric multiprocessing the program tasks (or threads) are strictly divided by type between processors and typically, each processor has its own memory address space. These features make asymmetric multiprocessing difficult to implement. Figure 3 and Figure 4 are two examples of asymmetric multiprocessor configurations.

The PS/2 Server 195 and Server 295 were examples of servers using asymmetric multiprocessing.

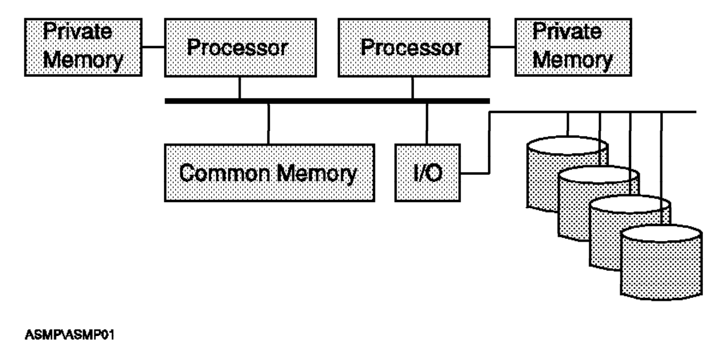


Figure 3. Asymmetric Multiprocessing Example 1. This configuration has multiple memory units with some of those not shared by all processors.

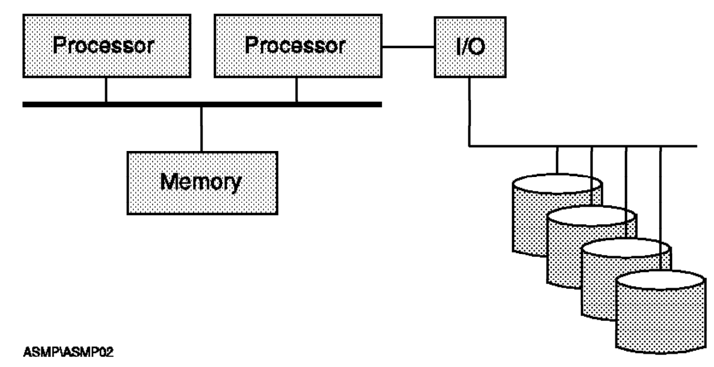


Figure 4. Asymmetric Multiprocessing Example 2. This configuration has one processor doing all I/O.

3.2.2 Symmetric Multiprocessing

Symmetric multiprocessing (SMP) is the most common configuration of multiple processors. A typical SMP configuration has the following items:

- It has multiple processors and exactly one of everything else: memory, I/O subsystem, operating system, etc.
- The processors are symmetric, that is, they can do anything the others can. Each can look at or alter any element of memory, and each can do any kind of I/O.

It is *symmetrical* because the view from any processor of the rest of the system is exactly the same. Figure 5 on page 23 shows a typical implementation of SMP.

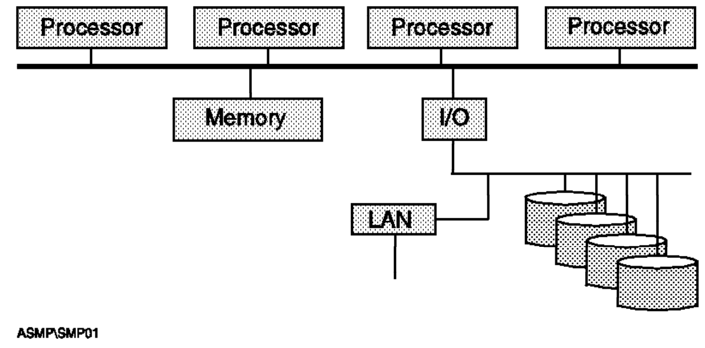


Figure 5. A Typical SMP Configuration

SMP is easier to implement in operating systems and is the method used most often in operating systems that support multiple processors. Operating systems that support SMP include:

- OS/2 Warp Server 4.0
- OS/2 for SMP 2.11
- Windows NT 4.0
- Novell NetWare 4.1 SMP
- Novell UnixWare SMP 2.0
- SCO Open Server 5.0 with SCO MPX 3.0
- Banyan Vines

SMP also yields some level of fault tolerance because if a processor fails, the others may still function. A reboot will still be required due to the fact that the design of an SMP configuration has every other component shared, including memory. If a processor fails, then it will invariably leave the memory in an unpredictable state, which could cause more problems for the remaining CPUs. Depending on the implementation of SMP on a server, the failed processor may or may not have to be physically removed before the system can restart.

A very important consideration of SMP systems is the bandwidth of the other system components. Consider an analogy of a dog with two heads (a savage multi-headed pooch, as it were).

It has the ability to eat (process) twice as much food (data) as a normal single-headed dog. However, to maximize the advantage of having two heads, it needs to have a throat (bus), stomach (disk) and digestive system (network) large enough to handle the output of the two heads. The alternative is for the heads to not eat as fast or to wait (wait state) for there to be sufficient swallowing capacity (bandwidth). If more heads were to be added, then the dog could eat a lot faster, but if the rest of the dog couldn′t handle the throughput, then the heads would have to wait even more and the benefits of the extra heads would not be realized (and if the dog′s owner can′t feed it fast enough, he may get his hand bitten off).

The point is that for SMP machines to be effective, the other subsystems have to have sufficient bandwidth to keep up with the multiple processors. For example, the IBM PC Server 720 implements a memory bus that has a bandwidth of 400 MBps. The IBM Streamer family of LAN adapters can process data from the network at the speed of the media without placing an extra burden on the CPUs. An IBM RAID subsystem utilizing data striping can provide very high bandwidth in SMP environments. These factors are very important when evaluating SMP systems.

The IBM PC Server 320, 325, 330, 520, 704 and 720 support SMP.

The Server 320 and 520 offer two-way SMP using a shared cache design. This design has one cache memory, which is accessed by the two processors.

The Server 720 uses a design where each processor has its own 512 KB (166 MHz models) or 2 MB (200 MHz models) cache.

The Server 325, 330 and 704 also use an independent cache design. However, the cache is integrated into the Pentium Pro processor. Refer to 4.1, "Cache" on page 34 for further information about the SMP designs of IBM servers and discussions of these alternatives in cache design.

3.3 Bus Architectures

There are a number of bus architectures implemented in IBM PC servers:

- IBM PC/AT Bus
- EISA
- Micro Channel
- PCI
- C-BUS II
- USB

Each of these is discussed in this section. The VESA local bus is also discussed here for completeness, although it is not implemented in any IBM server.

3.3.1 IBM PC/AT Bus

The IBM PC/AT bus, often called the Industry Standard Architecture (ISA) is not really an architecture at all but a de facto standard based on the original IBM PC/AT design. The main characteristics of this bus include a 16-bit data bus and a 24-bit address bus. The bus speed is limited to 8 MHz and it does not allow for DMA and bus masters in its original form. It does not support automatically configuring adapters and resolving resource conflicts among adapters, nor does it allow for sharing of interrupts. Nonetheless, it was an extremely successful design and even with these disadvantages, it is estimated that the ISA bus is in 70 percent of the PCs manufactured today.

A combination PCI/ISA bus architecture is used in the IBM PC Server 310.

3.3.2 EISA Bus

The Extended Industry Standard Bus Architecture (EISA) is a 32-bit superset of the ISA bus providing improved functionality and greater data rates while maintaining backward compatibility with the many ISA products already available.

The main advancements of the EISA bus are 32-bit addressing and 16-bit data transfer. It supports DMA and bus master devices. It is synchronized by an 8.33 MHz clock and can achieve data transfer of up to 33 MBps. A bus arbitration scheme is also provided that allows efficient sharing of multiple EISA bus devices. EISA systems can also automatically configure adapters.

Several models of the IBM PC Server 320, 325, 330, 520 and 704 implement the EISA architecture.

3.3.3 Micro Channel Bus

The Micro Channel Architecture (MCA) was introduced by IBM in 1987. Micro Channel is an improvement over ISA in all of the areas discussed in the previous section on EISA. In addition, it supports data streaming, which is an important performance feature of the MCA architecture, providing up to 80 MBps data transfers. The MCA bus operates at 10 MHz.

3.3.3.1 Data Streaming

The data streaming transfer offers considerably improved I/O performance. In order to understand data streaming transfers, we need to see how data is transferred between Micro Channel bus master adapters and memory.

The standard method of transfer across the Micro Channel is known as basic data transfer. In order to transfer a block of data in basic data transfer mode, an address is generated on the address bus to specify where the data should be stored; then the data is put on the data bus.

This process is repeated until the entire block of data has been transferred. Basic data transfer on the Micro Channel runs at 20 MBps (each cycle takes 200 nanoseconds, and 32 bits or 4 bytes of data are transferred at a time).

However, in many cases, blocks transferred to and from memory are stored in sequential addresses, so repeatedly sending the address for each 4 bytes is unnecessary. With data streaming transfers, the initial address is sent, and then the blocks of data are sent; it is then assumed that the data requests are sequential.

Micro Channel supports another mode of data streaming whereby the address bus can also be used to transfer data. In this mode, after the initial address is presented during the first bus cycle, the address bus is then multiplexed to carry an additional 32 bits of data. This results in an effective data transfer rate of 80 MBps.

Data streaming, as well as improving the data transfer rate, also provides a more efficient use of the Micro Channel. Since MCA operations complete in a shorter amount of time, the overall throughput of the system is increased.

Data streaming is useful for any adapters that perform block transfers across the Micro Channel such as the IBM SCSI-2 Fast/Wide Streaming RAID Adapter/A.

Micro Channel is implemented in the PC Server 720 and in some models of the IBM PC Server 320 and 520.

3.3.4 VESA Local Bus

As a result of shortcomings and development challenges of the ISA bus, the VESA local bus was released in August 1992, through efforts of the Video Electronics Standards Association (VESA). Because of its simplicity and ease of implementation, it quickly gained a great deal of industry support and momentum. While the VESA local bus was originally conceived as a way to boost graphics performance for desktop systems with the AT bus, manufacturers recognized the advantages of having an industry-standard local bus and crisp specifications for attaching other types of components. The VESA local bus is a true local bus; it is directly connected to the processor-memory bus using very simple bridge circuitry. Its frequency is synchronized with that of the external processor bus and many of its signals are virtually identical to those of the i486 microprocessor.

The VESA local bus specification also included a connector design. This capability solved the upgrade problem for local bus components in addition to providing for additional I/O expansion. The intent was to add to, rather than replace, the expansion bus. It shares many signals with the expansion bus, allowing VESA local bus devices to take advantage of the circuitry already in place. The expansion slots in systems with the VESA local bus can generally be used by either AT or VESA local bus adapters.

In practice, however, the expansion capability of the VESA local bus is limited. On 33 MHz systems, only two expansion slots can be used if the graphics controller is also on the local bus. As the frequency increases, the number of devices that can be supported drops. At 50 MHz, only one component device (that is, the graphics controller) can be attached and no expansion slots can be used. Clearly, systems with the VESA local bus must also have expansion bus capabilities because of this limitation.

There are some other limitations to the VESA local bus. Unlike Micro Channel architecture, the VESA local bus specification did not include provisions for parity, automatic configuration and other manageability features. As a consequence, board makers have primarily used DIP switches and jumpers to set up VESA local bus configurations and resolve conflicts with other devices. Further, since it was originally designed as a 32-bit bus, the ability to operate efficiently with more powerful 64-bit CISC and RS/6000 microprocessors, such as the Intel Pentium and the PowerPC, is also reduced.

The VESA committee has recognized the need for a more powerful local bus and has released a new set of specifications. The new VESA local bus specification (Version 2.0) defines a 64-bit interface. It will also allow for more VESA local bus slots at a given frequency provided that a new low-capacitance design is used. In terms of compatibility, the existing VESA local bus adapters can be used in systems with the newer VESA local bus design. In addition, new 64-bit adapters, operating in the 32-bit mode, can be used in the older 32-bit VESA local bus.

3.3.5 PCI Local Bus

The PCI Local Bus Specification was developed by the PCI Special Interest Group (PCI-SIG), led by a group of companies including Compaq, Digital, IBM, Intel, and NCR. Introduced in 1992, the PCI bus architecture has quickly gained widespread industry acceptance in the three years since its first release. There are now over 300 companies in the PCI-SIG supporting the architecture and currently producing PCI products. The PCI local bus is expected to become the preferred local bus architecture for PCs. PCI has accomplished this primarily by offering the advantages of a high-performance local bus capability, along with a bus architecture that is independent of the processor architecture and implementation.

The goal for PCI was to create an industry standard local bus that could be used in systems from laptops to servers. It was envisioned not only as a system local bus that would serve as a common design point supporting different system processor architectures, but would also support these various processors as they evolve over time. This is much like operating systems, which have defined Application Program Interfaces (APIs), so that applications need not change with each generation of the operating system.

The group initially intended the PCI bus to support high-performance basic system I/O devices such as graphics adapters, hard disk controllers, and/or LAN adapters which would be integrated on the system board (no pluggable connectors), and communicate through the PCI bus. However, the PCI-SIG soon realized that the PCI bus needed the capability of supporting connectors for pluggable adapter cards. For example, graphics controller evolution doesn′t necessarily match mother board development, so providing for an upgrade of the graphics controller became a requirement. The second release of the PCI specification (Revision 2.0 in April of 1993) included upgrade capability through expansion connectors.

The PCI I/O adapter card design, drawing heavily on Micro Channel technology, was a size that would be least disruptive to the physical package of systems containing ISA, EISA, or Micro Channel I/O boards. The PCI bus transfers either 4 bytes or 8 bytes of data at a time at a clock speed of 33 MHz. This yields a peak local bus performance of 132 MBps for 4-byte (32-bit) transfers, and 264 MBps for 8-byte (64-bit) transfers.

The latest specification of the PCI bus (Revision 2.1) was released in June of 1995 and is a compatible superset of PCI, defined to operate up to a maximum clock rate of 66 MHz. The purpose of defining the 66 MHz PCI capability was to provide connectivity for higher bandwidth devices. Typical systems that provide a 66 MHz bus would also include one or more 33 MHz PCI buses.

The differences between a 66 MHz and a 33 MHz PCI bus are relatively minimal. The signal definitions, signal protocols, and connector layouts are the same. One of the pins on the PCI connector is changed from a grounded connector on a 33 MHz bus to an ungrounded pin on the 66 MHz adapters. This provides for a technique to differentiate between the two types of boards, while still maintaining backward compatibility. The specification allows for 33 MHz adapters to operate (at 33 MHz) in a 66 MHz system, as well as allowing 66 MHz adapters to operate at 33 MHz in an older system.

There is also an addition to the Configuration Status register to indicate support of 66 MHz. The programming models for 66 MHz and 33 MHz devices are the same. For additional details concerning the 66 MHz PCI bus definition, please see Chapter 7 “66 MHz PCI Specification,” in the PCI Local Bus Specification Revision 2.1, which was released in June of 1995.

The PCI bus is standard on all IBM PC Servers. The PC Server 720 also has a PCI-to-PCI bridge enabling a total of 7 PCI slots.

3.3.5.1 PCI Architecture

By providing high performance capability and processor architecture independence, the PCI local bus architecture is also attractive for server I/O subsystems. Servers typically have need for multiple, high-bandwidth devices.

The PCI local bus is a clock synchronous bus that operates at up to 66 MHz. It allows low latency random accesses. For example, at 33 MHz, as little as 60 ns is required for a master on the bus to access another device. At 66 MHz, as little as 30 ns is required for this access.

The basic transfer operation for the PCI bus is a burst operation, which allows a contiguous block of data to be transferred on the bus following an address. This results in an address phase on the bus followed by one or more data phases. Figure 6 shows this type of operation.

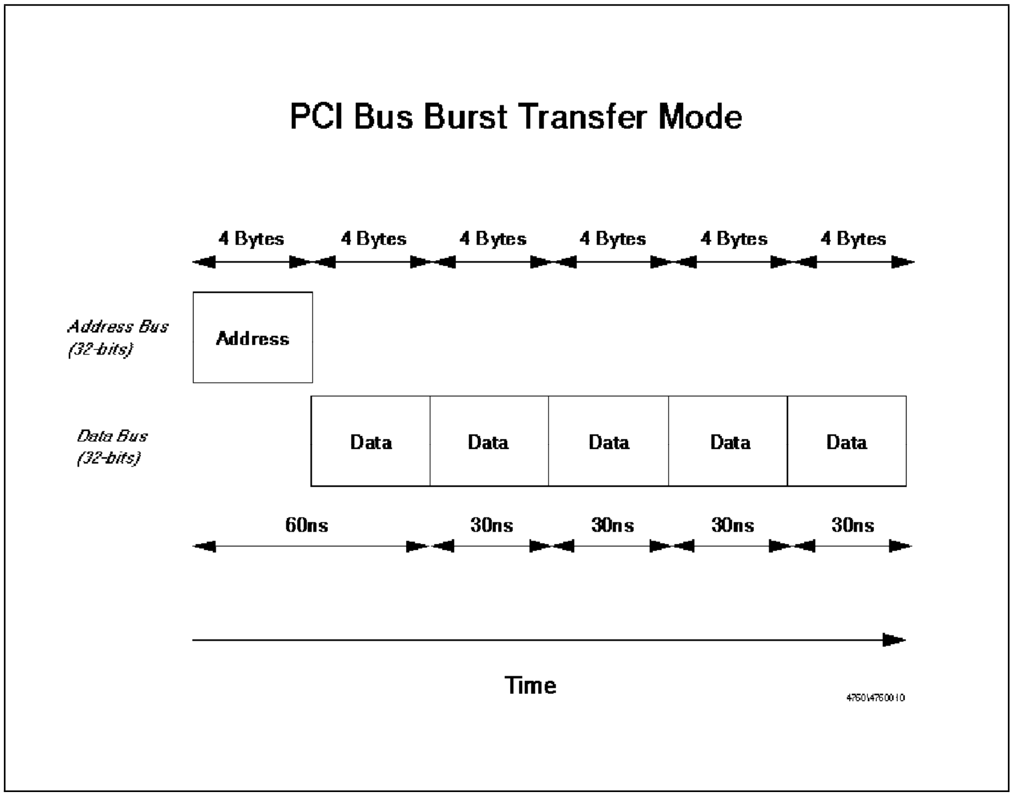


Figure 6. PCI Bus Burst Transfer at 33Mhz

The PCI local bus also provides for a streaming mode where the address and data bus are *multiplexed* to provide a 64-bit data path. Figure 7 on page 29 shows this type of operation.

The PCI local bus was optimized for direct attachment of single silicon chips to the bus (that is, no glue logic required) and provides connector definitions for pluggable adapter cards. The multiplexed address/data bus contributes

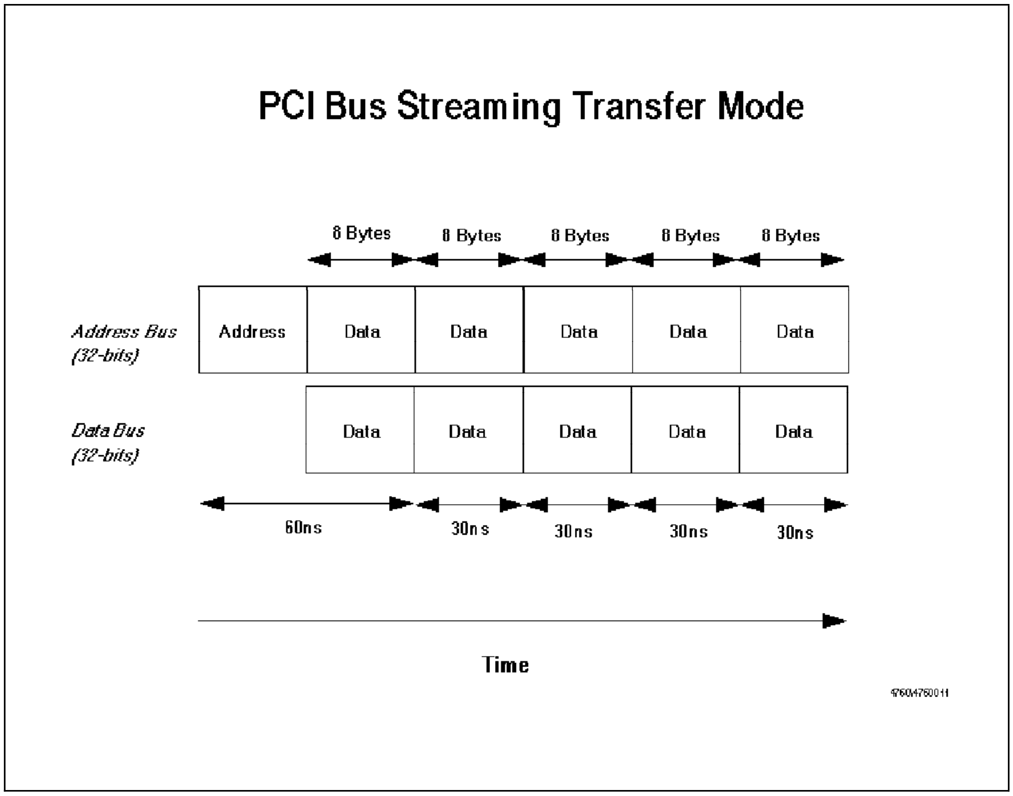


Figure 7. PCI Bus Streaming Mode at 33 Mhz

significantly to the low pin count allowing reduced packaging requirements. The physical dimensions for PCI adapter cards were defined to allow PCI cards to also fit into future ISA, EISA, and Micro Channel-based systems with minimal change to existing frame design concepts. Also for longevity, the PCI bus was defined with provision for processor architecture independence, support for 64-bit addressing, and both 5 volt and 3.3 volt signaling environments.

Because of its higher speeds, the PCI architecture only allows for 10 electrical loads on the bus which means, in practice, no more than four PCI adapter slots can be used. (A PCI adapter acts at two signal loads.)

System designers use two techniques that enable them to get around this limitation. The first is to drive multiple PCI buses off the processor bus. This is the technique used in the PC Server 704. Figure 8 on page 30 shows this implementation with the two Pentium Pro-to-PCI-Bus bridge chips and their corresponding PCI buses. With this implementation, each PCI bus offers the same level of performance.

The other method used to get around the bus-loading limitation is to use a PCI-to-PCI (PtP) bridge. This technique makes cascaded buses where each bridge creates another PCI bus that can handle another four I/O slots.

The PC Server 720 uses this method, which is illustrated in Figure 9 on page 31. Notice also that the Micro Channel bus is created via the same method using a PCI-to-Micro Channel bridge. One advantage of this type of design is that it works with several versions or generations of processors, without having to redesign the I/O subsystem for each new processor generation. The disadvantage is that each bridge chip creates additional logic propagation delay.

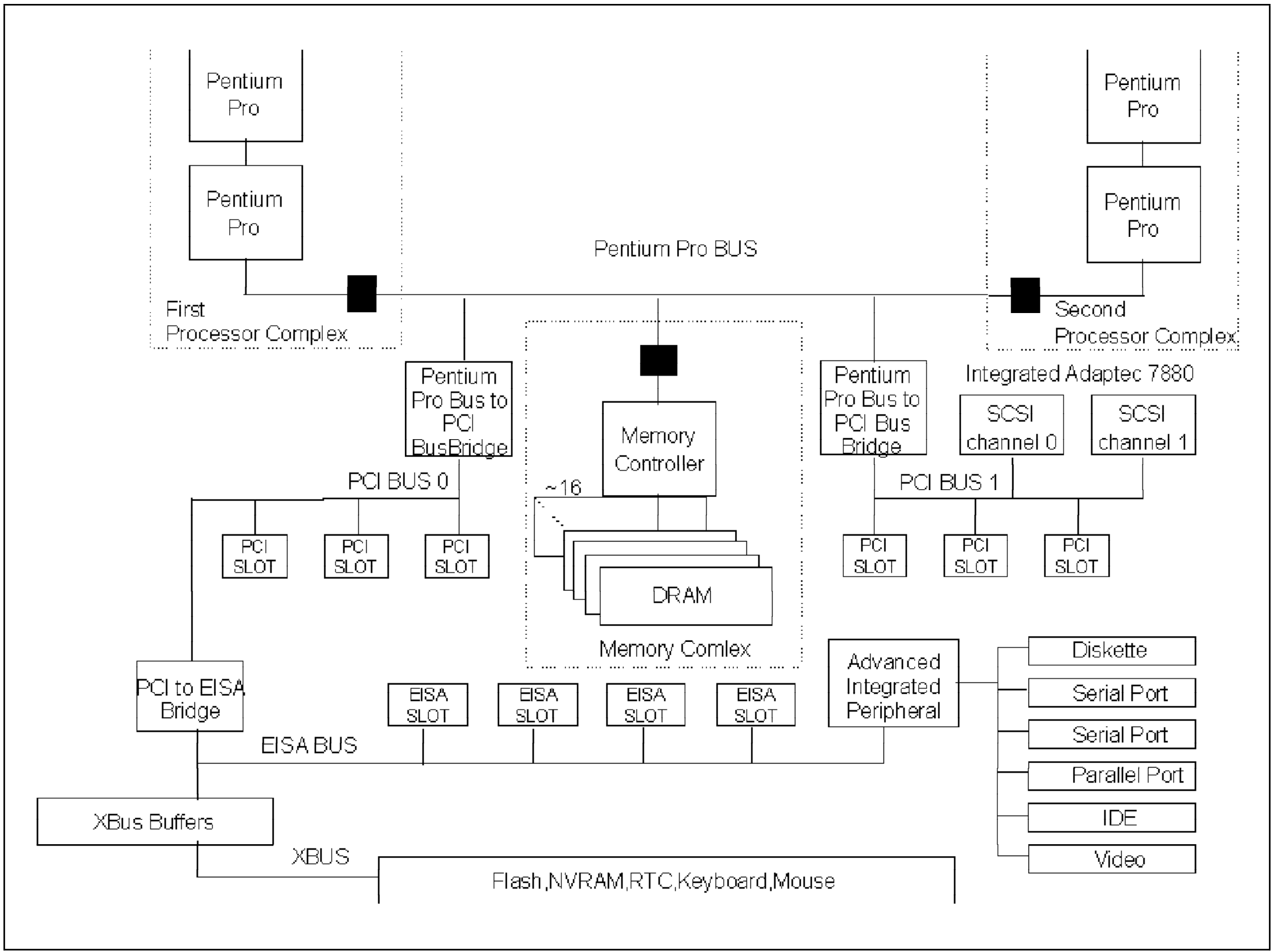


Figure 8. PCI Configuration in the PC Server 704

This is why devices attached to a bridged bus will operate at a slightly lower performance level than one on a PCI bus that is directly off the processor bus.

3.3.5.2 PCI Features

The PCI bus has many features which make it desirable to use. For example:

- PCI adapters are relatively simple to implement. For example, a bus master requires only 49 signals. This allows PCI bus interfaces to be implemented on smaller, lower cost logic chips which reduces adapter product costs. The PCI connector supports 3.3 volt power to the card, making the migration to higher density and lower power technologies easier.
- The high data rate and low latency are great for graphics subsystems. A 4-byte write can be completed in as few as two PCI clock periods (a PCI clock period is 30 nanoseconds at 33 MHz).
- System designers have flexibility when considering I/O expansion. As the number of electrical loads is a function of bus clock speed, if the number of pluggable boards is more important than bus data rate, the bus clock frequency can be reduced. This would mean, for example, that an additional one or two slots could be implemented if the bus clock speed was lowered from 33 to 25 MHz.
- The PCI bus is parity protected.

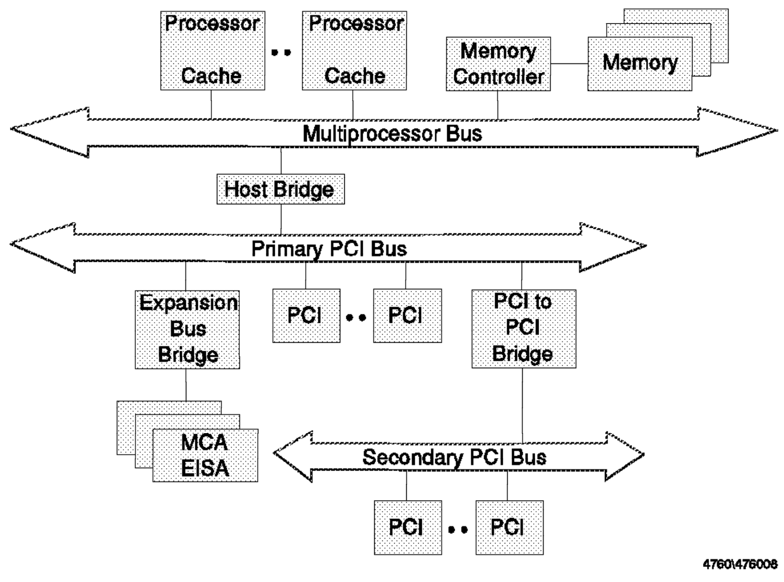


Figure 9. PCI Configuration in the PC Server 720

- Multiple bus masters are supported with the maximum number being determined by the arbitration logic implemented in the system.
- PCI masters have a latency timer that limits their bus ownership time when another master is requesting bus usage. The latency timer is set by the system when the device is configured.
- PCI devices are auto-configurable through a 256-byte configuration address space.

3.3.6 C-Bus II

The C-bus II is the multiprocessor bus implemented in the PC Server 720. It uses a symmetric multiprocessor (SMP) architecture allowing any process to execute on any CPU and for the process to be reassigned to another CPU to ensure effective load balancing. Refer to 3.2.2, “Symmetric Multiprocessing” on page 22 for more information on SMP.

The C-bus II in the PC Server 720 has the following overall characteristics:

- Processor Agility -- processing elements do not have to operate at the same speed
- Multiplexed, synchronous design
- 64-bit data bus operating at 50 MHz giving a bandwidth of 400 MBps
- Architectural limit of 32 GB of RAM
- Architectural limit of 15 system elements such as processor complexes, memory cards and bus bridges (these are usually circuit boards plugged into the passive backplane)
- Fully SMP, including I/O and interrupts
- Inherent cache coherency between processor complexes
- Provision for other I/O buses such as PCI and Micro Channel through bus bridges

Refer to 16.1.2, “Multiprocessor Bus” on page 154 for more information about how the C-bus II is implemented in the Server 720.

3.3.7 USB (Universal Serial Bus)

The Universal Serial Bus (USB) is a new serial interface standard for telephony and multimedia connections to the computer. The Universal Serial Bus port is a single connector for all types of devices that previously used serial, parallel and games ports. It will not accept standard serial devices. The USB connector uses Plug and Play technology. Each USB device is accessed by a unique USB address. A device called a hub is used to convert a single USB connector into multiple attachment points. A hub has seven ports where peripherals may be attached. The maximum speed of the USB is 12 Mbps.

The USB is implemented on some models of the PC Server 310.

Note: The USB is not designed to support high-speed disk subsystems.

Chapter 4. Memory and Cache Technology

The system design of PC servers (in fact all microprocessor-based systems) is centered around the basic memory access operation. System designers must always *tune* this operation to be as fast as possible in order to achieve the highest possible system performance.

Processor architectures always allow a certain number of clock cycles in order to read or write information to system memory. If the system design allows this to be completed in the given number of clock cycles, then this is called a zero wait state design.

If,for some,reason the operation does not complete in the given number of clock cycles, the processor must *wait* by inserting extra *states* into the basic operation. These are called *wait states* and are always an integer multiple of clock cycles. The challenge is that as each new generation of processors is clocked faster, it becomes more expensive to incorporate memory devices that have access times allowing zero wait designs.

For example, state-of-the-art Dynamic RAM (DRAM) has a typical access time of about 60 nanoseconds (ns). A 60 ns DRAM is not fast enough to permit a zero wait state design with a Pentium class processor. Static RAM (SRAM) has an access time of less than 10 ns. A 10 ns SRAM design would allow for zero waits at current processor speeds, but would be prohibitively expensive to implement as main memory. A basic trade-off that all system designers must face is simply that as the access time goes down, the price goes up.

The solution is to use DRAM for main memory, making a less expensive solution, and use a smaller amount of SRAM-based memory (cache) for higher-speed access of data, which is likely to be needed sooner. All servers (and most Pentium-based PCs and notebooks, for that matter) use cache in this manner. The size, speed and design of the cache varies but the chief criterion is to always reduce the *effective access time* of main system memory.

Another method used to increase access speed of main memory is to use EDO (Extended Data Out) DRAM. EDO DRAM increases performance by assuming that the next memory access will be in the same DRAM *row* as the previous one. This allows the chip to hold or *latch* the data at the chip′s output pins to be read by the processor at the same time that the processor is presenting the address to the next memory location.

The performance increase with EDO DRAM will vary from 5 to 30 percent for the memory subsystem. EDO DRAM memory may be implemented as non-parity or parity memory (see 4.3.1, “Parity Memory” on page 40 for more information about parity).

Other techniques, such as memory interleaving, are also used to increase CPU throughput and to achieve the required performance in high-end servers.

4.1 Cache

Caches are memory buffers that act as temporary storage places for instructions and data obtained from slower main memory. They use static RAM and are much faster than the dynamic RAM used for system memory (typically five to ten times faster). Thus, they reduce the number of clock cycles required for a memory access. However, SRAM is more expensive and requires more power, which is why it is not used for all memory.

Caches operate on the principle that many accesses to memory are sequential, for example, the execution of in-line code. Whenever the processor must perform external memory read accesses, the cache controller always pre-fetches extra bytes and loads them into the cache. When the processor needs the next piece of data, it is likely that it is already in the cache. If so, processor performance is enhanced; if not, the penalty is minimal.

Also, research has shown that when a system uses data, it will be likely to use it again. If this data resides in cache, then performance is further improved.

Caches are cost-effective because they are relatively small as compared to the amount of main memory.

There are several levels of cache implemented in IBM PC Servers:

1. Level 1 cache

   Level 1 cache is implemented directly into the processor chip. The Intel i486 incorporates a single 8 KB cache. The Intel Pentium and Pentium Pro processors have two 8 KB caches, one for instructions and one for data. Access to these on-board caches is very fast and consume only a fraction of the time required to access memory locations external to the chip.

2. Level 2 cache

   The second level of cache, called second-level cache or L2 cache, provides additional high speed memory to the L1 cache. If the processor cannot find what it needs in the processor cache (a first-level *cache miss*), it then looks in the additional cache memory. If it finds the code or data there (a second-level *cache hit*) the processor will use it and continue. If the data is in neither of the caches, an access to main memory must occur. The term *cache-hit ratio* refers to the percentage of memory requests that are satisfied by the cache versus those requests which have to be fulfilled by accesses to main memory. The higher the cache-hit ratio, the better.

   If a cache-miss occurs, then the data must be retrieved from main memory. If the cache is currently full, then one piece of cache (that is, one *line* of cache) must be removed from the cache so that a new piece can be brought in. The method to decide which line of cache to *evict* can be a complex algorithm but is generally based on a least recently used (LRU) algorithm.

   When cached data is modified, the master copy of that data must also be updated in main system memory. There are two ways that these *dirty* cache lines can be written to memory. This is called the write policy.

   a. Write-through

      With a write-through policy, when modified data is returned from the processor to cache, the cache controller will actually perform the write to main memory before signalling to the processor that the operation has completed. Write-through is used to ensure data integrity and would be used in conjunction with a write-through policy on disk controller cache to ensure complete data integrity from processor to disk.

b. Write-back (or lazy-write)

The write-back policy is the most commonly used write policy as it provides greater performance. Instead of writing the modified cache line back to memory immediately, it is written back at a later stage when the cache controller and cache-memory bus are idle. The advantage is that the processor does not have to wait for the write to memory operation to complete before continuing. This does not impact performance as much as you might think because on average, a system does about nine reads for every write operation.

L2 cache is standard in all IBM PC Server models. All servers with the Intel Pentium processor use SRAM chips for L2 cache implementation. The L2 cache in Pentium Pro systems such as the PC Server 704 is implemented on the processor itself. Within the Pentium Pro range, there are chips with either 256 KB or 512 KB of level 2 cache. See 3.1.3, "Intel Pentium Pro" on page 18 for more information on the Pentium Pro, and see Part 3, "Server Products" on page 99 for the PC Server range.

With all types of caching, more is not always better. For example, if the cache is too large or is not designed properly, the CPU will spend more time trying to find the data in cache than it would by accessing the data from main memory. Depending on the system, the optimum size of level 2 cache can be as little as 128 KB and upwards of 512 KB.

3. Level 3 cache

A third level of cache is implemented on the IBM PC Server 720. It is a small, highly customized cache used to increase the hit ratio of the L2 cache. It is 192 bytes organized as 6 lines of 32 bytes each. Because it stores recently evicted L2 cache entries, it is sometimes referred to as a *victim* cache.

Since this cache is directly connected to the L2 cache, it does not consume memory bus cycles in order to be read or written to. This allows data to be quickly cleared from L2 to L3 making room for new information and allowing modified data to be written to main memory later during free bus cycles. In addition, this information is still available to be quickly accessed if needed in future cycles.

This approach improves the overall cache hit ratio with little additional cost while saving significant clock cycles during the write-back or *eviction* process. Access to data in any of these three caches can be accomplished with zero wait states so the processor continues work on its applications without wasting time.

Note

This third level of cache implemented in the PC Server 720 is substantially different than the L3 cache that is implemented in other vendors′ systems. The L3 implementations serve completely different purposes and should not be confused.

4.1.1 Cache Implementations

Caches can be implemented using various techniques, some of which have significant performance implications. System performance of a cached design depends upon both how the cache was implemented as well as the application(s) which the machine will be running.

To say definitely whether one implementation is better than another for a given application, it is first necessary to understand how a cache works.

4.1.1.1 Cache Elements

The two main elements of a cache are:

- Cache lines
- Cache tags

Cache Lines: This is where the data is actually stored after it is fetched from main memory. The size of the line is always an integer multiple of the processor bus width in bytes. For example, in the PC Server 720, the cache line size is 32 bytes. This equates to four times the data bus width of 8 bytes (64 bits) for the Pentium and Pentium Pro processors. One of the main criterion for an efficient cache implementation is to balance the line size with the overall size of the cache.

Memory read and write operations between cache and main system storage usually occur in full cache lines although this is an implementation choice which must be made. It is more efficient to perform operations involving full cache lines if the bandwidth of the memory bus is adequate. This is of special concern in SMP environments where multiple CPUs may be loading cache lines at any one time. This is one of the reasons why the Corollary bus was chosen for the Server 720 design. The C-Bus II has a bandwidth of 400 MBps, which is more than adequate, even in a six-way SMP environment.

Cache Tags: Each line in the cache has a corresponding tag, that contains the high order address of the data in the cache line as well as status bits showing what kind of data it is and what the status of this data is. For example, this is the kind of data necessary to implement the MESI cache coherency protocol. Please see 4.2.2.1, “MESI Protocol” on page 39 for more information about this protocol.

The cache controller must look through the tags to see whether the desired information is in cache and whether it is valid for use by the processor.

The PC Server 720 uses a dual tag design in which two exact replicas of the tags are maintained. This allows the CPU and the bus interface unit to simultaneously examine the cache which allows for higher performance in the SMP environment.

4.1.1.2 Cache Mapping

Since there is always more main memory than cache, some scheme has to be employed to manage how the main memory gets allocated or *mapped* into available cache lines. The two most common schemes are:

- Direct mapping
- Set Associative mapping

Direct Mapping: Using this technique, every byte in main storage is directly mapped to a specific line in cache memory. This cache line is the only place where that byte of main memory can reside when cached. Since there is always much more main memory than cache memory, several main memory addresses must be mapped to the same cache address. (For example, if you have 16 MB of main memory and 256 KB of L2 cache, each cache address must be able to map to 16*1024 KB/256 KB = 64 different main memory addresses.)

This is simple to implement; you can easily see if a specific main storage address is in cache because that address can be in one and only one cache line. However, you can immediately see the problem. Only one of the 64 possible main memory addresses can be in that line of cache at any one time. If you have that line loaded with one of the 64 possible mappings and you need another mapping loaded, then you have to write the cache line back out (if it is modified data) and then load the cache line with the new data.

Note: In the C-Bus II architecture as implemented in the PC Server 720, the write of a dirty cache line is actually performed as part of the read operation. Therefore, it does not require a separate operation.

Set Associative Mapping: With a set associative cache, each block in memory can be stored in different locations in the cache. Two-way set associative gives each block in memory two cache line locations, and four-way set associative gives each block in memory four cache line locations. Fully associative allows each block in memory to be stored anywhere in cache.

The advantage of a set associative cache is that the most recently accessed information is more likely to be in cache. The disadvantage is that searching is slow since the processor must do more searching to see if the required data is in cache. The processor must now look in two, four, or all lines in L2 cache versus one line in a direct mapped design. It is also more complex to design and is therefore higher in cost.

Performance Trade-Offs: If data references are sequential, as in the execution of in-line code, then a direct-mapped cache is the desired approach. However, in multi-tasking environments, the processor is constantly switching between the different instruction streams needed to execute the multiple processes occurring in the machine. In this environment, a set associative cache normally yields better overall performance.

The caching scheme implemented in the PC Server 720 offers benefits of both types of mapping. The direct-mapped L2 cache allows zero-wait-state access to data in the cache. The L3 victim cache allows data to be cleared quickly from the L2 when switching between tasks and, at the same time, keeping that data available for future requests. This implementation allows the PC Server 720 to achieve the highest level of performance with manageable cost and complexity.

4.2 Cache in an SMP Environment

Within SMP designs, there are two ways in which a cache is handled:

1. Shared cache
2. Dedicated cache

4.2.1 Shared Cache

Sharing a single L2 cache among processors is the least expensive SMP design. However, the performance gains associated with a shared cache are not as great as with a dedicated cache. With the shared secondary cache design, adding a second processor can provide as much as a 30 percent performance improvement. Additional processors provide very little incremental gain. If too many processors are added, the system will run even slower due to memory bus bottlenecks caused by processor contention for access to system memory.

The IBM PC Server 320 and Server 520 support SMP with a shared cache. For performance reasons, IBM recommends that the L2 cache in the system be increased from 256 KB to 512 KB when a second processor is added.

Figure 10 shows a two-way SMP system with shared secondary cache.

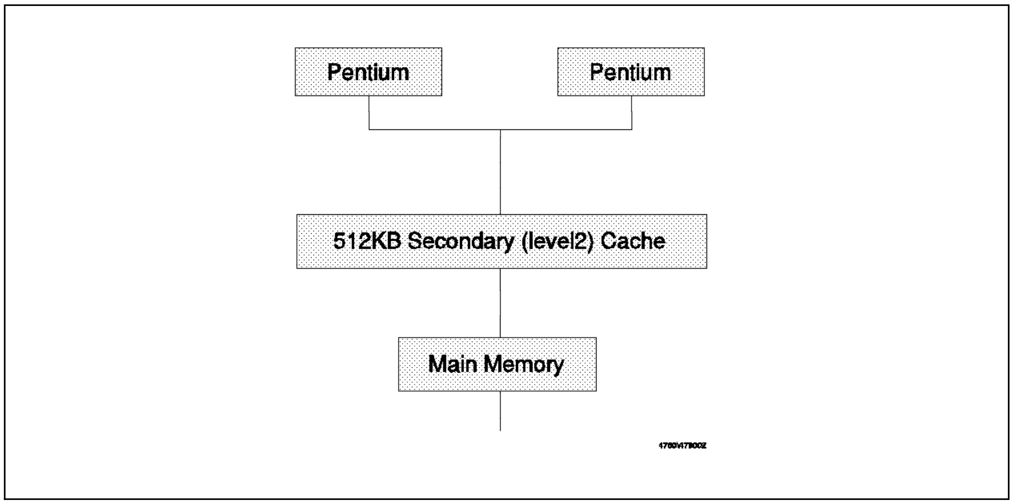


Figure 10. SMP Shared Secondary Cache

4.2.2 Dedicated Cache

Figure 11 on page 39 shows a two-way SMP system with dedicated secondary cache. This design provides for a significant performance boost over the shared cache for a couple of reasons. First, each processor having its own dedicated cache results in more cache hits than a shared L2 cache.

Second, and even more significant, if the L2 cache is integrated into the processor chip as is the case with the Pentium Pro, the speed to this cache is *much* faster than what can be obtained with a shared cache design. Remember, the main idea of the cache is to provide an area of storage that can be accessed more quickly than main storage.

The PC Servers 325, 330, 704 and 720 implement SMP with dedicated cache. The Server 325, 330 and 704 use on-chip L2 cache in the Pentium Pro processor. The Server 720 utilizes the Pentium processor with external L2 cache.

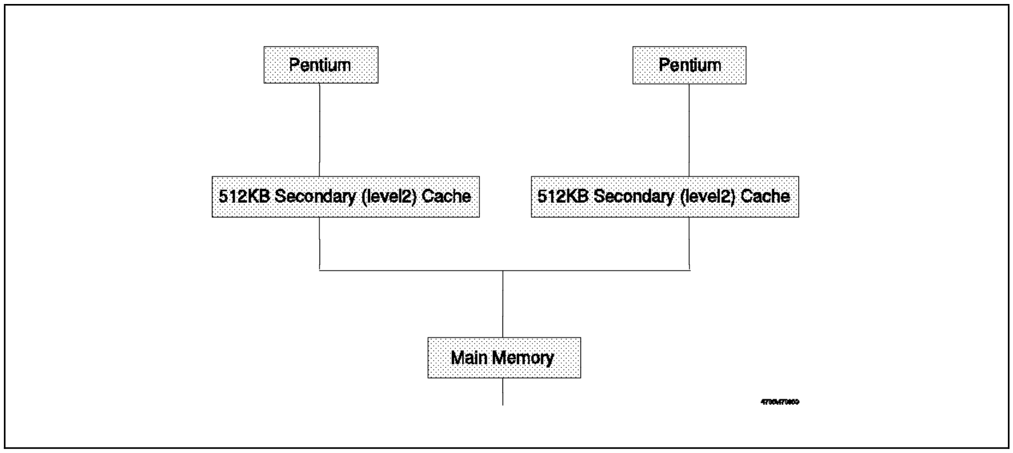


Figure 11. SMP with Dedicated Secondary Cache

However, dedicated caches are also more complicated to manage. Care needs to be taken to ensure that a processor needing data always gets the latest copy of that data. If this data happens to reside in another processor′s cache, then the two caches must be brought into sync with one another.

The cache controllers maintain this *coherency* by communicating with each other using a special protocol called MESI (Modified, Exclusive, Shared, Invalid).

4.2.2.1 MESI Protocol

One of the problems SMP introduces is that the same data from main memory can be held at any time in the caches of multiple processors. If one of these processors updates that information, the change needs to be communicated before another processor operates on what is now invalid data. MESI is a protocol that ensures that the data across the caches remains in sync. The name MESI stands for Modified, Exclusive, Shared, Invalid, which represent the four different states that cached data can be in at any time.

When data is first read out of main memory and into a processor′s cache, the data is flagged as *exclusive*. If another processor reads the same data into its cache, the data would then be flagged as *shared* in both caches. The data may be read by either processor without changing the state of the flag.

However, if a processor writes to (that is, modifies) the data in its cache, then the flag changes. If the data was previously marked as exclusive, it is now marked as *modified*. On the other hand if it was marked as shared, it is now marked as *invalid* and the modified data is written out from cache to main memory. When another processor tries to read the data that had previously been marked as shared in its cache, it will find that it is now marked as invalid and will have to fetch the modified data from main memory.

Note: On the C-Bus II, the operation is slightly different. During a write cycle to a shared line, the L2 cache sends out a Read Exclusive cycle on the C-Bus II (which invalidates all the other shared caches for that line) and reloads the data from main memory into the L2 again. The modified data is then written back to the L2. This extra step of reloading the cache line is necessary to inform the other caches that the line is about to be written. This is the only way the other L2s find out about a write.

MESI is designed to minimize the movement of data between main memory and cache. Modified data does not need to be written back to main memory until another processor requires access to the same data. The cache controllers use a technique called *bus snooping* to correctly flag the data held in the different caches and main memory. With snooping techniques, each processor and memory device on the bus monitors the address and data lines to determine when data held in any of the processors′ caches is being modified. They can then update the cache tags accordingly.

4.3 Memory

There are four different memory implementations used on IBM servers:

1. Parity memory
2. Error Correcting Code-Parity (ECC-P)
3. Error Correcting Code (ECC) memory
4. ECC Memory on SIMMs (EOS) Memory

4.3.1 Parity Memory

Standard parity memory, Single In-Line Memory Modules (SIMMs), have 32 bits of data space and 4 bits of parity information (one check bit/byte of data). The 4 bits of parity information are able to tell you an error has occurred but do not have enough information to locate which bit is in error. In the event of a parity error, the system generates a non-maskable interrupt (NMI) that halts the system. Double-bit errors are undetected with parity memory.

Parity memory is shipped with some models of the PC Server 310 although all of these can be upgraded to use EOS memory.

4.3.2 Error Correcting Code (ECC)

The requirements for system memory in PC servers has increased dramatically over the past few years. Several reasons include the availability of 32 bit operating systems and the caching of hard disk data on file servers.

As system memory is increased, the possibility for memory errors increase. Thus, protection against system memory failures becomes increasingly important. Traditionally, systems that implement only parity memory halt on single-bit errors and fail to detect double-bit errors entirely. Clearly, as memory is increased, better techniques are required.

To combat this problem, the IBM PC Servers employ schemes to detect and correct memory errors. These schemes are called Error Correcting Code (or sometimes Error Checking and Correcting but more commonly just ECC). ECC can detect and correct single-bit errors, detect double-bit errors, and detect some triple-bit errors.

ECC works like parity by generating extra check bits with the data as it is stored in memory. However, while parity uses only 1 check bit per byte of data, ECC uses 7 check bits for a 32-bit word and 8 bits for a 64-bit word. These extra check bits, along with a special hardware algorithm allow for single-bit errors to be detected and corrected in real time as the data is read from memory.

Figure 12 on page 41 shows how the ECC circuits operate. The data is scanned as it is written to memory. This scan generates a unique 7-bit pattern, which represents the data stored. This pattern is then stored in the 7-bit check space.

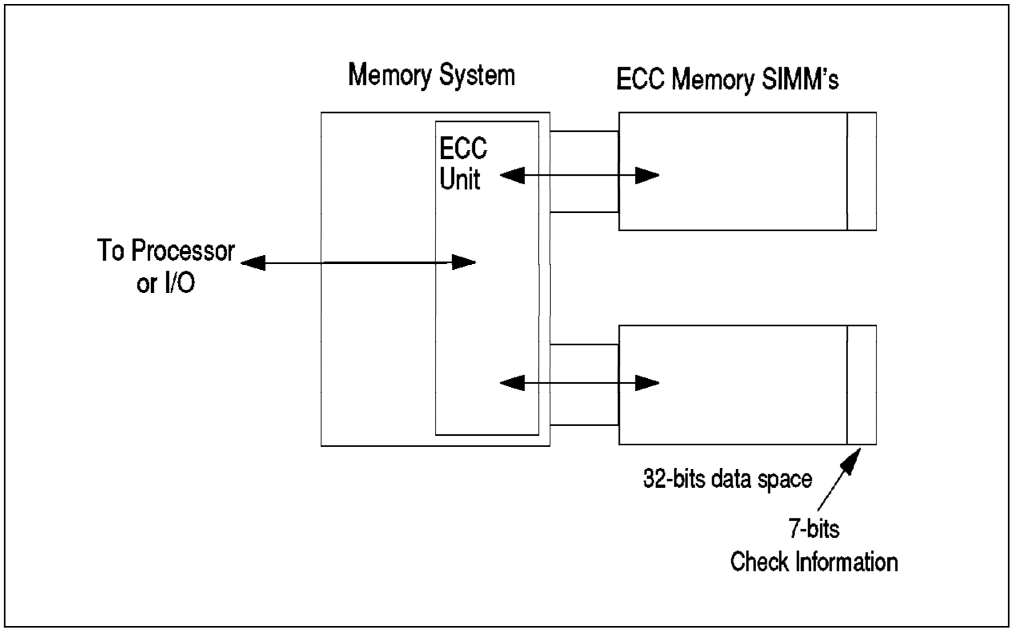


Figure 12. ECC Memory Operation

As the data is read from memory, the ECC circuit again performs a scan and compares the resulting pattern to the pattern which was stored in the check bits. If a single-bit error has occurred (the most common form of error), the scan will always detect it, automatically correct it and record its occurrence. In this case, system operation will not be affected.

The scan will also detect all double-bit errors, though they are much less common. With double-bit errors, the ECC unit will detect the error and record its occurrence in Non-Volatile RAM (NVRAM); it will then halt the system to avoid data corruption. The data in NVRAM can then be used to isolate the defective component.

In order to implement an ECC memory system, you need an ECC memory controller and ECC SIMMs. ECC SIMMs differ from standard memory SIMMs in that they have additional storage space to hold the check bits.

The IBM PC Server 720 has ECC circuitry and provides support for ECC memory SIMMs to give protection against memory errors. Tools such as PC SystemView (also known as NetFinity) can be used to signal the administrator of any single-bit or double-bit failures.

4.3.3 Error Correcting Code-Parity Memory (ECC-P)

ECC-P uses standard parity memory to implement ECC function. It takes advantage of the fact that a 64-bit word needs 8 bits of parity in order to detect single-bit errors (one bit/byte of data). Since it is also possible to use an ECC algorithm on 64 bits of data with 8 check bits, it is possible to design a memory controller which implements the ECC algorithm using the standard memory SIMMs, provided that the SIMMs are installed in pairs.

Figure 13 on page 42 shows the implementation of ECC-P. The ECC-P controller reads/writes two 32-bit words and eight bits of check information to standard parity memory. Since 8 check bits are available on a 64-bit word, the system is able to correct single-bit errors and detect double-bit errors just like ECC memory.

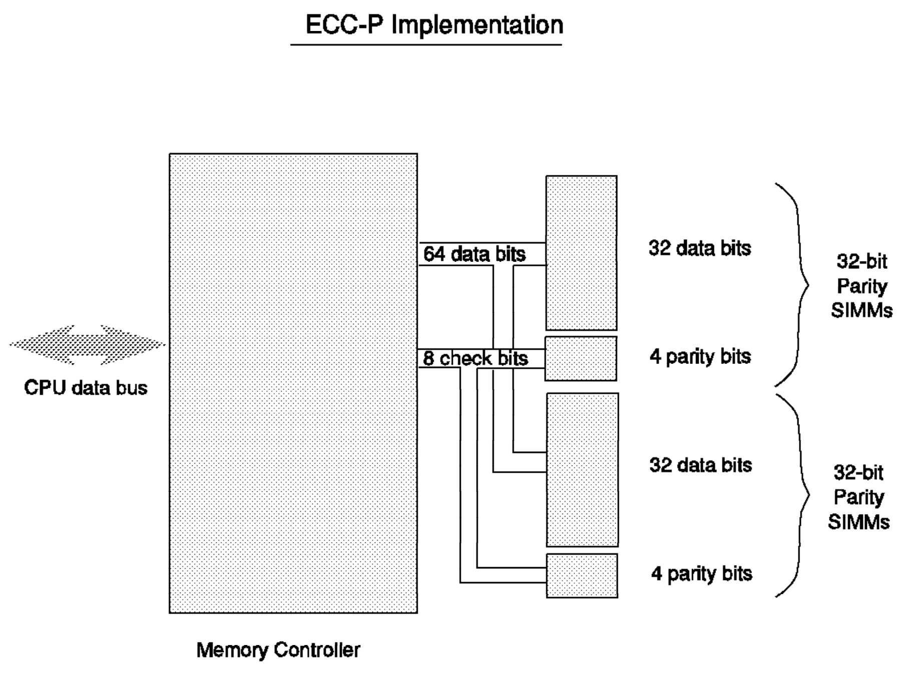


Figure 13. ECC-P Memory Implementation

While ECC-P uses standard non-expensive memory, it needs a specific memory controller that is able to read/write the two memory blocks and check and generate the check bits. Also, the additional logic necessary to implement the ECC circuitry make it slightly slower than true ECC memory.

The PC Server 704 uses ECC-P with 60 ns SIMMs while the PC Server 325 and 330 use 60 ns EDO DIMMs.

4.3.4 ECC-on-SIMM (EOS) Memory

A server that supports one hundred or more users can justify the additional cost necessary to implement ECC on the system. It is harder to justify this cost for smaller configurations. It would be desirable for a customer to be able to upgrade his system at a reasonable cost to take advantage of ECC memory as his business grows.

The problem is that the ECC and ECC-P techniques previously described use special memory controllers imbedded on the planar board which contain the ECC circuits. It is impossible to upgrade a system employing parity memory (with a parity memory controller) to ECC even if we upgrade the parity memory SIMMs to ECC memory SIMMs.

To answer this problem, IBM has introduced a new type of memory SIMM which has the ECC logic integrated on the SIMM. These are called ECC-on-SIMM (EOS) memory SIMMs. With these SIMMs, the memory error is detected and corrected directly on the SIMM before the data gets to the memory controller. This solution allows a standard memory controller to be used on the planar board and allows the customer to upgrade a server to support error checking memory.

ECC-on-SIMM is standard on the PC Server 320 and 520 and optional on the Server 310.

4.3.5 Memory Interleaving

Another technique used to reduce effective memory access time is interleaving. This technique greatly increases memory bandwidth when access to memory is sequential, such as in program instruction fetches.

In interleaved systems, memory is currently organized in either two or four banks, although this can be extended to eight, 16 and so on. Figure 14 shows a two-way interleaved memory implementation with a 32-bit data bus such as the one on i486DX processors.

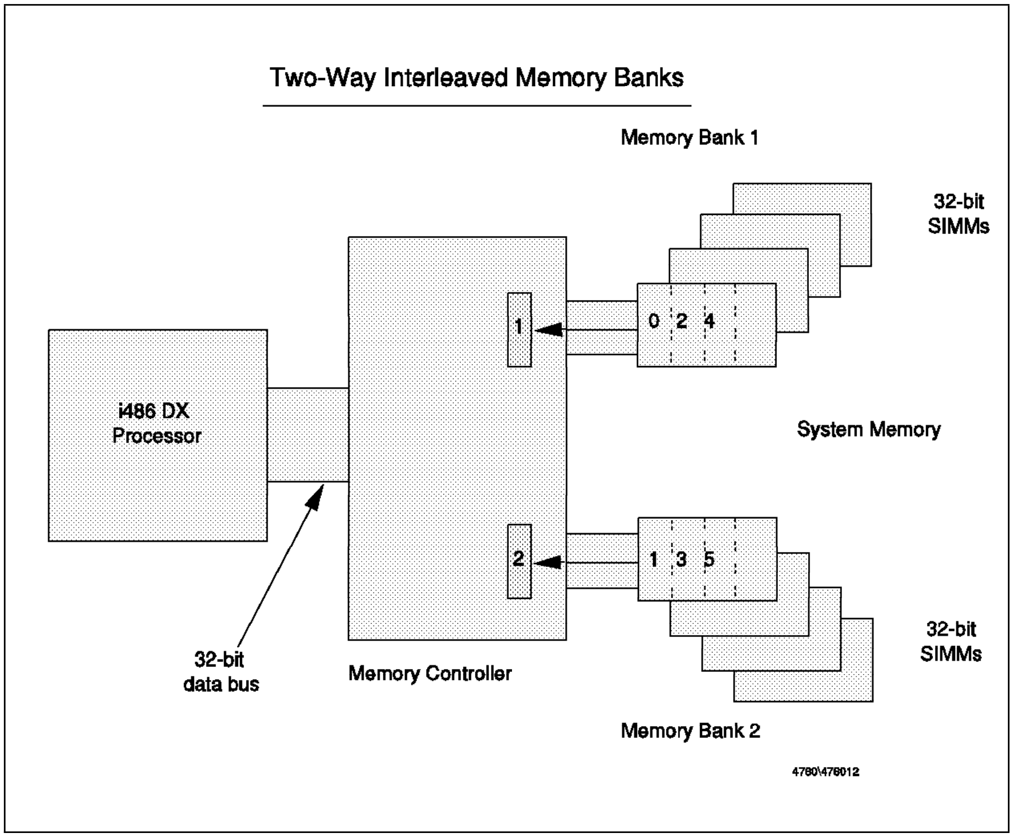


Figure 14. i486-based Two-Way Interleaved Memory System

Memory accesses are overlapped so that as the controller is reading/writing from bank 1, the address of the next word is presented to bank 2. This gives bank 2 a head start on the required access time. Similarly, when bank 2 is being read, bank 1 is fetching/storing the next word.

In systems implementing two-way interleaved memory, for example, memory must be installed such that there are two banks of appropriately wide memory available to the memory controller. For example, the i486DX has a 32-bit data bus. Therefore, for a i486DX-based system, a two-way interleaved system would require memory to be installed in pairs of 32-bit SIMMs.

The Pentium and Pentium Pro have a 64-bit data path. In a Pentium-based two-way system, (two banks, 64 bits wide) memory needs to be installed in fours.

The PC Server 704 is an example of a Pentium Pro-based two-way interleaved system which requires memory to be installed in sets of four SIMMs.

The PC Server 720 uses four-way interleaved memory so SIMMs have to be installed in sets of eight. In other words, four banks of 64-bit wide memory require eight 32-bit SIMMs. Figure 15 shows the implementation of four-way interleaving in the IBM PC Server 720.

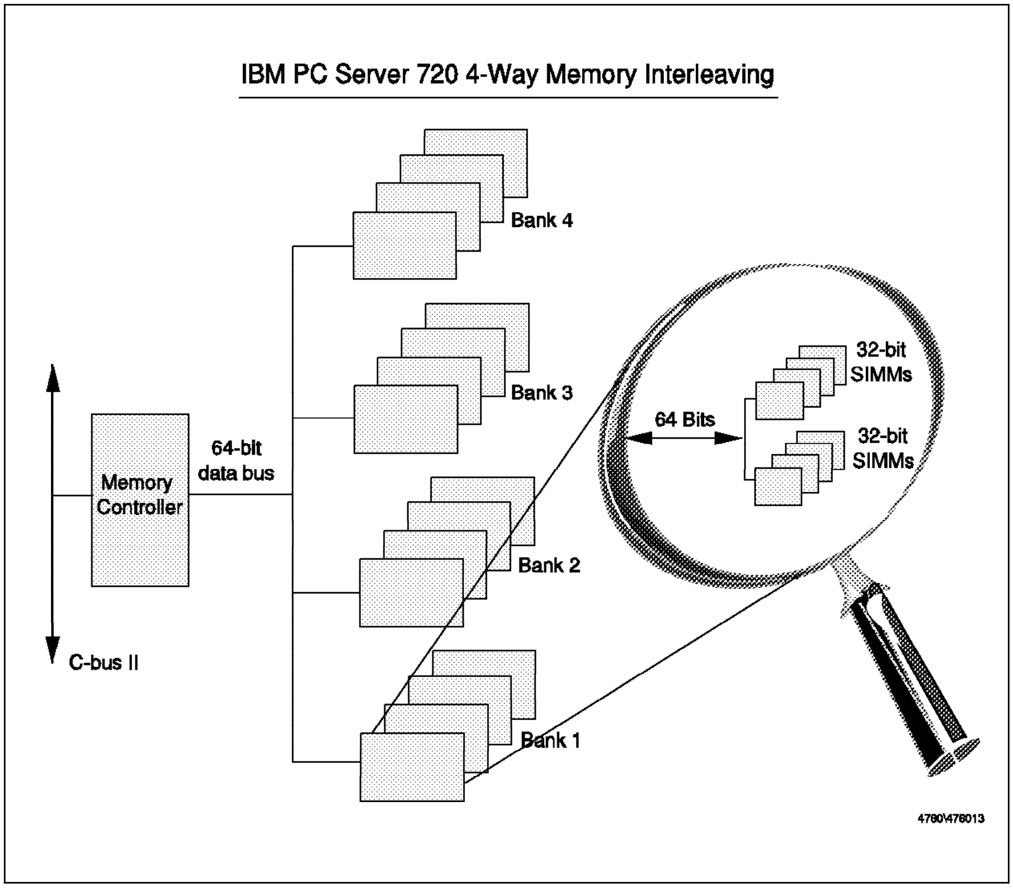


Figure 15. Four-way Interleaving in the IBM PC Server 720

In systems incorporating interleaved memory, each set of SIMMs installed has to be identical, that is the same speed, size and type. This is commonly referred to as installing memory in matched pairs or matched sets.

4.3.6 Dual Path Buses

A dual path bus allows both the processor and a bus master to access system memory simultaneously. Figure 16 on page 45 shows a dual path bus implementation.

Without a dual path bus, there is often contention for system resources such as main memory. When contention between the processor and a bus master occurs, one has to wait for the other to finish its memory cycle before it can proceed. Thus, fast devices like processors have to wait for much slower I/O devices, slowing down the performance of the entire system to the speed of the slowest device. This is very costly to the overall system performance.

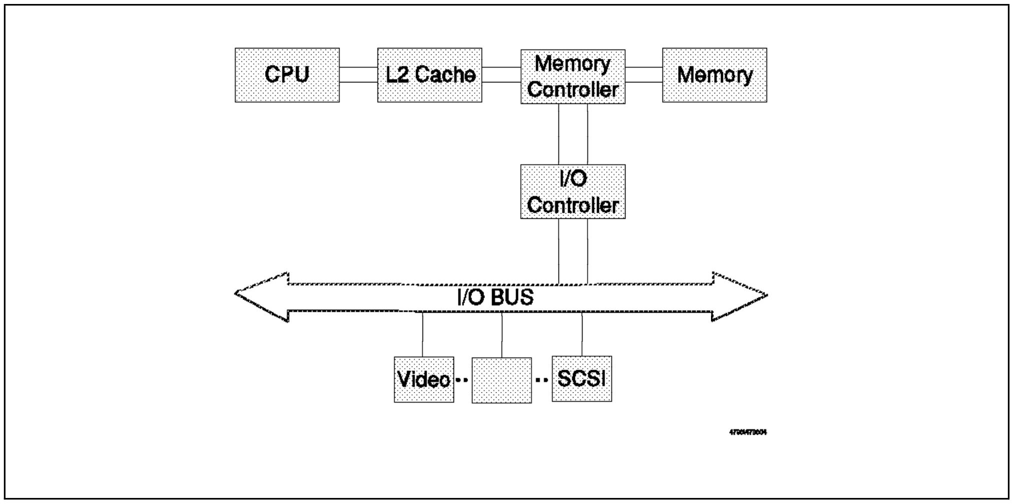


Figure 16. Dual Path Bus Implementation

4.3.7 SIMMs and DIMMs

Main memory is usually implemented as SIMMs or DIMMs. SIMMs (Single In-line Memory Modules) have tabs on both sides of the module that are shorted together in pairs so that the same signal can be read on both sides of the module. The original reason for this was that the connector was unreliable in transferring the signal from only one side of the SIMM. A 72-pin SIMM uses 72 signals but actually has 72 x 2 = 144 gold or tin tabs. A standard 72-pin SIMM has a 32-bit address and a 32-bit data bus.

Advancements in connector technology have made it now possible to read a separate signal from each side of the card. DIMMs (Dual In-line Memory Module) have tabs on both sides that carry different signals. A 168-pin DIMM has 84 tabs on each side for a total of 168 independent signals. The greater number of signals allows for a wider data bus (64-bit), which in turn allows for a higher performance design. The DIMM accomplishes this with a package which is only slightly longer than the standard SIMM. DIMMs also buffer critical signals on the module,which provides faster memory access and better signal quality.

4.3.8 General Tips on Memory Usage

IBM produces a large number of memory options for use with PC servers and desktop systems. It can sometimes be frustrating trying to determine the proper memory upgrade for your system. Further, your practical experience may tell you that there is a difference in what IBM will support and what actually works. The notes below are some general tips on the usage of different memory technology with the various IBM systems.

Notes:

1. Ensure you are meeting the requirements for interleaved systems and for systems with 64-bit data buses. For example, in two-way interleaved systems using a Pentium or Pentium Pro CPU, you would need to install memory in matched sets of 4 SIMMs. For more information, please see 4.3.5, “Memory Interleaving” on page 43.

2. An IBM memory option with the word kit in the description usually means there is more than one SIMM in the package, typically two SIMMs. For

example, part 11H0657 is a 32 MB memory upgrade kit which contains two 16 MB ECC-on-SIMM (EOS) SIMMs. This is an attempt by IBM to make the ordering process easier for systems which require matched pairs. The kit sizes may also change in the future to meet system requirements for matched fours and matched eights.

3. You can always use faster memory than the system requires but it will be more expensive. For example, you can use 60 ns memory in a system which requires 70 ns memory. You will not get any performance benefit by using memory faster than the controller.

4. Do not put a SIMM with lead/tin tabs into a system with gold-plated SIMM connectors. The system will function but the resulting corrosion will dramatically reduce the life of the SIMM and the connector.

5. You should be aware that some IBM systems implement a memory interface on the SIMM which is slightly different than the industry standard promulgated by JEDEC. (JEDEC stands for the Joint Electron Device Engineering Council and is the organization within the Electronics Standards Association (EIA) responsible for memory interfaces.) The description for memory SIMMs for such IBM systems usually specifies that it is IBM memory. The description for SIMMs which conform to the JEDEC standards usually contain industry standard in the text. The two implementations differ slightly in how the system detects the presence of additional memory. Take note whether your system requires IBM memory and do not use industry standard memory in a machine which requires IBM memory or vice versa.

6. Be aware of any memory configuration rules that apply. For example, the Server 720 has rules about which SIMM sizes can be installed on the same memory card.

Chapter 5. Storage Subsystem

The disk subsystem is a critical element of the server′s design. In this chapter, we examine the technologies used to implement IBM disk subsystems including device interfaces, hard drives, SCSI, SSA, RAID, and PFA. For product-specific implementation details, please see Chapter 17, “Storage Controllers” on page 163 and Chapter 20, “SCSI Disk Drives” on page 197.

5.1 Device Interfaces

The disk interface specifies the physical, electrical, and logical connections between the controller and the disk. There have been five main interfaces developed thus far. Each possesses different characteristics and performance levels. The interfaces are described in this section.

5.1.1 ST506

This interface was the original standard for microcomputers. It has a data transfer rate of 5 MBps between the controller and the disk and has a serial interface rather than a parallel interface. This interface is classified as a *device level interface* because the device itself has no logic to interpret commands. Its functions, such as formatting, head selection, and error detection, are directed by the controller, which typically is housed on an adapter card. A device level interface requires specific adapters and device drivers for each different type of device.

5.1.2 Enhanced Small Device Interface (ESDI)

This is an enhanced version of the ST506 interface. It provides a 10-MBps data transfer rate (15 MBps in some implementations). ESDI devices were the first to use a type of data encoding called *Run Length Limited* (RLL). RLL provides denser storage and faster data transfer than the older modified frequency modulation (MFM) technique. However, it is still a device level, serial interface.

5.1.3 Integrated Drive Electronics (IDE)

This is a *bus level interface* meaning that the device controller is built into the device itself. The IDE interface was designed for the low-cost PC market segment. The interface is flexible, although the original implementation allowed only two drives each of up to a maximum of 528 MB.

The IDE interface has been revised to *Enhanced IDE* (EIDE). It now includes caching at the adapter level, a CD-ROM interface, an increase from two to four drives and an increase in the maximum disk storage beyond the maximum of 528 MB. This revision is Revision 2e of the ANSI ATA-2 X3T9.2/948D specification.

EIDE provides a number of transfer modes that exceed the capabilities of the original IDE interface. These modes fall into two categories, PIO (programmable input/output) and DMA (direct memory access). These modes use different cycle times (in nanoseconds) to deliver the data from the platter faster. Table 5 on page 48 shows these faster modes. Most, although not all, EIDE drives will operate in one of these modes.

Mode	Cycle Time	Transfer Rate
PIO Mode 0	600 ns	3.3 MBps
PIO Mode 1	383 ns	5.2 MBps
PIO Mode 2	240 ns	8.3 MBps
PIO Mode 3	180 ns	11.1 MBps
PIO Mode 4	120 ns	16.6 MBps
DMA Single Word Mode 0	960 ns	2.1 MBps
DMA Single Word Mode 1	480 ns	4.2 MBps
DMA Single Word Mode 2	240 ns	8.3 MBps
DMA Double Word Mode 0	480 ns	4.2 MBps
DMA Double Word Mode 1	150 ns	13.3 MBps
DMA Double Word Mode 2	120 ns	16.6 MBps

Table 5. Enhanced IDE Transfer Modes

The theoretical maximum disk size for an EIDE drive is 8.4 GB and the theoretical maximum transfer rate is 20 MBps.

5.1.4 Small Computer System Interface (SCSI)

The SCSI interface is a high-speed parallel interface that transfers 8 or 16 bits at a time rather than one bit at a time for the ST506 and ESDI serial interfaces. Thus data transfer rates for SCSI are measured in mega*bytes* versus mega*bits* and are considerably faster than those of the older, serial interfaces.

SCSI is also a bus level interface which makes it very flexible. Since the commands are interpreted by the device and not the SCSI host bus adapter, new devices (with new commands) can be implemented and used with standard SCSI adapters. The device driver then interacts with the device via the new commands. An example of this would be a CD-ROM device sharing the same adapter as a hard disk drive.

The SCSI flexibility and high performance make it very suitable for the server environment. In fact, SCSI is the most widely used disk subsystem technology in advanced servers today. All the current IBM PC Servers use this technology.

5.2, “SCSI Technology” on page 49 looks at SCSI technology in greater detail.

5.1.5 Serial Storage Architecture (SSA)

Serial Storage Architecture (SSA) is a high-performance serial-interconnect technology used to connect I/O devices and host adapters. SSA is an open standard. SSA specifications have been approved by the SSA Industry Association and are in the process of being approved as an ANSI standard through the ANSI X3T10.1 subcommittee.

SSA subsystems are built up of loops of adapters and devices such as disk drives. This loop architecture allows full-duplex asynchronous data transfer between multiple devices simultaneously.

Theoretically, up to 127 devices are supported per loop. Devices in an SSA environment can be up to 25 meters apart with copper cabling and over 2 km with optical fiber connections.

The dual port technology with redundant paths guarantees high reliability and availability. Even if a device fails, the rest of the loop will continue working as two independent strings. See 5.3, “SSA Technology” on page 52 for more information about SSA technology.

5.2 SCSI Technology

As discussed in 5.1.4, “Small Computer System Interface (SCSI)” on page 48, SCSI is a bus level interface through which computers may communicate with a large number of devices of different types connected to the system unit via an SCSI controller and daisy-chained cable. The attached devices include such peripherals as fixed disks, CD-ROMs, printers, plotters, and scanners. The SCSI controller may be in the form of an adapter or may be integrated on the planar board.

There are several terms and concepts used in discussing SCSI technology that require definition.

5.2.1 SCSI-I, SCSI-II, and SCSI-III (UltraSCSI)

SCSI is a standard defined by the American National Standards Institute (ANSI). The original SCSI standard is defined in ANSI standard X3.131-1986. It defines an 8-bit interface with a burst-transfer rate of 5 MBps with a 5 MHz clock (1 byte transferred per clock cycle). It is sometimes referred to as SCSI-I to differentiate it from the generic term SCSI.

The SCSI-II specification gained final approval from ANSI in 1994 as standard X3T9.2/375R Revision 10K. It defines extensions to SCSI that allow for 16- and 32-bit devices, a 10 MBps synchronous transfer rate for 8-bit transfers and 20 MBps for 16-bit transfers, and other enhancements discussed in the next sections.

SCSI-III or UltraSCSI offers performance improvements over SCSI-II with two new synchronous transfer modes defined as:

- Fast-20, which supports 20 MBps on 8-bit data transfers and 40 MBps on 16-bit transfers.
- Fast-40, which supports 40 MBps on 8-bit data transfers and 80 MBps on 16-bit transfers.

SCSI-III also offers plug-and-play support through the use of a technique called SCSI configure auto-magically (SCAM).

Note: ANSI refers to the different SCSI specifications using the SCSI-I and SCSI-II type nomenclature. IBM uses SCSI-1 and SCSI-2 nomenclature in official product names.

5.2.2 Common Command Set

The SCSI standard defines a set of commands which must be interpreted by all devices that attach to an SCSI bus. This is called the common command set. Unique devices may implement their own commands, which can be sent by a device driver and interpreted by the device. The advantage of this architecture is that the SCSI adapter does not have to change when new devices with new capabilities are introduced.

5.2.3 Tagged Command Queuing (TCQ)

TCQ is an SCSI-II enhancement. It increases performance in disk-intensive server environments. With SCSI-I systems, only two commands could be sent to a fixed disk. The disk would store one command while operating on the other. With TCQ, it is possible to send multiple commands to the hard disk because the disk stores the commands and executes each command in the sequence that gives optimal performance.

Also with TCQ, the adapter has more control over the sequence of disk operations. For example, the adapter can tell the device to execute the next command immediately or it can instruct it to finish everything it already has been given before completing the new command.

5.2.4 Scatter/Gather

Scatter/gather allows devices to transfer data to and from non-contiguous or *scattered* areas of system memory and on-board cache independently of the CPU. The scatter/gather feature allows for high performance, even in systems that have fragmented memory buffers.

5.2.5 Fast/Wide Devices and Controllers

Fast refers to the doubling of the data transfer rate from the SCSI 5 MBps to 10 MBps by doubling the clock rate.

Wide is used in reference to the width of the SCSI parallel bus between the adapter and the device. Wide means wider than the original 8-bit path defined in SCSI-I, either 16 or 32-bit. Currently wide SCSI generally refers to 16-bit, as there are no 32-bit wide SCSI devices (due to the expense). With a 16-bit path, the data rate is double that of an 8-bit device.

Another advantage of wide SCSI is its ability to address 16 devices on the bus versus 8 on an 8-bit SCSI bus. This results from the additional control signals on the wide bus in addition to the wider data path.

Technically, what we know as wide SCSI is part of the SCSI III specification. SCSI II once had a wide specification, but it used two separate cables, each with its own clocking. It never became popular.

Fast/wide refers to adapters and devices that implement both the fast and wide interfaces previously defined. A fast/wide device has a maximum data transfer rate of 20 MBps.

Wide SCSI

Wide refers to the width of the bus between the SCSI adapter and the disk drive or other SCSI device. Do not confuse this with the width of the host bus interface (for example, a 32-bit MCA or PCI adapter).

5.2.6 Disconnect/Reconnect

Some commands take a relatively long time to complete (for example, a seek command takes roughly 11 ms). With this feature, the controller can disconnect from the bus while the device is positioning the heads (seeking). Then, when the seek is complete and data is ready to be transferred, the device can arbitrate for the bus and then reconnect with the controller to transfer the data.

If the device is really efficient, it will even begin reading the data and place it into a buffer before it reconnects. This allows it to burst the data across the SCSI bus, thereby minimizing the time it needs to use or *own* the bus.

These techniques result in a more efficient use of the available SCSI bus bandwidth. If the controller held onto the bus while waiting for the device to seek, then the other devices would be locked out.

Since, in effect, multiple operations can occur simultaneously, this is also sometimes referred to as overlapped operations or multi-threaded I/O on the SCSI bus. This feature is very important in multitasking environments.

If you only have one SCSI device, then disable this feature. You will gain a small amount of performance since there is a slight overhead associated with the disconnect/reconnect sequence.

5.2.7 Synchronous versus Asynchronous

An asynchronous device must acknowledge each byte as it comes from the controller. Synchronous devices may transfer data in bursts and the acknowledgments happen after the fact. Synchronous is much faster than asynchronous and most newer devices support this mode of operation. The adapters negotiate with devices on the SCSI bus to ensure that the mode and data transfer rates are acceptable to both the host adapter and the devices. This process prevents data from being lost and ensures that data transmission is error free.

5.2.8 Physical and Logical Unit Numbers

Each device on the bus has a unique, two-part address consisting of a physical and a logical address. The physical address is sometimes referred to as a PUN (physical unit number) while the logical address is referred to as a LUN (logical unit number). The addresses take the form of PUN.LUN.

Each physical device has its own physical address. A narrow (8-bit) SCSI bus can have up to eight addresses and a wide bus (16-bit) can have up to 16 addresses. Examples of physical devices include adapter cards, SCSI disk drives, CD-ROM drives, and tape drives. You can think of the physical unit as any piece of hardware that requires you to connect it via an SCSI cable.

A logical unit is the device inside the physical unit that holds the data. LUNs are needed because there can be multiple logical devices inside of one physical unit. For example, a CD-ROM juke box has one LUN for each platter inside the unit. This allows the system to address each platter uniquely.

Each physical unit can have up to eight logical devices which are addressed as LUN 0-7. The first LUN is always zero. Most devices have only one LUN. However, some tape devices and CD-ROM units have multiple LUNs.

Both physical and logical addresses begin at address zero. The higher addresses are higher priority. Adapter cards are generally address seven. On an 8-bit SCSI bus, address 7.7 would have the highest priority with address 7.6 having the next highest and so on until address 0.0 which would have the lowest priority.

Both the physical and logical addresses are used to determine which SCSI device will be used to boot the system. System BIOS has a pre-defined order in which it will search to find a boot record. Different vendors use different schemes. For example, the BIOS on IBM Micro Channel adapter cards require the boot drive to be at physical address 6. The BIOS on most other adapter cards require the boot drive to be address 0, or sometimes 0 or 1.

5.3 SSA Technology

As briefly discussed in 5.1.5, “Serial Storage Architecture (SSA)” on page 48, SSA is a high performance serial technology used to connect devices to host adapters. SSA subsystems are typically built up by loops of adapters and devices. A simple example is shown in Figure 17.

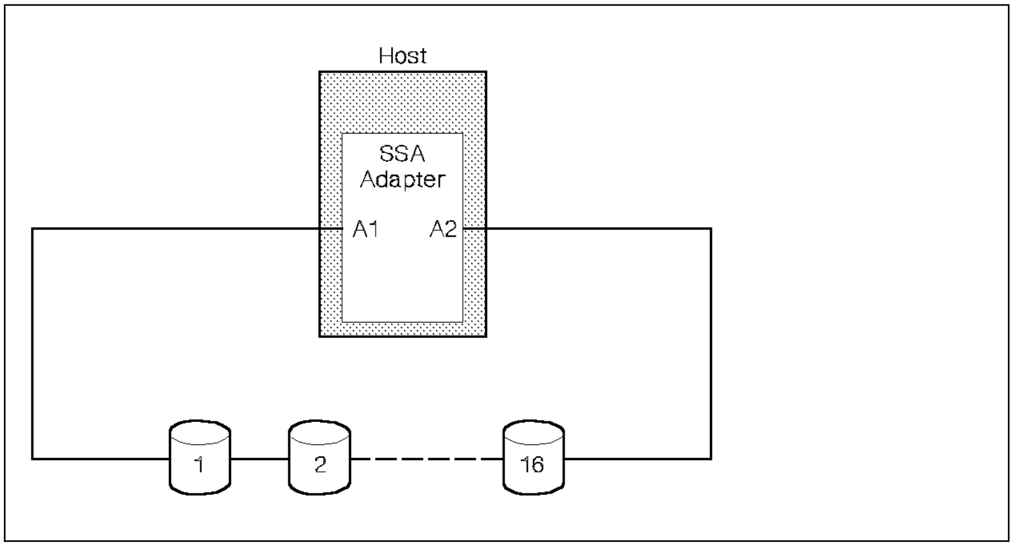


Figure 17. A Simple SSA Disk Subsystem

A single adapter controls an SSA loop of 16 disks. The host adapter is *dual-ported*, that is, it has two physical ports (in this example, A1 and A2).

Before we go any further, we provide you with some important definitions:

Port For all intents and purposes, the port is the physical connector on the device. Typical SSA devices are *dual-ported*, that is, they have two ports or connectors.

Node An SSA device. This can be a host adapter or disk or the like. There are three types of nodes: single-port, dual-port or switch-port. Switch-port nodes are covered later in this section.

Link The cable between ports. The connection typically contains six wires for internal cables and nine wires for external cables. Only four wires are used for transmitting information, two in and two out.

Loop Figure 17 is an example of a loop. All the nodes in a loop are dual-ported. The advantage of a loop over a bus is that if any of the links between nodes break, access to all the data is still possible via alternate paths (SSA does this re-routing automatically).

String A string is much like an SCSI chain or an SSA loop with one of the two links connecting the host adapter broken. The nodes on the ends of the string are single-port, while the interior nodes are dual-port. Loops are more common than Strings as they allow for a single link breakage.

Now consider Figure 18.

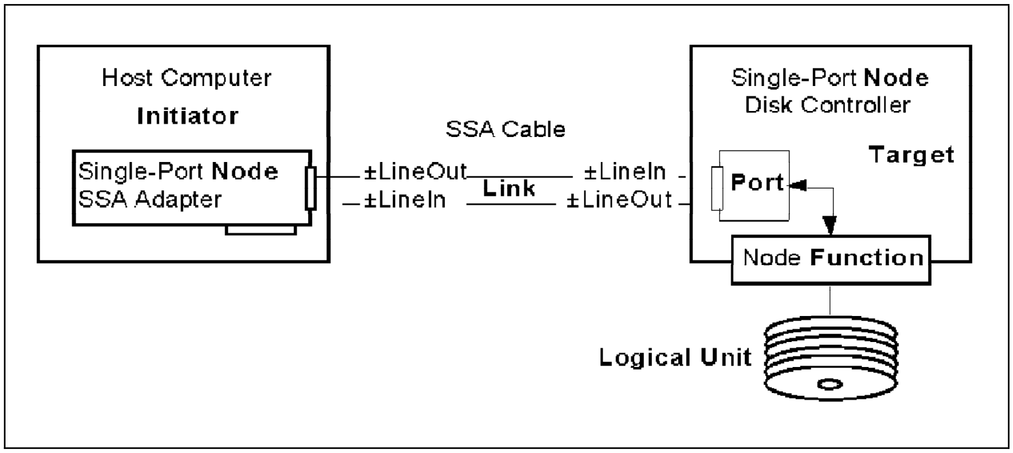


Figure 18. SSA Link - a Closer View

In this figure, the host adapter is a single-port node connected to a disk drive via an SSA cable. The SSA cable has four data wires represented as ±LineOut and ±LineIn. Both LineOut and LineIn are pairs of cables known as *differential pairs*. You should treat these pairs as one signal line, that is, the SSA cable is comprised of two signal lines (differential-pairs are discussed later in this section).

Initially, SSA allows data to be transmitted at 20 MBps *on each signal line*. Thus in Figure 18, the single-port SSA adapter can support up to 20 MBps both inbound and outbound *at the same time*. In other words, the single-port adapter supports up to 40 MBps total bandwidth in a full-duplex mode. As most SSA adapters are dual ported, potential bandwidth is 80 MBps per adapter.

The transmission rate of 20 MBps per signal line is only the initial limit. The SSA architecture is open-ended so higher speeds such as 40 MBps or 100 MBps per direction per port can be supported.

5.3.1 Serial versus Parallel Connections

The current mainstream architecture for transmitting data to high-bandwidth high-capacity devices is SCSI. SCSI allows up to 40 MBps through the use of a 68-wire *parallel* connection. It is recognized, however, that further increases in bandwidth are limited.

To increase the bandwidth of a parallel interface, there are two options:

1. Increase the speed or *clock rate* of the transmission (currently limited to 20 MHz maximum).

2. Increase the number of data signals (width of the data bus) which is currently 16 bits for an SCSI wide interface.

An alternative approach is to change from a parallel to a *serial* interface.

This would seem to be a backward step given the performance advantages of a wide data bus. However, a serial I/O connection can take advantage of the many advances in high-speed data communications that have been developed. These techniques allow the serial link to be clocked much faster than the parallel interface and actually achieve a higher bandwidth than the parallel interface even though it has fewer data signals.

Some of the techniques that make this possible are:

- Balanced interface

  The most effective transmission of high frequency signals using copper wire is achieved by using *differential pairs* where the same signal is transmitted with the opposite polarity on the other wire of the pair. The advantage of a differential pair is that any induced noise in the signal is canceled when one signal is subtracted from the other. This is referred to as a *balanced* interface.

  A similar scheme is used in differential SCSI implementations mainly to increase the distance between controller and device. However, the clock rate is unchanged from that in *single-ended* SCSI.

- Higher quality cables

  Because of the fewer number of signals required to implement a serial connection, the quality of the electrical conductors (cables) used to carry the signals can be of much greater quality without becoming cost prohibitive as they would with a parallel interface.

  A good quality serial cable consists of high-quality twisted pairs that are individually shielded for maximum noise immunity. This eliminates data corruption caused by *cross talk* where unwanted energy from another signal is induced on an adjacent conductor in the cable. Cross talk worsens as the clock rate is increased.

  Contrast a cable such as this with the standard ribbon cable used to implement an internal SCSI connection. The ribbon cable has no shielding and no twists.

In addition to the higher data rate, the serial interface provides other advantages:

- Other transmission mediums, such as optical, radio and telephone, can be used to continue the I/O connection; these are not practical in parallel transfers.
- The Application Specific Integrated Circuits (ASICs) that implement the interface require fewer Input/Output signals resulting in fewer pins on the chip. In most technologies, it is the number of pins rather than the number of logic cells that dictate the cost of the integrated circuit.
- Correspondingly, the boards that house these chips require less line drivers and receivers to implement the interface. These are expensive, difficult to place, and consume more power.

Although in theory a serial link could just require one wire, the nature of the transmission medium, the speed of the data and the requirements of the connection protocol make this unrealistic. For example, the RS232 serial link typically has nine connections. The reasons are as follows:

- A reliable transmission protocol usually demands two-way, or full-duplex communication. So the second communication path can be used for acknowledgment, bidirectional concurrent data transfer or both. When this is used in a loop, the second path may be used to double the bandwidth to the device and to provide an alternative route to the device should one path fail.
- As previously discussed, SSA uses a balanced interface that requires two conductors in the cable for every signal. This, in effect, doubles the number of conductors required in the cable.

SSA uses both of these features in its physical design.

5.3.2 SSA Characteristics

Components on SSA loops are called nodes: a node can be either an *initiator* or a *target*. An initiator issues commands and targets respond with data and status. The SSA nodes in the adapter are, therefore, initiators and the SSA nodes in the disk drives are targets. Each SSA node is given a unique address (UID) at the time of its manufacture which allows the initiators in the loop to determine what other SSA nodes are attached to the loop.

The SSA architecture allows more than one initiator to be present in a loop. In that case, commands and data from multiple initiators can be directed to the same or different targets and intermixed freely. Error recording in an SSA network is coordinated by one particular initiator, known as the *master initiator*. This is the initiator with the highest UID. If a new initiator is added to the network with a higher UID than those currently present, it takes over the master responsibilities for that network. Similarly, if a master initiator is removed from the SSA network, the initiator with the next highest UID takes over master responsibility. This master handover occurs automatically without any user intervention.

The master initiator has two primary functions:

1. The node where all *asynchronous alerts* are sent
2. Coordinator of error recovery procedures

Asynchronous alerts are used to notify the master initiator *asynchronously* of high-severity errors or conditions. Examples of such conditions are when there is an unrecoverable link error, or a new node is added to the network. Refer to 5.3.4, “Error Handling” on page 58 for more information.

The basic unit of data transferred between SSA nodes is a frame (Figure 19 on page 56) which contains 1 byte describing the frame type, 1 byte of CRC, up to 6 bytes of addressing and up to 128 bytes of data. The data block contains either an SSA message or the user data.

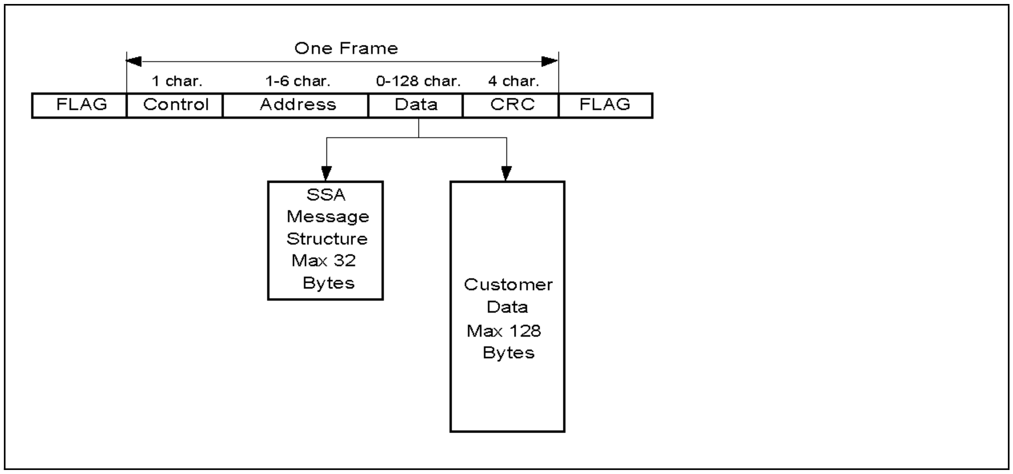


Figure 19. The SSA Frame

The SSA loop arrangement has the following characteristics:

- Full duplex support is provided on each link, so that traffic can be present in both directions on the loop simultaneously.
- 40 MBps total per port (20MBps inbound and 20MBps outbound) resulting in 80 MBps total per loop bandwidth.
- The loop supports spatial reuse; that is, different frames can be moving between different devices on the loop concurrently. For instance, a frame could be moving from disk 1 to disk 2 at the same time as the adapter is sending a frame to disk 1 (refer to 5.3.5.2, “Spatial Reuse” on page 60).
- If a break occurs in the loop (for example, if a cable is disconnected), each device on the loop adjusts its routing methods, under direction from the master initiator, so that frames are automatically rerouted to avoid this break. This allows devices to be removed from or added to the loop while the subsystem continues to operate without interruption.
- Hot pluggable cables and disks.
- Very high capacity per adapter of up to 128 devices per loop (although most adapter implementations limit this).
- Large distance between devices (up to 25 m with copper cables, more than one kilometer with optical links).
- As all SSA devices have unique addresses an SSA setup is auto-configuring.
- SSA is an open standard.

5.3.3 SSA Topology

Currently, SSA has three possible configurations:

- Loop
- String
- Switch

5.3.3.1 Loop Configuration

The loop configuration is the most common arrangement of SSA devices. In Figure 20, Host 1 and disks 1, 2 and 3 form a loop. The connecting line between each device is a *differential pair* (as defined in 5.3.1, "Serial versus Parallel Connections" on page 53). *All* nodes in the loop must be dual-ported (for example, in Figure 20, disk 1 is a dual port node, but disk 5 is not.)

If any connection between two devices fails, all devices can still be accessed by the host adapter. SSA provides for automatic reconfiguration if such an event occurs. Similarly, a node can be dynamically inserted into the loop or removed from the loop without preventing the other nodes from functioning.

Up to 128 nodes (that is, 256 ports) can be connected in a loop.

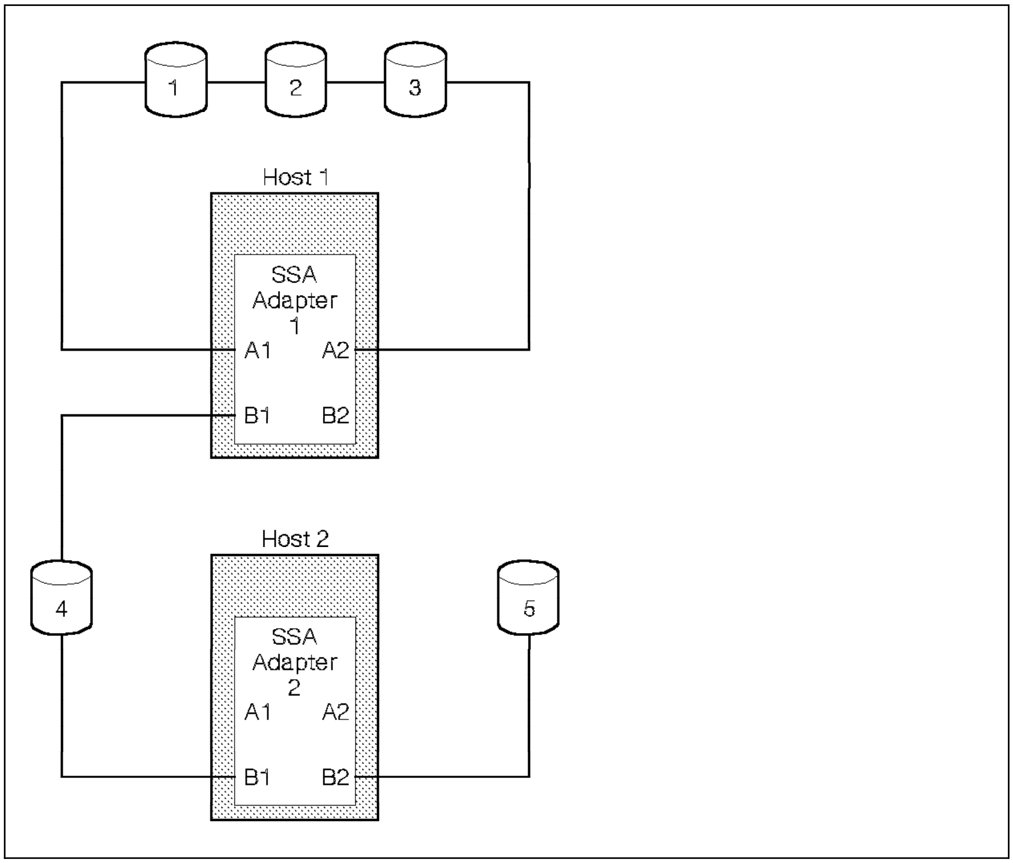


Figure 20. An Example of an SSA Loop and an SSA String. Disks 1, 2 and 3 form a loop with Host 1; Disk 5 forms a string with Host 2; Disk 4 forms a string with both Host 1 and 2 providing multi-host access to its data.

5.3.3.2 String Configuration

The string configuration is less commonly used than the loop as it does not provide redundant access to all disks. The host adapter and the end node are single-port whereas the other nodes are dual-port nodes. This configuration will fail if there is a break in any connection.

In Figure 20, disk 5 is connected to host 2 in a basic string configuration. Disk 4 also is connected to hosts 1 and 2 in a string formation. This configuration allows for the connectivity of multiple hosts (systems or adapters) providing host redundancy.

Up to 129 nodes (that is, 256 ports) can be connected in a string.

5.3.3.3 Switch Configuration

Switches allow a number of strings to be connected to achieve an almost unlimited number of nodes. Switches are typically four-port nodes. A switch allows large numbers of nodes to be connected together and also enables alternative paths to be established to provide fault tolerance. A switch network can also include other *cyclic paths*. Cyclic paths are not loops if they include a switch.

A switch allows connectivity to at most, 96 ports, although adding additional switches allows for a larger network.

5.3.4 Error Handling

As discussed briefly in 5.3.2, "SSA Characteristics" on page 55, the *master initiator* (or master node) is responsible for receiving notification of, and coordinating the recovery from, high-importance errors and conditions. However, not all errors are handled by the master node.

SSA handles errors in two ways, depending on the importance or severity of the error or condition:

1. Low severity errors are handled by the node itself.
2. High severity errors are handled by the master initiator.

5.3.4.1 Low Severity Errors

These errors typically affect only a single node or between two adjacent nodes. An example of such an error would be line noise or other transmission errors. The errors are local and do not affect other nearby nodes nor the whole network. A procedure known as Link Error Recovery Procedure, or *Link ERP* specifies what the node must do to recover from the condition. The Link ERP is usually implemented in the microcode of the node itself.

The SSA X3T10.1 standard defines what the Link ERP is and how every node should progress through that procedure. In summary, the Link ERP is:

1. Local port detects an error, enters a check state.
2. Local port transmits a *Link Reset* frame to the remote port and waits for an acknowledgement from the remote port.
3. If remote port receives the Link Reset frame with the correct CRC and successfully transmits the ACK, it enters a check state and transmits a Link Reset frame back to the local port.
4. The local port should then receive the Link Reset frame, verify its CRC, and then transmit an ACK back to the remote port.
5. Both ports then verify the Link Reset frames as being valid.
6. Both ports reset their registers back to initial states and re-transmit all lost frames.

As an aid to detecting errors, all SSA ports transmit a heartbeat when they are not sending or receiving data.

If any of the steps fail, then the master initiator is alerted by an asynchronous alert.

5.3.4.2 High Severity Errors

High severity errors are handled by the master initiator which is notified by an *asynchronous alert*. The master initiator coordinates the recovery for these errors for the network. Events that will trigger an asynchronous alert are:

- The escalation of a low severity error by the Link ERP
- A router in a node receives a frame addressed to one of its ports that is not operating
- A node receives an invalid message
- A new link is connected to the network via a previously non-operational port of an existing node
- The number of times the Link ERP has been invoked on a node has exceeded a predetermined threshold

The format of the message that is transmitted by the node to the master initiator is defined in the SSA standard. The procedure that the master initiator follows is, in summary:

1. Node (or nodes) generate asynchronous alerts which are routed to the master initiator.
2. The master initiator notifies the other primary initiators.
3. All affected commands are stopped.
4. Each initiator in the network performs a reconfiguration of the nodes under its control, as required
5. All affected nodes are returned to normal state.
6. All stopped commands are reissued, using alternative paths as required.

5.3.5 Performance Aids

Two components of the SSA architecture provide greater performance:

- Cut-through routing
- Spatial reuse

5.3.5.1 Cut-Through Routing

It would appear at a glance that with all the error checks and flow control between links that a frame that crossed several links would make a very sluggish journey toward its intended destination, held up at each node while its credentials were checked. However, the expected error rate across SSA links is so low that such delays can be avoided by a method known as cut-through routing.

Cut-through routing (or worm-hole routing) means that a node may forward a frame character-by-character as it is received; it does not have to wait for confirmation that the frame passed its CRC check. Using this method, the delay can be as little as 5-10 characters or 0.5 microseconds at 20 MB per second.

If an error is detected on an inbound frame after a router has already started to forward the frame, then it sends an ABORT character (followed by a FLAG) to the receiver. The receiver will send an ABORT if it, too, has already begun forwarding. An ABORT character tells a receiver to discard the frame in which it occurs; however, the receiver that detected the error must now attempt to recover the error.

5.3.5.2 Spatial Reuse

One of the characteristics that distinguishes a ring (where each link between nodes is a separate connection) and a bus (where each node connects to the same pieces of wire) is that a ring allows the possibility of Spatial Reuse. This is the technique whereby links that are not involved in a particular data transaction are available for use in another transaction. This is not possible on a bus because the entire bus carries the same information.

An example of spatial reuse is shown in Figure 21. The single host adapter can transmit data X to or from drives A, B and C at the same time as it is transmitting data Y to or from drives D, E and F.

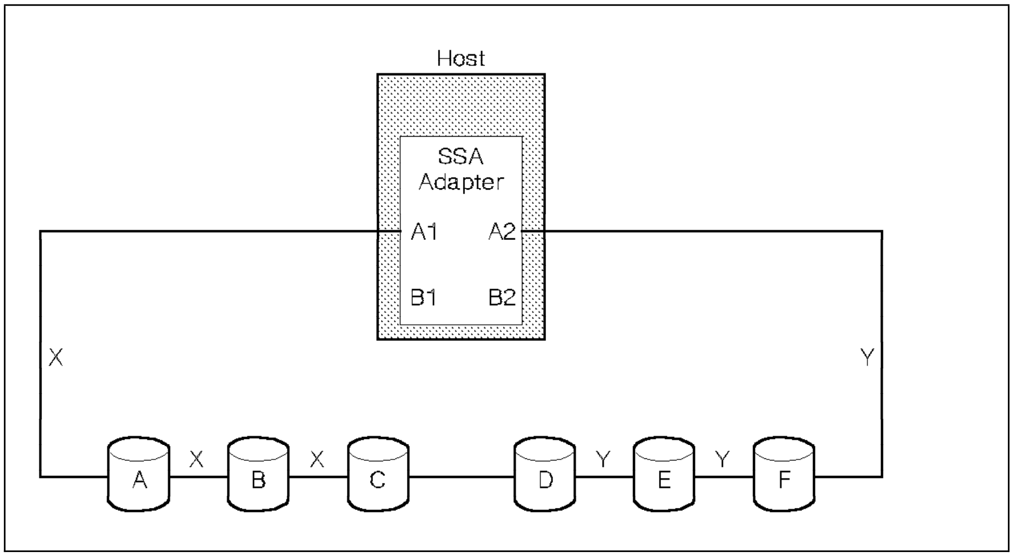


Figure 21. An Example of SSA Spatial Reuse. Disks 1, 2 and 3 form a loop with Host 1; Disk 5 forms a string with Host 2; Disk 4 forms a string with both Host 1 and 2 providing multi-host access to its data.

5.3.6 SSA Futures

IBM experts are also teaming with others at Seagate and Adaptec to prepare an architectural proposal for future serial storage. The proposal combines attributes of SSA and Fiber Channel-Arbitrated Loop (FC-AL) and would support operating speeds of 100 MBps or higher.

Design work on this proposed new technology has not yet begun, and products that would include it are not expected before 2000.

5.4 Hard Disk Drives

Ultimately, the hard disk drive is the component that has the most effect on subsystem performance. The following specifications should be considered when evaluating hard disks in order to optimize performance:

- Average access time
- Maximum transfer rate
- Disk Rotation Speed
- On-board cache size

Average Access Time: This is one of the standard indicators of hard drive performance. It is the amount of time required for the drive to deliver data after the computer sends a read request. It is composed of two factors, the *seek time* and the *latency*. The seek time is the time necessary to position the heads to the desired cylinder of the disk. The latency is the amount of time it takes for the disk to rotate to the proper sector on that cylinder.

It should be noted that two disks of the same physical size, for example, 3.5-inch disks, will differ in their access times with the larger capacity disk having a better access time. This is due to the fact that the distance between cylinders is shorter on the larger disk and, therefore, seek time is reduced. This is the primary reason that disk access times have been reduced as capacities have been increased.

Maximum Transfer Rate: Maximum or burst transfer rate is the rate at which the device can deliver data back to the SCSI adapter. It mainly depends on the processor/DMA controller integrated on the device but can be no more than the SCSI maximum data transfer rate, for example, 20 MBps for an SCSI-II Fast/Wide interface.

Rotation Speed: Rotation speed is literally the speed at which the disk platters spin. To some extent, the faster the disk spins, the shorter the access time and the faster the data is read from the disk. There are other factors that control this such as the number of sectors per track so you cannot say absolutely that a 7200 RPM drive will outperform a drive that spins at 5400 RPM. The faster drive will have a shorter latency since the *half-track average* (the time it takes the disk to spin one half the way around) will be shorter.

IBM currently makes drives that operate at both 5400 and 7200 RPM.

Caching: Cache is important for the same reason it is important on other subsystems. It speeds up the time it takes to perform routine operations. Like processor cache, disk drives usually offer two write policies, *write-through* and *write-back*.

With write-back cache, when a write operation is performed on the disk, the disk stores the modified data in its cache and signals the controller that the operation is complete before it actually is. This disk can then do the actual write operation when idle (or within a threshold time as set by the administrator). A gain in performance is achieved as the controller does not have to wait for the write operation to occur to completion. The disadvantage to this is that should the power fail before the modified data is written to disk, then data integrity problems will occur.

With the write-through policy, the data is physically written to disk before the controller is signaled that the operation is complete. This ensures the data is correctly written, but the trade-off is a reduction in performance.

Note: High-performance SCSI controllers often have a controller cache in addition to this disk cache (often as much as 4 MB). This controller cache can be battery-backed to improve data integrity.

5.5 RAID Technology

RAID is an architecture designed to improve data availability by using arrays of disks in conjunction with data striping methodologies. The idea of an array, a collection of disks which appear as a single device to the system, has been around for a long time. In fact, IBM was doing initial development of disk arrays as early as the 1970s. In 1977, IBM filed a patent for a disk array subsystem and was issued the patent in 1978. At that time, however, the cost of technology precluded the use of RAID in products.

In 1987, IBM co-sponsored a study by three researchers at the University of California at Berkeley on the potential use of arrays. This study by David A. Patterson, Garth Gibson, and Randy H. Katz resulted in a paper entitled "A Case for Redundant Arrays of Inexpensive Disks (RAID)."

The Berkeley study on arrays was a response to rapidly improving processor performance, which was increasing at 30 to 40 percent per year, while disk storage was improving at only 7 percent per year. The Berkeley team was looking for ways to unleash processor power by providing more data in the shortest period of time.

Patterson, Gibson, and Katz theorized that an array of small, inexpensive devices with relatively low capacity, reliability, and performance could provide improved performance by compounding the data rates of the individual devices in the array. The resulting performance would be better than that of larger, faster, more expensive, more reliable, offerings available at that time. The overall cost per unit of storage would drop because inexpensive disks were being used. Better yet, by designing logical data redundancy into the array, availability was expected to equal or exceed that of the then current high-end storage products.

The original Berkeley paper emphasized performance and cost. The authors were trying to fix a performance bottleneck while lowering costs at the same time. In their efforts to improve reliability, they designed the fault-tolerance and logical data redundancy, which was the origin of RAID. The paper defined five RAID architectures, RAID Levels 1 through 5. Each of these architectures has its own strengths and weaknesses, and the levels do not necessarily indicate a ranking of performance, cost, or availability.

While *inexpensive* disks were the focus of the initial RAID study, subsequent work showed that large aggregates of inexpensive generic devices are not well suited for real applications. The word inexpensive in the RAID acronym was later changed by the Berkeley team to *independent*, which is consistent with the fact that several independent disks in aggregate form an array.

Today, the interest in RAID is spurred by the prospects of high availability. However, there are some common *misconceptions* about RAID:

- *RAID makes backup of online data unnecessary*

  Data must still be backed up because when data is deleted by mistake, it cannot be recovered even with RAID.

- *RAID means the same as a disk array*

  RAID is an architecture that improves data availability in a disk array.

- *RAID products never lose data*

Most RAID architectures can tolerate one disk failure, but two disk failures could result in data loss.

- *RAID products have lower performance*

  Performance claims are hard to make because performance depends on the specific RAID architecture and the type of application. Some RAID Levels can improve performance for applications accessing data in a sequential fashion, but may degrade performance when used for transaction-oriented applications.

In order to know RAID strengths and limitations, a clear understanding of the different RAID levels or classifications is required.

5.5.1 RAID Classifications

The original RAID classification described five levels of RAID (RAID-1 to RAID-5). RAID-0 (data-striping), RAID-1 Enhanced (data stripe mirroring) and RAID-6 (data striping with double skewed parity) have been added since the original levels were defined. RAID-0 is not a RAID level as such, since it does not provide any data redundancy.

Different designs of arrays perform optimally in different environments. The two main environments are those where high transfer rates are very important, and those where a high I/O rate is needed, that is, applications requesting short length random records.

To maximize storage capacity using any of these RAID levels, it is wise to use drives of the same physical capacity, for example, all 2.25 GB drives. If a mixture of drives is used in an array, then only the portion of the drives equal to the smallest drive in the array is used and the rest is unused. For example, if an array is configured using three 2.25 GB drives and one 1.12 GB drive, then the array will only use 1.12 GB on each drive.

Table 6 shows the RAID array classifications and is followed by brief descriptions of their designs and capabilities. Table 7 on page 71 is a summary of the RAID classifications.

Note: The RAID levels (0 to 6) are not meant to indicate that one RAID level is newer or better than the others. They merely provide an easy naming convention.

Table 6. RAID Classifications

RAID Level	Description
RAID-0	Data striping
RAID-1	Disk Mirroring/dual copy
RAID-1 (Enhanced)	Data stripe mirroring
RAID-2	Bit interleave data striping with hamming code
RAID-3	Bit interleave data striping with parity disk
RAID-4	Block interleave data striping with one parity disk
RAID-5	Block interleave data striping with skewed parity
RAID-6	Block interleave data striping with double skewed parity

5.5.1.1 RAID-0 - Block Interleave Data Striping without Parity

Striping of data across multiple disk drives without parity protection is a disk data organization technique sometimes employed to maximize performance of the disk subsystem (for example, Novell NetWare's *data scatter* option).

An additional benefit of this data organization is *drive spanning*. With data striped across multiple drives in an array, the logical drive size is the sum of the individual drive capacities. The maximum file size may be limited by the operating system.

An example of RAID-0 is shown in Figure 22.

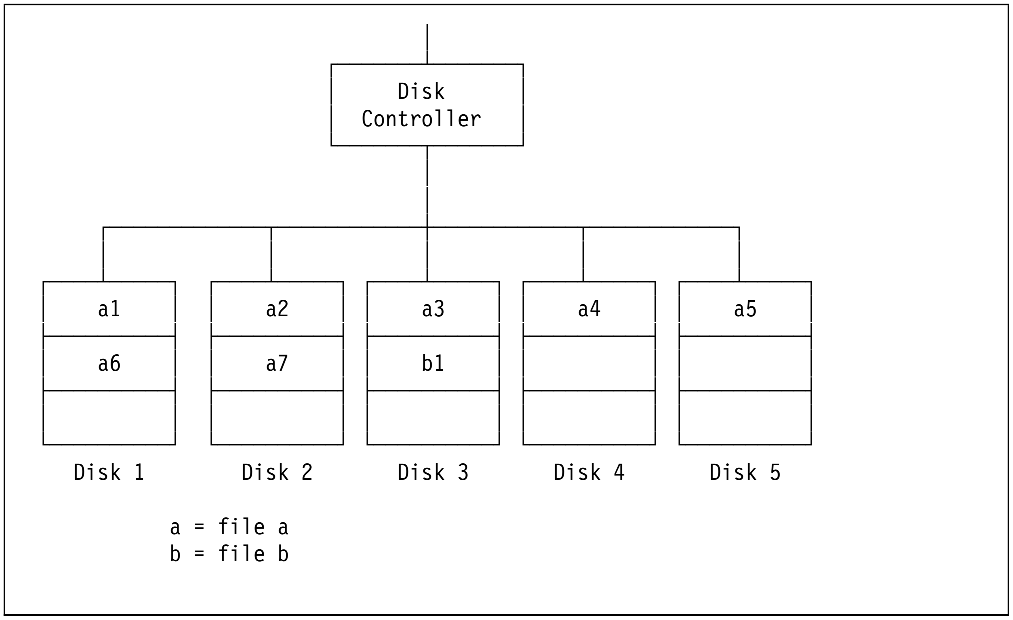


Figure 22. RAID-0 (Block Interleave Data Striping without Parity)

Data striping improves the performance with large files since reads/writes are overlapped across all disks. However, reliability is decreased as the failure of one disk will result in a complete failure of the disk subsystem according to the formula:

$$\frac{1}{\text{RAID-0 MTBF}} = \frac{1}{\text{Disk 1 MTBF}} + \frac{1}{\text{Disk 2 MTBF}} + \ldots$$

5.5.1.2 RAID-1 - Disk Mirroring

This approach keeps two complete copies of all data. Whenever the system makes an update to a disk, it duplicates that update to a second disk, thus mirroring the original. Either disk can fail, and the data is still accessible. Additionally, because there are two disks, a read request can be satisfied from either device, thus leading to improved performance and throughput. Some implementations optimize this by keeping the two disks 180 degrees out of phase with each other, thus minimizing latency.

However, mirroring is an expensive way of providing protection against data loss, because it doubles the amount of disk storage needed (as only 50 percent of the installed disk capacity is available for data storage) Figure 23 on page 65 shows an example of RAID-1 mirroring.

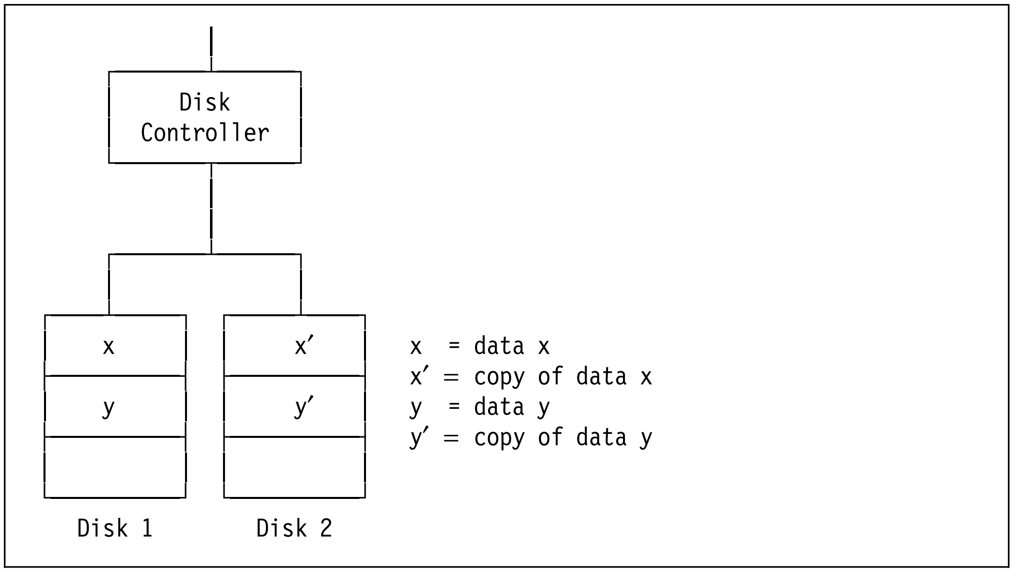


Figure 23. RAID-1 (Disk Mirroring)

5.5.1.3 RAID-1 Enhanced - Data Stripe Mirroring

RAID level 1 supported by the IBM PC Server array models provides an enhanced feature for disk mirroring that stripes data and copies of the data across all the drives of the array. IBM calls this RAID-1 *enhanced* and an example is shown in Figure 24. The first stripe is the data stripe; the second stripe is the mirror (copy) of the first data stripe, but it is shifted over one drive. Because the data is mirrored, the capacity of the logical drive, when assigned to RAID 1 Enhanced, is 50 percent of the physical capacity of the hard disk drives in the array.

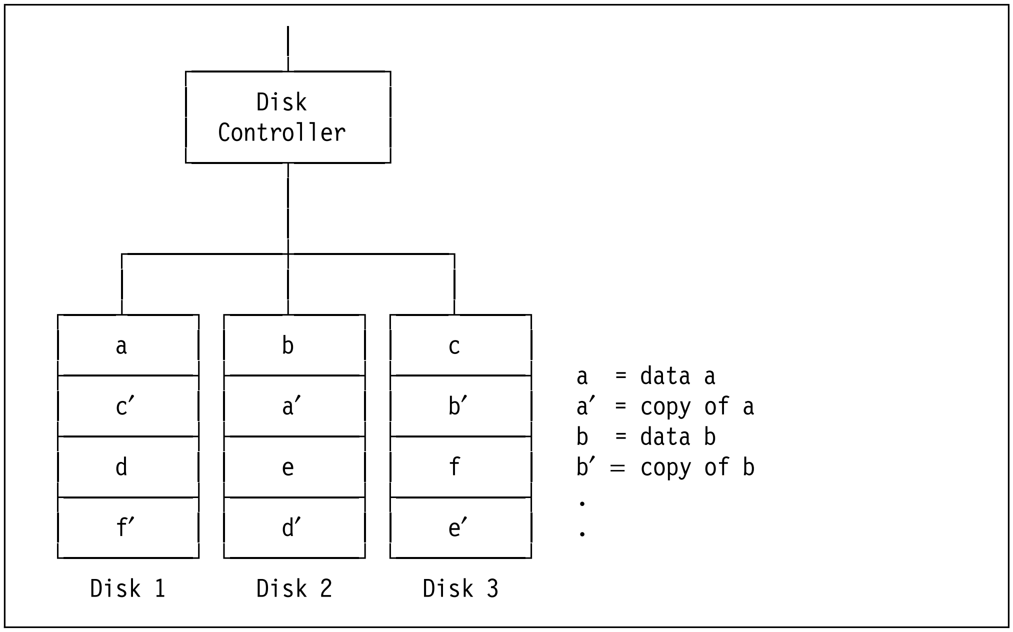


Figure 24. RAID-1 Enhanced, Stripe Mirroring

Some vendors have also implemented another slight variation of RAID-1 and refer to it as RAID-10 since it combines features of RAID-1 and RAID-0.

As shown in Figure 25 on page 66, this solution consists of mirroring a striped (RAID-0) configuration. In this example, a RAID-0 configuration consisting of 2 drives (drive 1 and 2) is mirrored to drives 3 and 4.

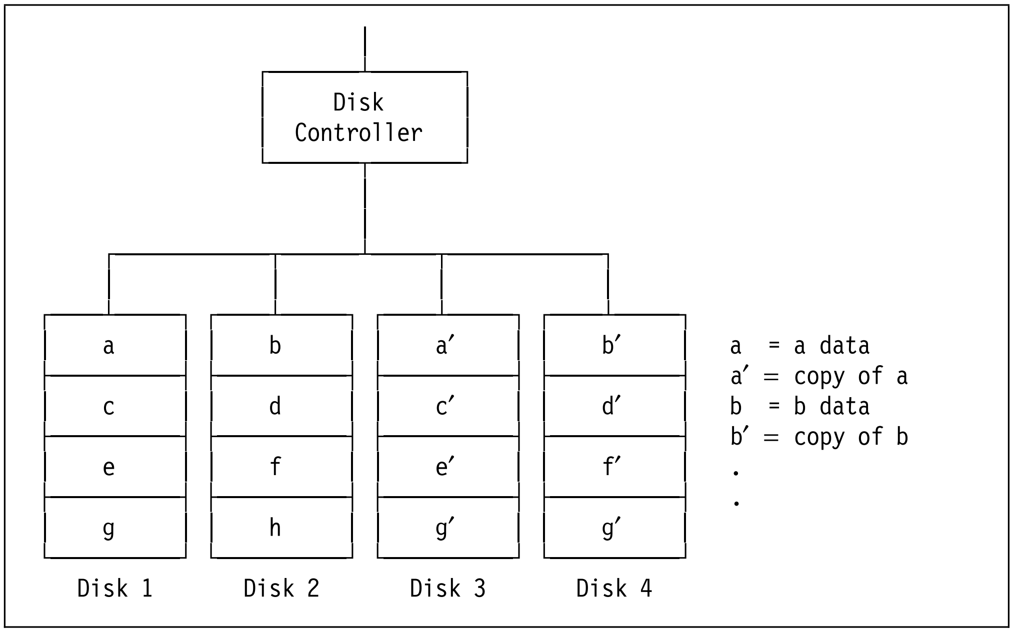


Figure 25. RAID-10 - Mirroring of RAID-0 Drives

Performance and capacity of RAID-10 are similar to RAID-1 with 50 percen of the total disk capacity usable. However, this solution always uses an *even* number of disks. RAID-1 enhanced can use both odd and even number of disks.

5.5.1.4 RAID-2 - Bit Interleave Striping with Hamming Code

This type of array design is another form of data striping: it spreads the data across the disks one bit or one byte at a time in parallel. This is called bit (or byte) interleaving.

Thus, if there were five disks in the array, a sector on the first drive will contain bits 0 and 5, and so on of the data block; the same sector of the second drive will contain bits 1 and 6, and so on as shown in Figure 26 on page 67.

RAID-2 improves on the 50 percent disk overhead in RAID-1 but still provides redundancy by using the Hamming code. This is the same algorithm used in ECC memory. The check bits can be generated on a nibble (4 bits), a byte (8 bits), a halfword (16 bits) or a word (32 bits) basis, but the technique works most efficiently with 32-bit words. Just like with ECC, it takes 7 check bits to implement the Hamming code on 32 bits of data.

Generating check bits by byte is probably the process used most frequently. For example, if data were grouped into bytes, 11 drives in total would be required, 8 for data and 3 for the check bits. An 8-3 configuration reduces the overhead to 27 percent.

Note: For clarity, the Hamming code drives are not shown in Figure 26 on page 67.

An array of this design will perform optimally when large data transfers are being performed. The host will see the array as one logical drive. The data transfer rate, however, will be the product of the number of drives in the array and the transfer rate of the individual drives.

This design is unable to handle multiple, simultaneous small requests for data, unlike the previous design. So, it is unlikely that it will satisfy the requirements for a transaction processing system that needs a high transaction rate.

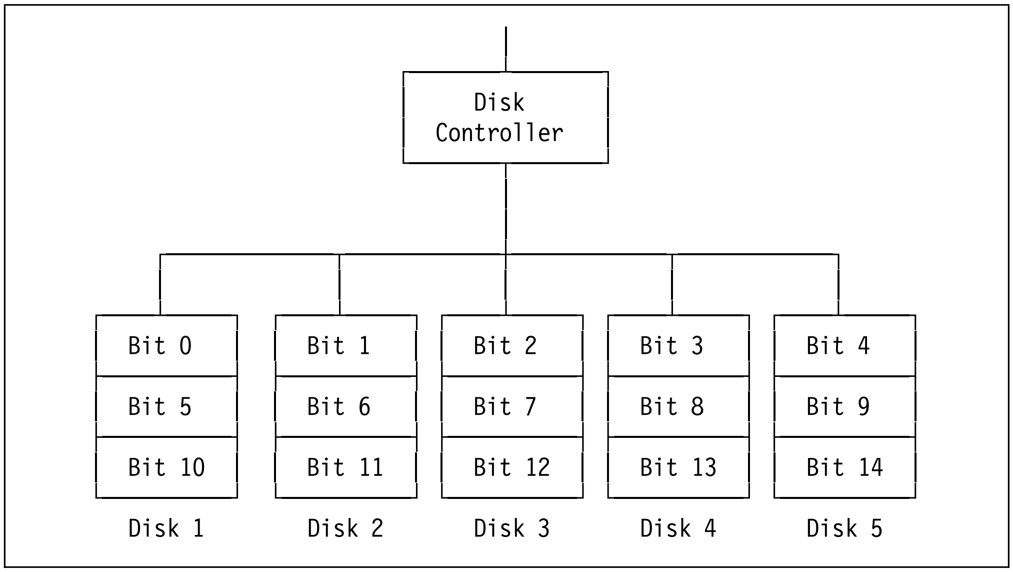


Figure 26. RAID-2 (Bit Interleave Striping with Hamming Code)

5.5.1.5 RAID-3 - Bit Interleave Data Striping with Parity Disk

RAID-3 stripes or distributes data sequentially across several disks. The data is written or retrieved in one parallel movement of all of the access arms. RAID-3 uses a single dedicated disk to store parity information.

In the example in Figure 5, there is an array of four disks. Three of the disks are used to store data, and the fourth is used to store parity for the three data disks. If one of the data disks should fail, the parity disk can be used, along with the remaining data disks, to regenerate the data. If the parity disk fails, access to data is not affected.

Because of the single parallel movement of all access arms, only one I/O can be active in the array subsystem at any one time. Because data is striped sequentially across the disks, the parallel arm movement yields excellent transfer rates for large blocks of sequential data, but renders RAID-3 impractical for transaction processing or other high throughput applications needing random access to data. When random processing does take place, the parity disk becomes a bottleneck for write operations.

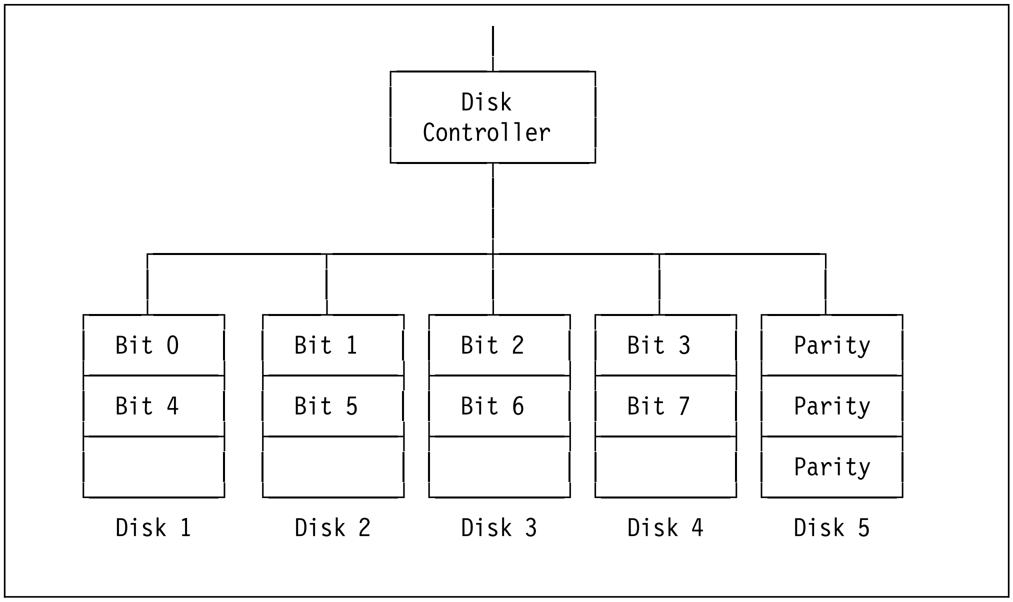


Figure 27. RAID-3 (Bit Interleave Data Striping with Parity Disk)

5.5.1.6 RAID-4 - Block Interleave Data Striping with One Parity Disk

The performance of bit-interleaved arrays in a transaction processing environment, where small records are being simultaneously read and written, is very poor. This can be compensated for by altering the striping technique, such that files are striped in block sizes that correspond to the record size being read. This will vary in different environments. Block sizes can be of the order of 4 KB to 64 KB.

RAID-4 provides n−1 drives of actual data storage (where n is the number of drives in the array).

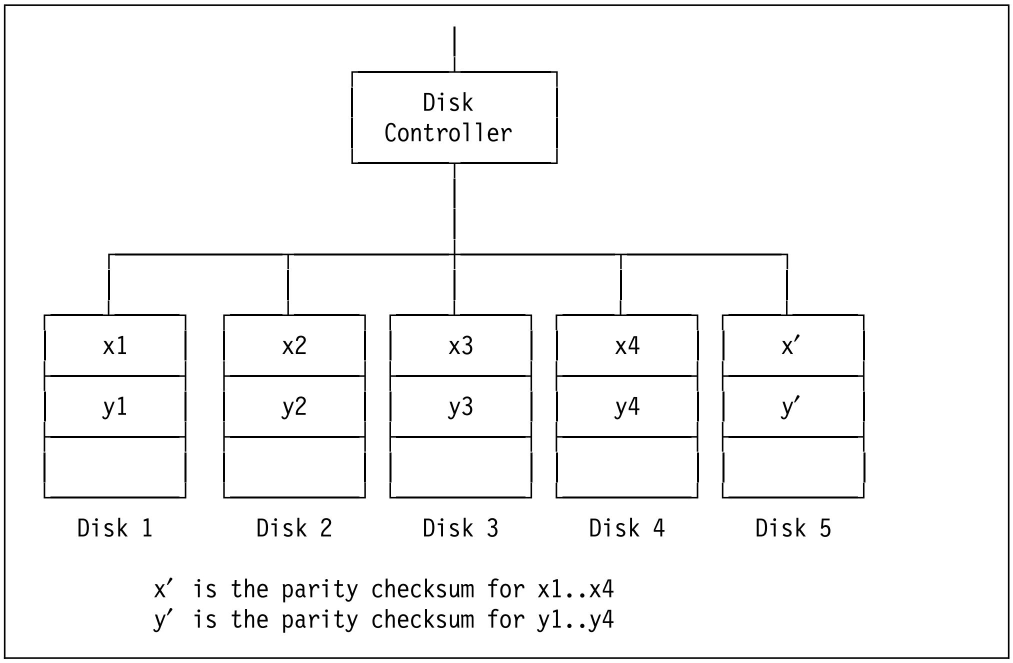


Figure 28. RAID-4 (Block Interleave Data Striping with One Parity Disk)

5.5.1.7 RAID-5 - Block Interleave Data Striping with Skewed Parity

RAID-5 does not have a dedicated parity disk, but instead interleaves both data and parity on all disks.

In RAID-5 the access arms can move independently of one another. This enables multiple concurrent accesses to the array devices, thereby satisfying multiple concurrent I/O requests and providing higher transaction throughput. RAID-5 is best suited for small blocksize, random access data.

One important difference between RAID-3 and RAID-5 is that in RAID-3 every transfer involves all of the disks. In RAID-5, most transfers involve only one data disk, thus allowing operations in parallel and giving higher throughput for transaction processing.

There is a write penalty associated with RAID-5. Every write I/O will result in four actual I/O operations: two to read the old data and parity and two to write the new data and parity.

Unlike RAID-4, the parity is skewed across all drives to remove the bottleneck that is created by storing all the parity information on a single drive.

Like RAID-4, RAID-5 provides n-1 drives of actual data storage (where n is the number of drives in the array). A minimum of three physical drives are required to form a RAID-5 array.

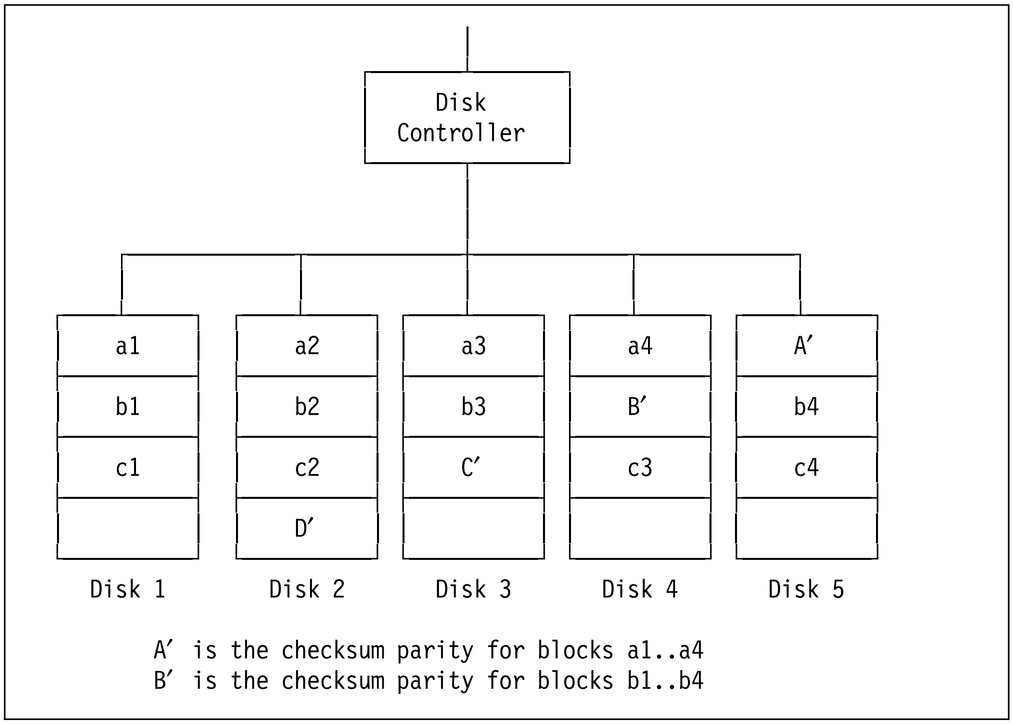


Figure 29. RAID-5 (Block Interleave Data Striping with Skewed Parity)

5.5.1.8 RAID-6 - Block Interleave Striping with Double Skewed Parity

RAID-6, which was not among the original Berkeley RAID levels, adds a second, independent parity block to RAID-5.

Figure 30 is an example of RAID-6 with an array of five disks. Three disks are used for data and two disks are used for parity, with data and parity rotating and interleaving within the array.

With two independent parity schemes, each using a different algorithm, data availability is extremely good and is uninterrupted even when two disk failures occur at the same time. However, more disk space is required for parity, and there is an even greater write penalty than with RAID-5. For this reason, the write performance of RAID-6 is extremely low. The lower performance and the complexity of implementation have made RAID-6 impractical for most applications.

RAID-6 provides n−2 drives of actual data storage (where n is the number of drives in the array). A minimum of four physical drives are required to form a RAID-6 array.

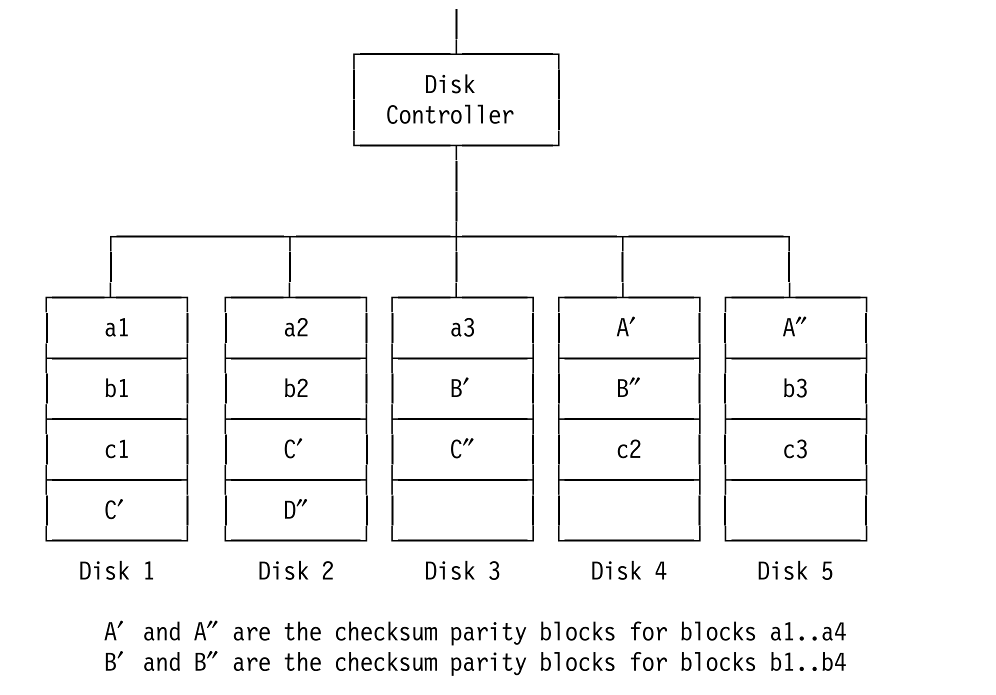


Figure 30. RAID-6 (Block Interleave Striping with Double Skewed Parity)

5.5.2 Summary of RAID Performance Characteristics

RAID-0: Block Interleave Data Striping without parity

- Fastest data-rate performance
- Allows seek and drive latency to be performed in parallel
- Significantly outperforms single large disk

RAID-1: Disk Mirroring/Disk Duplexing and Data Stripe mirroring (RAID-1, Enhanced)

- Fast and reliable, but requires 100 percent disk space overhead
- Data copied to each set of drives
- No performance degradation with a single disk failure
- RAID-1 enhanced provides mirroring with an odd number of drives

RAID-2: Bit Interleave Data Striping with Hamming Code

- Very fast for sequential applications, such as graphics modelling
- Almost never used with PC-based systems

RAID-3: Bit Interleave Data Striping with Parity

- Access to all drives to retrieve one record
- Best for large sequential reads
- Very poor for random transactions
- Poor for any write operations
- Faster than a single drive, but much slower than RAID-0 or RAID-1 in random environments

RAID-4: Block Interleave Data Striping with one Parity Disk

- Best for large sequential I/O
- Very poor write performance
- Faster than a single drive, but usually much slower than RAID-0 or RAID-1

RAID-5: Block Interleave Data Striping with Skewed Parity

- Best for random transactions
- Poor for large sequential reads if request is larger than block size
- Better write performance than RAID-3 and RAID-4
- Block size is key to performance, must be larger than typical request size
- Performance degrades in recovery mode (when a single drive has failed)

RAID-6: Block Interleave Data Striping with Double Skewed Parity

- Best for data available as it can suffer a two disk failure and still operate
- Best for random transactions
- Poor for large sequential reads if request is larger than block size
- Better write performance than RAID-3 and RAID-4 but worse than RAID-5
- Block size is key to performance, must be larger than typical request size
- Performance degrades in recovery mode (when a single drive has failed)

Table 7 (Page 1 of 2). Summary of RAID Performance Characteristics

RAID Level	Data Capacity	Large Transfers	High I/O Rate	Data Availability
Single Disk	n	Good	Good	varies
RAID-0	n	Very Good	Very Good	Poor
RAID-1	n/2	Good	Good	Good
RAID-2	n-1	Good	Poor	Good
RAID-3	n-1	Very Good	Poor	Good
RAID-4	n-1	Very Good	Poor	Good
RAID-5	n-1	Very Good	Good	Good

Table 7 (Page 2 of 2). Summary of RAID Performance Characteristics				
RAID Level	**Data Capacity**	**Large Transfers**	**High I/O Rate**	**Data Availability**
RAID-6	n-2	Very Good	Good	Better
Note: In data capacity, *n* refers to the number of equally-sized disks in the array.				

5.6 Predictive Failure Analysis

Predictive Failure Analysis (PFA) is a feature of IBM hard drives that monitors key device performance indicators that change over time or exceed specified limits. The disk notifies the system when an indicator surpasses a pre-determined threshold. This allows the administrator to replace a drive *before* it fails, thereby reducing unscheduled down-time.

Pending drive failures discovered by PFA can be passed to TME 10 NetFinity management software to alert the LAN administrator. Any drive notified by NetFinity in this way will be replaced by IBM under standard warranty conditions before the drive actually fails.

As with any electrical/mechanical device, there are two basic failure types:

1. On/off failures

   A cable breaking, a component burning out and a solder connection failing are all examples of unpredictable catastrophic failures. As assembly and component processes have improved, these types of defects have been reduced but not eliminated. PFA cannot provide warning for on/off unpredictable failures.

2. Degradation

   The second type of failure is the gradual performance degradation of components. Predictive Failure Analysis has been developed to monitor performance of the disk drive, analyze data from periodic internal measurements, and recommend replacement when specific thresholds are exceeded. The thresholds have been determined by examining the history logs of disk drives that have failed in actual customer operation.

Predictive Failure Analysis monitors performance two ways (see Figure 31 on page 73):

1. Measurement driven method

   The measurement driven process is based on IBM′s exclusive Generalized Error Measurement (GEM) feature. At periodic intervals, GEM automatically performs a suite of self-diagnostic tests which measure changes in the disk drive′s component characteristics.

   IBM leads the disk drive industry with this two-step condition monitoring approach. To accomplish this task, GEM directly measures various magnetic parameters of the head and disk, as well as figures of merit for the channel electronics. The GEM circuit monitors head fly height on all data surfaces (see 5.6.1.1, “Flying Height Monitoring” on page 73), channel noise, signal coherence, signal amplitude and writing parameters.

   Then, a digital signal processor (DSP) that is housed on the drive assembly runs an algorithm called *partial response maximum likelihood* or PRML. The

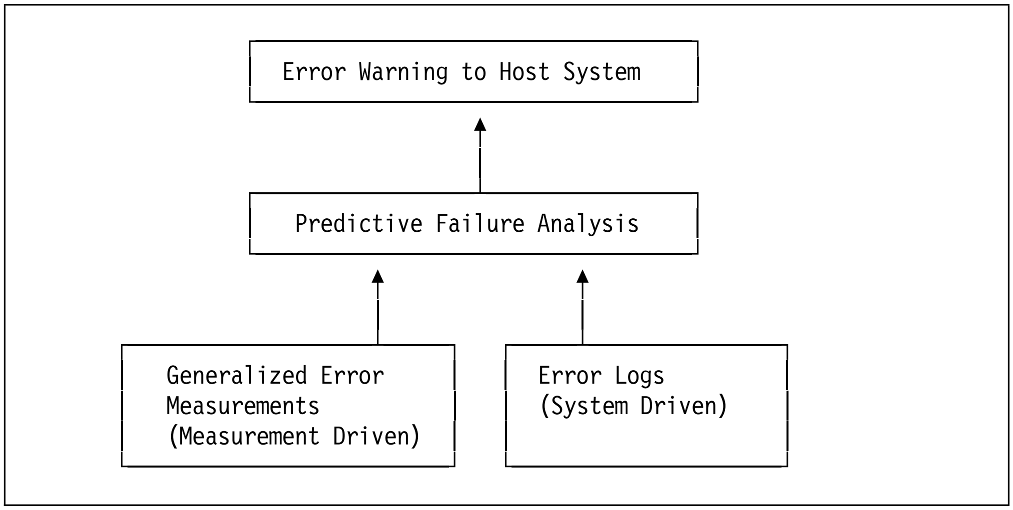


Figure 31. Predictive Failure Analysis

PRML algorithm determines from the GEM data what the likelihood will be that future disk operations can be completed successfully.

With this feature, specific mechanisms that can precede a disk drive failure can be directly detected. This results in a more accurate prediction of drive failure and can give a much earlier failure warning than can be obtained with conventional techniques using error monitors. This type of monitoring can give a mean warning time of 240 hours, or in other words, 10 days advance notice of a drive failure.

2. Symptom driven method

This PFA technique uses the output of data, non-data, and motor start error recovery logs. The analysis of the error log information is performed periodically during idle periods. When the analysis detects a failure that exceeds the predetermined threshold, the host system is notified.

This technique has recently been adopted by the industry and is referred to as SMART for Self Monitoring, Analysis, and Reporting Technology. This technique can typically provide up to 24 hours warning time before a drive fails.

5.6.1.1 Flying Height Monitoring

Consider an aircraft flying only inches above the ground at over 1000 Kmph. This analogy is frequently used to describe a recording head flying above a disk surface in a disk drive. Disk drives are designed so that no head-to-disk contact occurs during operation. Should contact occur, it can damage the magnetic layer on the disk surface. Such an event is known as a head crash. An important requirement for high reliability of the head-disk interface, therefore, is how much flying height exists. IBM has pioneered a way to measure head flying height using only the drive electronics.

From the time disk drives were invented by IBM, head flying heights have steadily decreased. Lower flying heights mean increased storage capacity. To achieve these lower flying heights requires good control of head tolerances. IBM has developed advanced manufacturing tests that ensure every head has the correct flying height before a drive is shipped.

What IBM has done with flying height monitoring is to measure flying height throughout the entire life of a drive. This means that when the flying height becomes unsuitable, a warning is sent by the drive. This warning criterion is designed so there is time to respond. Data on the drive can be transferred before the disk drive fails.

Flying height is a critical parameter for disk drive reliability. Heads flying too low are in danger of crashing. Heads flying too high are likely to experience higher data error rates. With flying height monitoring, potential drive failures due to flying height change can be predicted.

At periodic intervals, the Generalized Error Management circuitry automatically measures the flying height on all data surfaces. Measurements are made at many locations on each disk surface and these results are saved on the drive.

Using the measured flying heights, significant changes are detected using a statistical procedure to detect anomalies. Thus, heads that show a significant change in flying height, as compared to the other heads in the same drive, are suspect. Next, the magnitude of the change is analyzed to determine if it exceeds a predetermined threshold. If a head appears to be an anomaly and exceeds the threshold, a warning is issued.

Using flying heights as part of an early warning hard disk failure system can offer significant benefits. Methods based on error rate must rely on changes in error rate that also change with flying height. This correlation is very poor, however, because there are so many factors in a drive that affect error rate besides flying height.

Generalized Error Management and flying height monitoring are IBM exclusives and offer the best early-warning system for hard drive failure.

5.7 Overall Storage Subsystem Recommendations

- Use Fast/Wide controllers *and* drives.

  The hard disk is the slowest component in a server for accessing online data (excluding devices such as CD-ROMs, tape drives and the like); it is important that this access is as fast as possible. Ensure that you have fast/wide drives if you use a fast/wide adapter (and vice versa).

- Use PCI or Data Streaming MCA adapters.

  The PCI bus can achieve a burst transfer rate of 132 MBps and Micro Channel can achieve up to 80 MBps data transfer. However, EISA can only achieve 33 MBps and ISA only 16 MBps. Once the data is off the disks and into the controller, a fast bus is required to achieve the throughput required of fast processors. Use bus master devices.

- Use RAID technology to protect your data.

  When reliability is a must, use an IBM RAID controller with hot swap drives to protect your data. Disks are often among the least reliable components of the system, yet the failure of a disk can result in the irrecoverable loss of vital business data, or at the very least a need to restore from tape with consequent delays.

- Use IDE on smaller systems.

  IDE actually outperforms SCSI on systems where only one or two devices are attached. Several models of the IBM PC Server 310 and 320 lines implement

IDE as an integrated controller on the planar board. This is more than adequate if no more than two hard disks will be used.

- Use multiple SCSI channels.

  An IBM 2.25 GB SCSI-2 Fast/Wide drive can retrieve data from disk at a sustained rate of 5 MBps or more. With SCSI-2, the maximum transfer rate is 20 MBps. Therefore one SCSI channel could be fully used by four drives. Admittedly, a real-world system would usually not have a system such as this, but it shows that, for example, 15 SCSI-2 Fast/Wide drives on one channel would be very much under used. However, for maximum performance, use no more than four SCSI-2 Fast/Wide drives per SCSI channel.

- Distribute the workload on large systems.

  Research has shown that a single 66 MHz Pentium processor doing database transactions needs as many as 6-10 drives to optimize system performance. Therefore, do not determine the number of drives you need by simply adding up your total storage requirements and dividing this by the capacity of your drives. Instead, distribute the disk-intensive workload from a single physical disk drive to multiple disk drives and use the striping features of RAID technology.

Chapter 6. LAN Subsystem

The LAN adapter is another important component in the file server design. While there are many different types of LAN adapters, for file servers they fall into two main categories: bus master and shared RAM (non-bus master). The following discussion centers on the benefits of using bus master LAN adapters, although for small, lightly loaded LANs, non-bus master LAN adapters are quite adequate.

6.1 Shared RAM Adapters

Shared RAM adapters derive their name from the fact that they carry on-board RAM that is shared with the system processor. The memory on the adapter card is mapped into a reserved block of system address space known as the upper memory block (UMB) area. The UMB area is reserved for I/O adapters and is addressed between the addresses of 640 KB and 1 MB. The server processor can access this memory in the adapter in the same manner in which it accesses system memory.

Shared RAM can be 8, 16, 32, or 64 KB in size depending on which adapter is used and how it is configured. Adapter cards with 64 KB support RAM paging which allows the system to view the 64 KB of memory on the card in four 16 KB pages. This scenario only requires 16 KB of contiguous system memory instead of the 64 KB required when not using RAM paging. All IBM NetBIOS products support RAM paging.

The starting address of the shared RAM area is determined by the adapter device driver, switch settings, or in the case of an EISA or MCA adapter, via the setup utility or the reference diskette, respectively.

The main disadvantage of shared RAM architecture is that any data movement between the shared RAM area and system memory must be done under direct control of the system′s CPU. This movement of data to and from the shared RAM must be done because applications cannot operate on data while it resides in the shared RAM area. To compound matters, MOVE instructions from/to the shared RAM are much slower than the same MOVE instruction from/to the system memory because they occur across an I/O expansion bus. This means that when shared RAM adapters are involved, the CPU spends a significant amount of time doing the primitive task of moving data from point A to point B.

On lightly loaded servers providing traditional productivity applications such as word-processing, spreadsheets, and print sharing, this is not really a problem. But for applications such as databases or for more heavily loaded file servers, this can be a major source of performance degradation.

The IBM Token Ring Network 16/4 Adapters I and II for MCA and ISA are examples of shared RAM adapters.

6.2 Bus Master Adapters

Bus master adapters use on-board Direct Memory Access (DMA) controllers to transfer data directly between the adapter and the system memory without involving the system processor. The primary advantage of this architecture is that it frees up the system processor to perform other tasks, which is especially important in the server environment.

6.2.1.1 The IBM 16/4 Token-Ring Bus Master Adapter/A

This adapter was the first generation of bus master LAN adapters from IBM. It employed the 64 KB on-board adapter memory as a frame buffer that was used to assemble frames before they were sent to the server or sent from the server to the network. The time elasticity provided by this buffer allowed the token-ring chip set to complete its processing and forwarding of the frame before the frame was lost; this is a condition known as overrun (receive) or underrun (transmit).

This adapter was a 16-bit Micro Channel bus master capable of burst mode DMA. Due to the 24-bit addressing capabilities of the adapter, it was limited to using only the first 16 MB of system address memory.

6.2.1.2 IBM LANStreamer Family of Adapter Cards

The LANStreamer technology employs a completely different design to previous IBM LAN adapters. The LANStreamer utilizes a revolutionary chip set that is capable of processing token-ring frames without using memory as a frame buffer. It does it *on-the-fly* as the frames are passing through the adapter. Therefore, the latency of assembling frames from an on-card buffer is eliminated.

This low latency chip set is the key to the small-frame performance characteristics of the LANStreamer adapter.

The throughput for the LANStreamer Token-Ring MC 32 Adapter/A is quite high relative to its predecessors, especially for small frames. This is extremely important in client/server environments where research has shown that the vast majority of frames on the network are less than 128 bytes.

Another advantage of this technology is that since adapter memory buffers are no longer required, the adapter is less expensive to produce.

A consequence of the high LANStreamer throughput is that the LAN adapter is not usually the bottleneck in the system. Also, a side effect of using LANStreamer technology could be the higher CPU utilization. This sometimes happens because the LANStreamer adapter can pass significantly more data to the server than earlier adapters. This corresponds to more frames per second that must be processed by the server network operating system. Higher throughput is the desired effect but what this also means is that the bottleneck sometimes moves quickly to the CPU when servers are upgraded to incorporate LANStreamer technology.

Of course, other components can emerge as the bottleneck as throughput increases. The wire (network bandwidth) itself can become a bottleneck if throughput requirements overwhelm the ability of the network technology being used. For example, if an application requires 3 MBps of throughput, then a token-ring at 16 Mbps will not perform the task. In this case a different network technology must be employed.

The LANStreamer technology is used in the IBM Auto LANStreamer Adapters for PCI and MCA as well as the EtherStreamer and Dual EtherStreamer MC 32 LAN adapters.

EtherStreamer

The EtherStreamer LAN adapter supports duplex mode, which allows the adapter to transmit as well as receive at the same time. This provides an effective throughput of 20 Mbps (10 Mbps on the receive channel and 10 Mbps on the transmit channel). To implement this feature, an external switching unit is required.

6.3 PeerMaster Technology

The PeerMaster technology takes LAN adapters one step forward by incorporating an on-board Intel i960 processor. This processing power is used to implement per port switching on the adapter without the need for an external switch. With this capability, frames can be switched between ports on the adapter, bypassing the file server CPU totally.

If more than one card is installed, packets can be switched both within cards and between cards. The adapters utilize the Micro Channel to switch inter-card and can transfer data at the very high speed of 640 Mbps.

The IBM Quad PeerMaster Adapter is a four-port Ethernet adapter that utilizes this technology. It is a 32-bit Micro Channel bus master adapter capable of utilizing the 80 MBps data streaming mode across the bus either to/from system memory or peer-to-peer with another PeerMaster adapter.

The Quad PeerMaster is a *type 5* Micro Channel adapter. This refers to the physical size of the adapter. A type 5 adapter is 13.1 x 4.825 inches and is larger than normal MCA adapters (11.5 x 3.475 inches). It fits in specific servers and only in certain slots. Servers that support the type 5 adapters include the Server 320, 500 and 520. Refer to Part 3, “Server Products” on page 99 for more information on these servers.

It ships with 1 MB of memory. Each port on an adapter serves a separate Ethernet segment. Up to six of these adapters can reside on a single server and up to 24 segments can be defined in a single server.

This adapter can also be used to create virtual networks (VNETs). Using VNETs, the NOS sees multiple adapter ports as a single network, eliminating the need to implement the traditional router function either internal or external to the file server.

The Ethernet Quad PeerMaster Adapter is particularly appropriate when there is a need for:

- Switching/Bridging traffic among multiple Ethernet segments
- Attaching more than eight Ethernet 10Base-T segments to the server
- Attaching more than four Ethernet 10Base-2 segments to the server
- Providing switching between 10Base-T and 10Base-2 segments
- Conserving server slots

An add-on to PC SystemView (NetFinity) provides an advanced Ethernet subsystem management tool. Parameters such as packets/second or total throughput can be monitored for each port, for traffic within an adapter, or for traffic between adapters.

By using PC SystemView, you can graphically view the data, monitor for predefined thresholds, and optionally generate SNMP alerts.

Chapter 7. Security Features

This section discusses some technologies used in IBM PC Servers to comply with the United States Department of Defense (DoD) security requirements. Security features in the IBM PC Server line vary by model and all models do not have all the security features described here.

DoD requirements have been very influential in defining security standards used on computer system (both hardware and software) implementations around the world. The source for these requirements is the *Department of Defense, Trusted Computer System Evaluation Criteria, DoD 5200.28 STD*, dated 12/85. The essence of the requirements is contained in the Assurance section, Requirement 6: a “a trusted mechanism must be continuously protected against tampering and/or unauthorized changes...” The National Computer Security Center (NCSC) evaluates computer system security products with respect to the criteria defined by the U.S. Department of Defense.

There are seven computer system security product classifications in the DoD requirements: A1, B3, B2, B1, C2, C1, and D. The requirements for these classifications fall into four basic groups: security policy, accountability, assurance, and documentation. Several criteria, which vary by security classification, are specified in each of these groups. Currently, A1 is the highest classification, followed by B3, B2, and so on. The C2 classification satisfies most of the security requirements for personal computing environments.

7.1 LogicLock

On the IBM MCA PC Servers, IBM implements a collection of security features referred to as the LogicLock security system. LogicLock is designed to be hardware compliant with the C2 security classification. It goes far beyond basic PC security systems in its design to protect data against unauthorized access.

LogicLock security features include:

- Tamper-evident switches
- Optional secure I/O cables
- Privileged-access password
- Optional secure removable media
- Selectable drive startup
- Unattended start mode

7.1.1 Tamper-Evident Cover

Systems equipped with a tamper-evident cover have a key lock for their covers and internal I/O devices. In the locked position, it mechanically prevents the covers from being removed. The key has been changed to a type that can be duplicated only by the manufacturer.

If the covers are forced open, an electro-mechanical switch and perimeter sensor detect the intrusion. If the computer was on during the break-in attempt, depending on options specified during system setup, it will either defer action until the next IPL, lock up, or pass a non-maskable interrupt (NMI) to the software.

The next time the computer is started, the power-on self-test (POST) routine displays a message informing the user of the intrusion and requires that the automatic configuration program be run before the computer can be used. This is done to flag any configuration changes that may have occurred due to the intrusion (for example, removal of a disk drive). In addition, the system cannot be used without the privileged-access password if it has been set. There is a provision for maintenance that allows the system to be used without the covers in place. However, to use this feature, the key must have been used to remove the covers.

Other systems may have lockable covers. However, it is not that difficult to pry the system unit cover off, disable or unplug the key mechanism, and get inside the system. The tamper-evident mechanism is an important feature that flags the intrusion and prevents the operation of the system after a forced entry has occurred. This detection feature is very valuable for detecting the person most likely to break into the secured workstation, the user. Once the machine has been disabled, the system owner or administrator must be contacted to reset the system.

7.1.2 Secure I/O Cables

This rear-panel security option is an enclosure that is secured to the back of the computer by the cover lock. Its function is to prevent the cables from being removed and other cables from being attached. This effectively secures the serial, parallel, and SCSI cables, as well as other ports and cables provided by adapters. This is because it prevents someone from attaching a device through these connectors and gaining access to the data in the system.

The cable cover also has a tamper-evident feature.

7.1.3 Passwords

IBM PC Servers are equipped with several layers of password protection. The most basic is the power-on password. The power-on password must be entered correctly each time the system is turned on. After three incorrect attempts, the system must be turned off and back on in order to try again.

The keyboard password is another level of password protection and is used to lock the keyboard without turning the computer off. It also prevents rebooting the system by pressing the Ctrl+Alt+Del keys.

IBM PC Servers also provide an unattended server mode (or network server mode). This mode allows other computers to access a fixed disk drive on a server even though the keyboard is locked. This is useful, for example, when there is a power failure; the machine is able to recover with the keyboard lock still in place.

7.1.3.1 Privileged-Access Password

Because the power-on and keyboard passwords can be defeated by deactivating the battery inside the system, another level of password protection is provided. This security feature is called the privileged-access password. It provides a much higher level of security. The privileged-access password restricts access to system programs, prevents the IPL source and sequence from being changed, and effectively deters unauthorized modifications to the hardware. Also, if a forced entry is detected by the tamper-evident cover switch, the privileged-access password (if it has been set) must be used in order to make the system operational again.

The privileged-access password is stored in a special type of read only memory called flash EEPROM. EEPROM is an acronym for electrically erasable programmable read only memory.

Systems are shipped with the privileged-access password disabled. To set this password, a jumper on the system board must be moved in order to put the system in the change state. Once this password is set, it cannot be overridden or removed by an unauthorized person.

Forgotten Password

If the administrator misplaces or forgets the privileged-access password, the system board will have to be replaced. There is no way to reset a forgotten privileged-access password.

7.1.4 Secure Removable Media

An optional 2.88 MB diskette drive with security features is available on some IBM PC Server systems. The diskette drive is a 3.5-inch, one-inch high drive with media sense capability for the standard diskette capacities of 720 KB, 1.44 MB, and 2.88 MB. It can read and write data up to a formatted capacity of 2.88 MB, while maintaining read and write capability with 720 KB and 1.44 MB diskette drives.

A control signal has been added to the diskette interface that supports LOCK, UNLOCK, and EJECT commands issued by the operating system. If the privileged-access password is not set, the diskette is unlocked during POST. If the password is set, the boot process does not unlock the diskette drive unless it is the designated IPL source. In this case, the LOCK and UNLOCK state is controlled by an operating system utility. For SCSI devices, there is a proposed standard UNLOCK command. In this case, the operating system will control the LOCK command if the privileged-access password is set. Access to the unlocking function with specific user authorization can be controlled by secured system software.

In the event of power loss, the system retains its state (secured or unsecured) independent of the state of the battery. A diskette can be inserted in the drive, but it cannot be removed if the power is off. When the drive is turned on and locked, the media cannot be inserted or removed.

7.1.5 Selectable Drive Startup

Selectable drive startup allows the system owner or administrator to select the IPL source and sequence. This allows the system owner to control the IPL source, but prevents the user from modifying the source and sequence. For example, the diskette drive can be excluded as an IPL source. This feature helps to ensure that the system owner′s specified operating system is loaded.

The IPL sequence is stored in the system EEPROM and can only be changed using the privileged-access password. Storage of the IPL sequence in the EEPROM protects it from being deactivated by removing the battery. The setup routine ensures that at least one IPL source is specified if the privileged-access password is used.

7.1.6 Unattended Start Mode

The unattended start mode automatically restarts the server after a power failure and resumes normal operation, without operator intervention.

It locks the keyboard when the system is powered on, but it allows the operating system and startup files to be loaded. The keyboard remains locked until the power-on password is entered.

This mode is useful for unattended operations because it allows authorized network user access to information on the server but prohibits unauthorized access via the system keyboard.

When the system is in the unattended mode, the password prompt will not appear unless an attempt to start the system from a diskette or other removable media is issued. If you start the system from a removable media, the password prompt will appear and you must enter the correct power-on password to continue.

Chapter 8. Systems Management

Systems management is an important element of a successful LAN. Because of this, the IBM PC Server brand ships with TME 10 NetFinity (also known as NetFinity and PC SystemView). TME 10 NetFinity is a very powerful systems and network management tool.

In this chapter, we look at the capabilities of TME 10 NetFinity as well as some of the underlying technologies such as DMI and SNMP that TME 10 NetFinity has incorporated. DMI is an emerging standard for managing desktop machines and SNMP is an established network management protocol.

We also describe IBM ServerGuide, a package of CD-ROMs whose purpose is to make installing network operating systems a simple task.

Note

TME 10 NetFinity is the new name for PC SystemView and NetFinity. With the IBM acquisition of Tivoli Systems, Inc., several IBM products including PC SystemView are being renamed to fall in line with the new Tivoli brand, TME 10. If you are already familiar with NetFinity or PC SystemView, please rest assured that the industry-leading technology within these two products are still intact within TME 10 NetFinity.

For detailed information on the use of NetFinity, please see the IBM Redbook, *Systems Management from an NT Server Point of View*, SG24-4723. While the book is targeted for an NT audience, the NetFinity function discussed is the same for all the NetFinity supported environments.

8.1 Desktop Management Interface

The Desktop Management Interface (DMI) is a standard developed by an industry consortium that simplifies management of hardware and software products attached to, or installed in, a computer system. The computer system can be a stand-alone desktop system, a node on a network, or a network server. DMI is designed to work across desktop operating systems, environments, hardware platforms, and architectures.

DMI provides a way to obtain, in a standardized format, information about the hardware and software products installed in the system. Once this data is obtained, management applications written to the DMI specs can use this data to manage those products. As DMI technology evolves, installation and management in desktop PCs and servers will become easier.

It should be noted that the DMI specification says nothing about the transport protocol that is used between the manageable products and the management applications. Both of these elements of a DMI compliant system can be implemented using any native transport protocol available in the system.

The DMI architecture includes:

- Communicating service layer
- Management information format (MIF)
- Management interface (MI)

- Component interface (CI)

8.1.1 Communicating Service Layer

The service layer is the desktop resident program that is responsible for all DMI activities. Service layer communication is a permanent background task or process that is always ready for an asynchronous request.

The service layer is an information broker handling commands from management applications, retrieving the requested information from the MIF database or passing the request on to manageable products as needed via the CI. The service layer also handles indications from manageable products and passes that information on to the management applications.

Management Applications: These are remote or local programs for changing, interrogating, controlling, tracking and listing the elements of a desktop system and its components.

A management application can be a local diagnostic or installation program, a simple browser that walks through the MIF database on the local system or any other agent that redirects information from the DMI over a network.

Manageable Products: These include hardware, software or peripherals that occupy or are attached to a desktop computer or network server, such as hard disks, word processors, CD-ROMs, printers, mother boards, operating systems, spreadsheets, graphics cards, sound cards, modems, etc.

Each manageable product provides information to the MIF database by means of a file that contains the pertinent management information for that product. Manageable products, once installed, communicate with the service layer through the component interface. They receive management commands from the service layer and then return information about their status back to the service layer.

8.1.2 Management Information Format (MIF)

A management information format (MIF) is a simple ASCII text file describing a product′s manageable attributes, grouped in ways that make sense. The MIF has a defined grammar and syntax. Each product has its own MIF file.

When a manageable product is initially installed into the system, the information in its MIF file is added to the MIF database and is available to the service layer and thus to management applications.

The simplest MIF file contains only the component ID group, but MIFs can become as complex as needed for any given product.

8.1.3 Management Interface (MI)

The management interface (MI) shields management applications from the different mechanisms used to obtain management information for products within a desktop system.

The MI allows a management application to query for a list of manageable products, access specific components and get and set individual attributes. Additionally, the MI allows a management application to tell the service layer to send back information about indications from manageable products.

The MI commands provide three types of operations to control manageable products:

- **Get:** allows a management application to get the current value of individual attributes or group of attributes
- **Set:** allows writable attributes to be changed
- **List:** allows management applications to read the MIF descriptions of manageable products, without having to retrieve the attribute values for that product

Thus, a management application can query a system and retrieve useful information about the contents of the system, with no previous knowledge of that system.

8.1.4 Component Interface (CI)

The component interface (CI) handles communication between manageable products and the service layer. The CI communicates with manageable products for get and set operations. It also receives indications from manageable products and passes those to the MI. Active instrumentation allows components to provide accurate, real-time information whenever the value is requested. A single component attribute can have a single value, or it can be obtained from a table using index keys.

8.2 SNMP

Simple Network Management Protocol (SNMP) is a network management protocol defined within the TCP/IP transport protocol standard. It is a rather generic protocol by which management information for a wide variety of network elements may be inspected or altered by logically remote users. It is a transaction-oriented protocol based on an interaction between managers and agents. The SNMP manager communicates with its agents. Agents gather management data and store it, while managers solicit this data and process it.

The SNMP architectural model has been a collection of network management stations and network elements such as gateways, routers and hosts. These elements act as servers and contain management agents that perform the network management functions requested by the network elements. The network management stations act as clients; they run the management applications that monitor and control network elements.

SNMP provides a means of communicating between the network management stations and the agents in the network resources. This information can be status information, counters, identifiers, etc.

The SNMP manager continuously polls the agents for error and statistical data. The performance of the network will depend upon the setting of the polling interval.

8.2.1.1 Management Information Base (MIB)

The Management Information Base (MIB) is a collection of information about physical and logical characteristics of network objects. The individual pieces of information that comprise a MIB are called MIB objects and they reside on the agent system. These MIB objects can be accessed and changed by the agent at the manager′s request.

The MIB is usually made up of two components:

1. MIB II

   This is standard definition that defines the data layout (length of fields, what the field is to contain, etc.) for the management data for the resource. An example would be the resource name and address.

2. MIB Extension

   This incorporates unique information about a resource. It is defined by the manufacturer of the resource that is being managed. These are usually unique and proprietary in nature.

8.2.1.2 SNMP Agent

The SNMP agent is responsible for managed resources and keeps data about the resources in a MIB. The SNMP agent has two responsibilities:

1. To place error and statistical data into the MIB fields
2. To react to changes in certain fields made by the manager

8.2.1.3 SNMP Manager

An SNMP manager has the ability to issue the SNMP commands and be the endpoint for traps being sent by the agent. Commands are sent to the agent using the MIB as a communication vehicle.

8.2.1.4 Traps

In a network managed with SNMP, network events are called traps. A trap is generally a network condition detected by an SNMP agent that requires immediate attention by the system administrator. It is a message sent from an agent to a manager without a specific request for the manager.

SNMP defines six generic types of traps and allows definitions of enterprise-specific traps. This trap structure provides the following information:

- The particular agent object that was affected
- Event description (including trap number)
- Time stamp
- Optional enterprise-specific trap identification
- List of variables describing the trap

In summary, the following events describe the interactions that take place in an SNMP-managed network:

- Agents maintain vital information about their respective devices and networks. This information is stored in a MIB.

- The SNMP manager polls each agent for MIB information and stores and displays this information at the SNMP manager station. In this manner, the system administrator can manage the entire network from one management station.

- Agents also have the ability to send unsolicited data to the SNMP manager. This is called a trap.

8.3 TME 10 NetFinity

TME 10 NetFinity, or simply NetFinity, is IBM′s comprehensive hardware systems management environment for IBM PC Servers. It provides an easy-to-use graphical set of local and remote services designed to make the PC Server and client systems simple and affordable to manage. It offers many features such as:

- Agent (services) and manager support for common operating environments
- Remote systems management
- Power-On error detection and analysis
- Predictive failure analysis (PFA) for disk drives
- Event scheduler
- Software inventory
- Full database export

TME 10 NetFinity has a flexible, modular design that allows for a variety of system-specific configurations. NetFinity is able to manage IBM and non-IBM desktop PCs and servers and supports most client operating systems. Agents (or services in NetFinity parlance) exist for Warp Server, NetWare, and Windows NT Server as well. The NetFinity manager code can run on OS/2, Windows, and Windows NT. NetFinity is designed to work with the existing protocols on the network and includes support for NetBIOS, IPX, TCP/IP, and even ASYNC/Serial modem LAN protocols.

TME 10 NetFinity is delivered as two components:

- NetFinity Services
- NetFinity Manager

8.3.1.1 TME 10 NetFinity Services

TME 10 NetFinity Services is the client portion of the system. This is a foundation that provides the underlying services for several levels of administration, including remote system and user management facilities.

NetFinity Services provides the following functions:

- *System Information* - provides details regarding specific hardware and software configurations and user/application settings
- *System Profile* - allows the systems administrator to define additional information for each system, such as location
- *System Monitor* - provides system performance monitoring utilities, such as CPU, DASD, and Memory
- *Critical File Monitor* - can generate alerts when critical files are changed or deleted
- *System Partition Access* - allows an administrator to access the system partition on remote computers
- *Predictive Failure Analysis* - monitors PFA-enabled drives for errors
- *ServerGuard Service* - monitors server environmental parameters such as temperature and voltage and can remotely boot the system (requires the ServerGuard Adapter)

- *RAID Manager* - allows the administrator to view the array configuration
- *ECC Configuration* - allows the administrator to set thresholds for ECC scrubbing, counting, and NMIs
- *Security Manager* - controls which NetFinity Services each manager can access for a given system
- *Alert Manager* - fully customizable alert generation, logging, and forwarding, and also has the ability to filter and forward SNMP alerts
- *Serial Control Service* - supports an ASYNC connection between two NetFinity systems, either managers or clients

Note

The graphical user interface (GUI) program files for ECC configuration, System Partition Access, RAID administration and Predictive Failure Analysis will be installed regardless of whether your system has the associated hardware features installed. This enables a network administrator to remotely access these services on other systems.

TME 10 NetFinity Services supports both IBM and non-IBM systems. It supports PCI, Micro Channel, and EISA bus-based systems. It supports most client operating systems including DOS/Windows, Windows for Workgroups, Windows 95, Windows NT, OS/2 2.X, OS/2 Warp, OS/2 Warp Connect, and OS/2 SMP.

It also supports Novell NetWare. This means that there is a version of TME 10 NetFinity Services that installs as a NetWare NLM on the file server and allows the server to be managed by a NetFinity Manager station.

NetFinity Services can also be installed on a Windows NT server and used to manage this platform as well.

TME 10 NetFinity Services can be configured in three client modes of operation:

- Stand-alone client

  Stand-alone mode allows an individual user, who is not connected to a network, to effectively manage or monitor their own system including hardware, resources and performance.

- Passive client

  With the passive client installed on a LAN workstation, a NetFinity Manager is able to fully manage and monitor the resources and configuration setting of the workstation. However, with the passive mode installed, that same client is not able to perform its own management task locally. This mode is most effective for LAN administrators who do not want individual users to have management capability on an individual basis.

- Active client

  The active client allows the TME 10 NetFinity Manager to manage and monitor the resources and configuration setting of the workstation. In comparison to the passive client mode, the active client mode allows local users to perform their own subset of local systems management tasks.

8.3.1.2 TME 10 NetFinity Manager

The NetFinity Manager is the set of applications that is installed on the managing platform. It automates the collection of data from managed clients and archives it into a database, which maintains specific, unique workstation data and configuration settings. TME 10 NetFinity also supports database exports into Lotus Notes or DB2/2.

In addition to logging the information in a database, an administrator may dynamically monitor performance on client workstations. An administrator may also identify resource parameters to monitor and maintain.

NetFinity Manager has the ability to discover LAN-attached client workstations automatically. For example, if a new NetFinity client appears on the LAN, it will be sensed by the manager services and, from that point on, will be automatically included as a managed device within the profile.

A profile is a set of managed devices grouped by a set of unique attributes such as system processor types, operating systems, installed transport protocols, and administrator defined keywords. The keywords can be descriptors of systems, users or profiles. These NetFinity profiles can be dynamically declared, reset and maintained on an *as needed* basis by the administrator.

TME 10 NetFinity Manager includes the following functions:

- *Remote Systems Manager* - allows managers to access remote TME 10 NetFinity-managed machines on a LAN, WAN, or a serial link. The manager can access the NetFinity Services as if the manager was at that machine.
- *Manageability via WWW* - manage LAN-connected PCs across the Internet from any PC or workstation with a web browser, including Apple MacIntosh and UNIX workstations.
- *File Transfer* - send/receive files to the remote system.
- *Remote Session* - open a remote console to the managed device.
- *Screen View* - take a snapshot of any screen on the remote device.
- *DMI Browser* - view information about DMI-compliant hardware and software.
- *Process Manager* - start/stop/view processes running on the managed device.
- *Software Inventory* - scan remote device for installed software using a software dictionary.
- *POST Error Detect* - detect and log errors at Power on System Test (POST) time on managed devices.
- *Event Scheduler* - automate the execution of a service on one or multiple systems in a profile.

8.3.1.3 TME 10 NetFinity Architecture

Each NetFinity service is comprised of two separate components. One is a unique graphical user interface for the applicable operating system. The second is a native operating system executable, which is known as the *base executable*. The base executable is the code that performs the client management and monitoring tasks for each unique workstation. Communication between the GUI and the base executable is handled by the NetFinity IPC (inter-process communication) mechanism.

Using this IPC within the LAN, NetFinity was designed to provide a peer-to-peer platform architecture, which does not require a management server or a dedicated management console. From this design, a manager may take control of the NetFinity client system to perform all NetFinity administrative and problem reconciliation tasks as if they were the local user′s tasks. Additionally, IBM has been able to isolate NetFinity from any network, protocol or operating system layer dependencies. In essence, IBM uses the existing transport layers within the installed network to allow NetFinity to communicate between managers and services. Since IPC resides on top of the Media Access Control (MAC) layer, it simply communicates between the installed modules and services, utilizing the transport mechanism within the workstation.

If the transport layer between the two NetFinity workstations is dissimilar, then NetFinity utilizes a mapper (within a NetFinity manager), which receives data packets from one transport and, using NetFinity Manager, is able to re-wrap the packets for transport into the foreign network.

When two TME 10 NetFinity systems are connected in a networked environment, they communicate using the IPC into the mapper and then subsequently into a NetFinity Manager services and system module. This feature provides an extensive capability to merge dissimilar LANs into a single view of NetFinity managed assets.

8.3.1.4 DMI Support

TME 10 NetFinity is the first product available to customers that includes DMI support. NetFinity implementation of DMI support provides instrumentation from its System Information Tool to the DMI service layer for both OS/2 and Windows clients. To accomplish this, IBM has delivered a DMI component agent that allows a manager to access a client desktop MIF database to deliver system specific information back into the DMI browser. Today, NetFinity not only supports local DMI browsing capabilities, but also DMI alerting and a remote browser service.

8.3.1.5 Interoperability with Other Management Tools

NetFinity supports coexistence with almost any other LAN or enterprise management product, whether from IBM or other vendors. To provide for this integration, NetFinity Alert Manager was developed to allow its alerts to be captured and forwarded into any SNMP compliant management application. SNMP alerts are recognizable by the vast majority of different management tools in the market today. With this NetFinity feature, administrators can integrate sophisticated systems management functions with their existing SNMP-based systems management applications, such as IBM SystemView and HP OpenView. In direct support of heterogeneous LAN management environments, NetFinity is also launchable from within NetView for OS/2 and Novell NMS (NetWare Management Services).

8.3.1.6 TME 10 NetFinity 4.0 New Functions

The following new functions have been added to TME 10 NetFinity Version 4.0:

- Full support for Windows NT Server Version 3.51 in manager and services (agent) roles.
- A Remote Manager function for management of TME 10 NetFinity systems over an Internet connection (requires a Web browser and HTML support)

- ODBC database support to allow data export from TME 10 NetFinity to a wide range of popular databases, including Microsoft SQL.
- An enhanced Desktop Management Interface (DMI) to include an MIF for monitor information, for local or remote browsing.
- An enhanced RAID utility that allows remote configuration of RAID arrays inside a server or an external RAID enclosure.
- An improved Alert Manager to ease the selection and monitoring of alerts received by the TME 10 NetFinity Manager.
- Support for APC PowerChute Plus that allows you to control and monitor an APC Smart-UPS (available as an option through IBM). In addition, APC provides extensions to TME 10 NetFinity to allow full management of the UPS through NetFinity.
- IBM AntiVirus 2.5 software that protects the server from over 8000 viruses.

For detailed information on the use of NetFinity, please see *Systems Management from an NT Server Point of View*, SG24-4723. While the book is targeted for an NT audience, the NetFinity function discussed is the same for all the NetFinity supported environments.

8.4 ServerGuide

ServerGuide is a set of CD-ROMs containing programs and utilities designed to make the NOS installation easier. It is provided free of charge with each new IBM PC Server.

Operation of ServerGuide is as simple as inserting the first CD from the ServerGuide package into the CD-ROM drive, then inserting the diskette into the diskette drive, and starting your IBM PC Server. An easy-to-use graphical interface guides you through the simple installation steps.

During installation, ServerGuide automatically detects and identifies specific hardware configurations and fills in all configuration choices with sensible defaults. You can simply accept ServerGuide′s suggestions, or if you want, change them to reflect different requirements. ServerGuide then automatically installs the necessary drivers.

After using ServerGuide to install the operating system and network operating system of your choice, you can use ServerGuide′s additional utilities and extensive online reference documentation and diskette images to support and enhance the day-to-day running of your LAN.

You can use ServerGuide to install and configure either the encrypted programs contained on ServerGuide CDs or retail versions you already own. You can purchase software keys from IBM to unlock the encrypted programs. For these retail versions, no encryption key is required.

ServerGuide Version 3.0 is the current version as of the date of this publication. ServerGuide 3.0 offers installation assistance for the following retail packages:

- OS/2 Warp Server Version 4.0
- OS/2 Warp Server Advanced Version 4.0
- Windows NT Server 4.0
- NetWare from IBM Version 3.12

- NetWare from IBM Version 4.1
- NetWare from IBM Version 4.1 SMP for 1 to 4 processors
- NetWare from IBM Version 4.1 SMP for 4 additional processors
- NetWare from IBM Version 4.11

The following sections describe the programs and utilities that are supplied with ServerGuide.

8.4.1 Network Operating System (NOS) Installation

Use ServerGuide to help you install an operating system from an existing license, or (optionally) using a key purchased through ServerGuide. Using ServerGuide ensures that you have all the device drivers and fixes you need for a successful installation on your server. ServerGuide supports the latest releases from OS/2, Novell, Microsoft, and SCO. Figure 32 shows the main menu for NOS installation.

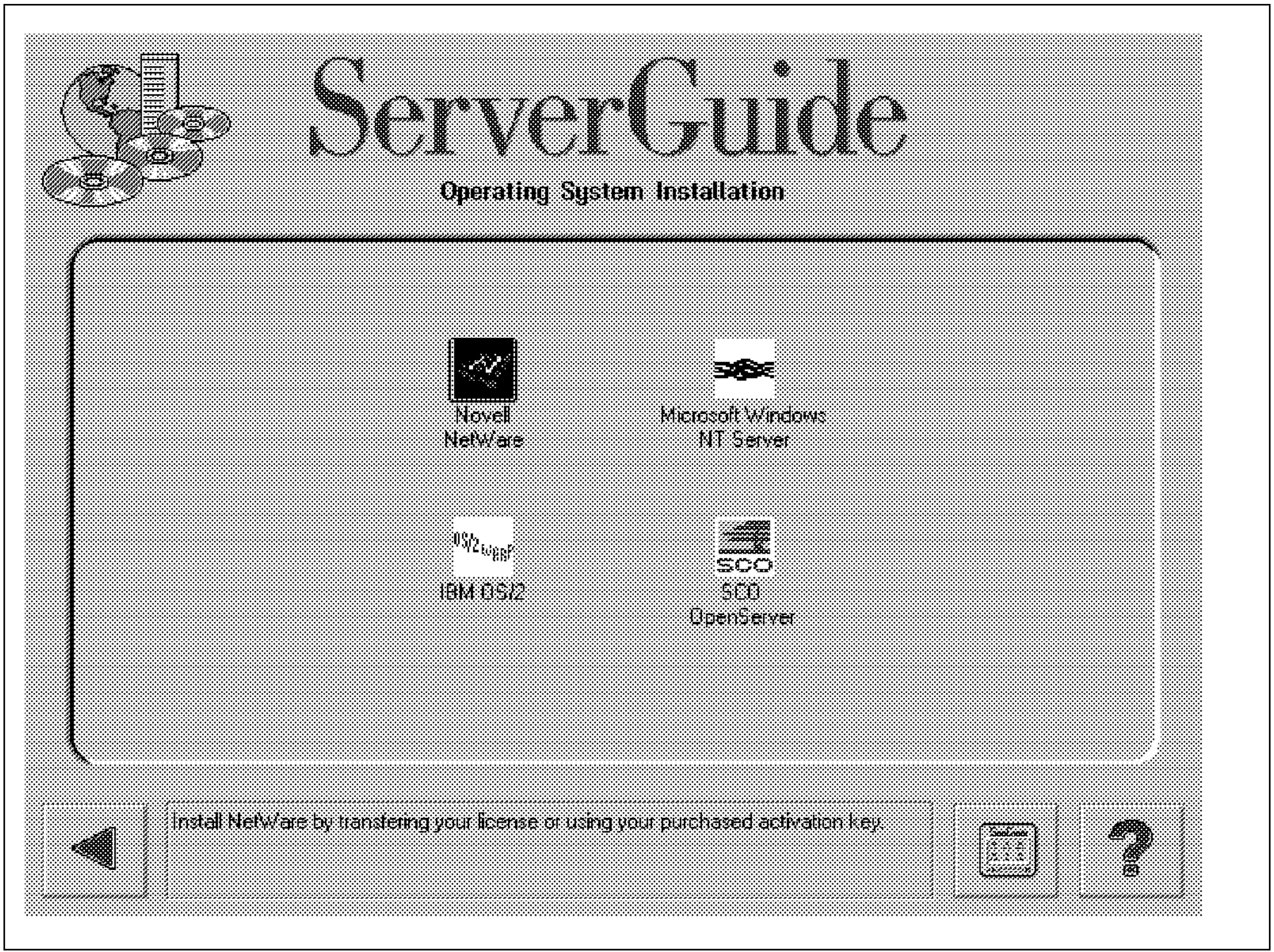


Figure 32. Main Menu for NOS Installation

8.4.2 AutoPilot

AutoPilot allows you to learn about ServerGuide and be guided step-by-step through the hardware configuration of the server. It includes checking (and if necessary updating) the BIOS level, configuring the adapters, and configuring the disk arrays. Online help is readily available to guide you through the process, if necessary.

Figure 33 shows the AutoPilot main menu.

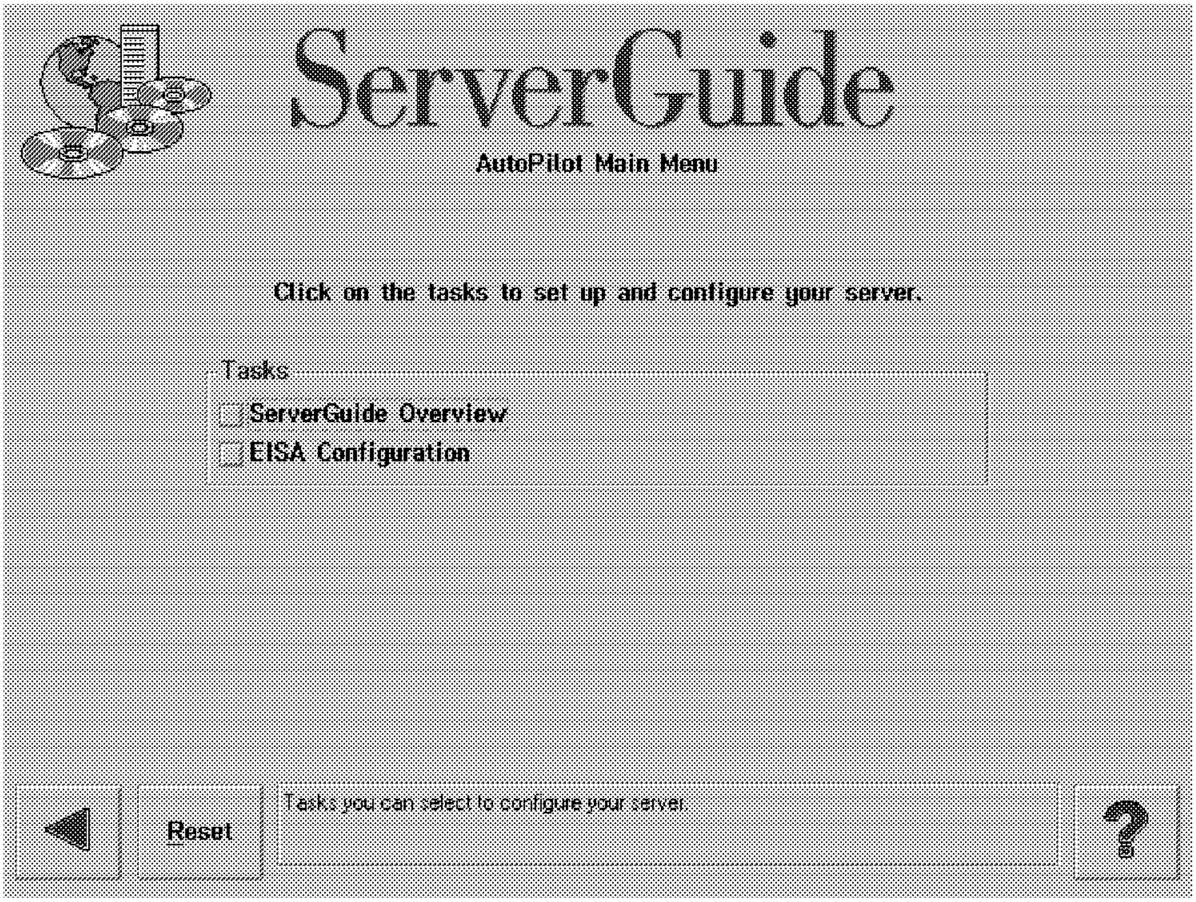


Figure 33. AutoPilot Main Menu

8.4.3 Diskette Factory

Diskette Factory is used to quickly create diskettes containing device drivers, configuration programs, and additional utilities and options for your PC Server. Figure 34 on page 96 shows the Diskette Factory main menu.

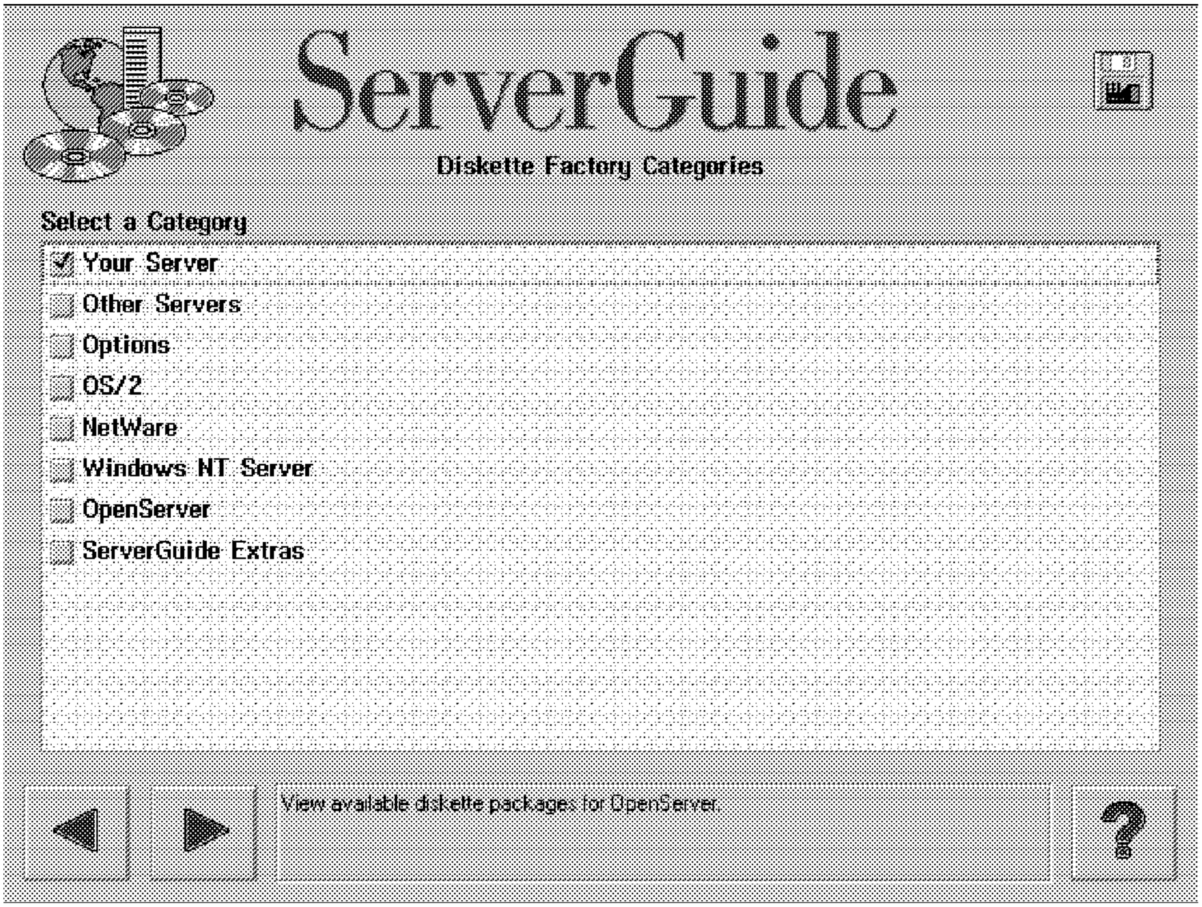


Figure 34. Diskette Factory Main Menu

8.4.4 CoPilot

CoPilot is a set of CDs that can be used after installation of the NOS. The CDs contain a range of free and encrypted software. For example, when you purchase an IBM PC Server you receive Lotus Notes Release 4.1 Server, Single Processor Edition, at no additional charge. This release of Notes contains national language versions also at no additional charge.

Lotus Notes Release 4.1 provides built-in Internet integration. The end user can browse the Web and seamlessly incorporate HTML documents into Notes documents using the InterNotes Web Navigator. InterNotes Web Publisher enables businesses to create, manage, and administer their internal intranet and public Web sites by using application development facilities in Notes to easily build and host home pages and applications on the Web.

8.4.5 TME 10 NetFinity 4.0

NetFinity is a highly-flexible systems management program designed to allow LAN administrators to easily view, initiate, and exploit management services for LAN-connected clients and servers. Please see 8.3, “TME 10 NetFinity” on page 89 for more information on NetFinity.

8.4.6 Documentation

ServerGuide includes the following online documentation:

- Hardware manuals for PC Server 310, 320, 325, 330, 500, 520, 704 and 720
- Integration guides for NT and Novell NetWare
- TME 10 NetFinity documentation

These online documents can be viewed or printed to a PostScript printer.

8.4.7 Packaging

The ServerGuide Version 3.0 package includes the following:

- One ServerGuide License diskette
- Four blank diskettes
- 13 ServerGuide CDs:
  - ServerGuide Main CD
  - OS/2 for Symmetric Multiprocessing 1 - 2 processors CD
  - OS/2 for Symmetric Multiprocessing 1 - 7 processors CD
  - Windows NT Server 3.51 CD
  - NetWare from IBM Versions 3.12 and 4.1 and SMP CD
  - NetWare Non-English Documentation CD
  - SCO Open Server Enterprise 5.02 CD
  - Diskette Factory CD
  - Book Factory CD
  - Fix Pack CD
  - CoPilot CD 1
  - CoPilot CD 2
  - CoPilot CD 3
- ServerGuide Publications
  - ServerGuide Getting Started
  - ServerGuide product order form
  - ServerGuide Program License Agreement booklet

The program package for different geographies will be the above package plus specific inserts for conditions of use and local requirements.

Part 3. Server Products

Chapter 9. Introduction

The IBM PC Server family contains three product lines that offer different features and capabilities:

1. The PC Server 300 series

   This series is targeted at small enterprises or workgroup LANs. These machines offer leading technology and are very price competitive. They are more limited in terms of upgrade and expansion capabilities than the other two lines in the family. Current models in the 300 series include the PC Server 310, 320, 325, and 330.

2. The PC Server 500 series

   This series is targeted for medium to large enterprises who need more power and more expansion capabilities. With 22 storage bays available (18 of which are hot swappable), these machines are very suitable for the enterprise server. The PC Server 520 is the current model in this series.

3. The PC Server 700 series

   The super servers in this series are targeted for large enterprises and for those who need the maximum computing power in a PC Server environment. The 704 and 720 are the two models currently in this product line. The PC Server 704 offers four-way Pentium Pro power for compute-intensive applications. The PC Server 720, with its 6-way Pentium multiprocessing capability, is very suitable for the application server environment. Both offer state of the art technology and wide expansion capabilities.

In conjunction with these Server series, there are external expansion solutions designed to enhance the storage capacity of your servers:

- 3517 SCSI Multi-Storage Enclosure - Offering 7 drive bays for up to 22.5 GB of storage.
- 3518 PC Server Enterprise Expansion Enclosure - Offering 18 hot-swap drive bays for up to 40 GB of storage.
- 3527 SSA Entry Storage Subsystem - Offering 5 bays for SSA devices for up to 22.5 GB of storage.

To allow efficient site management, the 9306 PC Server Rack Enclosures provide the ability to house your servers more efficiently by stacking the systems within a 24″ enclosure, allowing the sharing of devices such as keyboard, screen and mouse.

Each of these products is discussed in this chapter. Special mention will also be made of the range of SCSI adapters, SCSI hard drives and SCSI tape drives as they apply to the PC Server environment.

Note: The servers described here may not be available in all countries. Similarly, other servers (for example, the PC Server 300) may still be available in the country where you live.

The server family has a number of features common to all of its members:

- Pentium and Pentium Pro microprocessors - Each of the servers is based on Pentium and Pentium Pro technology from a singe Pentium 133 MHz processor in the entry-level machines to 6-way Pentium 166 MHz and four-way Pentium Pro 200 MHz processor-based systems at the high end.

- SCSI performance - Each server has an SCSI-2 or an UltraSCSI storage subsystem. RAID controllers are standard on some models for added performance and security. Serial Storage Architecture (SSA) is available as an option.
- SVGA video - All models in the family offer 0.5 MB or 1MB super video graphics array (SVGA) subsystems for displaying high resolutions and colors. This is a benefit especially where systems and network management are performed from the server itself.
- CD-ROM drive - Each server is configured with either a double-speed (2x), quad-speed (4x), or 8x CD-ROM drive to make it easier to install software.
- TME 10 NetFinity - This is a comprehensive systems management tool that allows LAN administrators to monitor and manage servers and workstations. For more information about TME 10 NetFinity, please refer to 8.3, "TME 10 NetFinity" on page 89.
- ServerGuide - This is a set of CD-ROM disks that contain the most popular operating systems and management tools such as NetFinity. It provides a simple interface to install and configure the operating system and tools. For more information on IBM ServerGuide, please see 8.4, "ServerGuide" on page 93.
- Enhanced keyboard and mouse - Supplied standard with each server.

Each of the server products is described in the following chapters.

Chapter 10. IBM PC Server 310

The Server 310 is positioned as an entry-level workgroup server for small organizations or departments requiring file and print services. It is equipped with all the necessary features for this purpose including an Ethernet adapter which is standard on all models except one.

The models in the Server 310 range fall into two categories:

1. Pentium 133 MHz models
2. Pentium 166 MHz models

10.1 Technical Description

The following sections describe each of the major subsystems.

10.1.1.1 Intel Pentium Processor

All models are equipped with either a 133 MHz or a 166 MHz Pentium processor. The 166 MHz models are upgradable to a 200 MHz processor.

10.1.1.2 Cache

A 256 KB level 2 cache is standard on all models:

- xDx models use adaptive write-back cache.
- xEx models use synchronous write-back and are upgradable to 512 KB.
- All models use 12 ns SRAM chips.

10.1.1.3 Memory

On Models 0E0 and 0EV:

- 16 MB 60 ns RAM is standard.
- ECC is implemented in the memory controller.
- One 168-pin gold-plated DIMM and four 72-pin tin/lead SIMM sockets are present.
- A total of 160 MB is installable using one 32 MB DIMM and four 32 MB SIMMs.

All other models:

- 16 MB 70 ns parity RAM is standard.
- Six 72-pin gold-plated SIMM sockets are present.
- Two sockets are used by standard 8 MB SIMMs with four available.
- A total of 192 MB is installable using six 32 MB SIMMs.
- ECC-on-SIMM memory is optional. (Refer to 4.3.4, "ECC-on-SIMM (EOS) Memory" on page 42 for more information on this technology.)

Note: Table 8 on page 107 shows the IBM supported memory options for use with the Server 310. For more information concerning the usage of various memory SIMM technologies, please see 4.3.8, "General Tips on Memory Usage" on page 45.

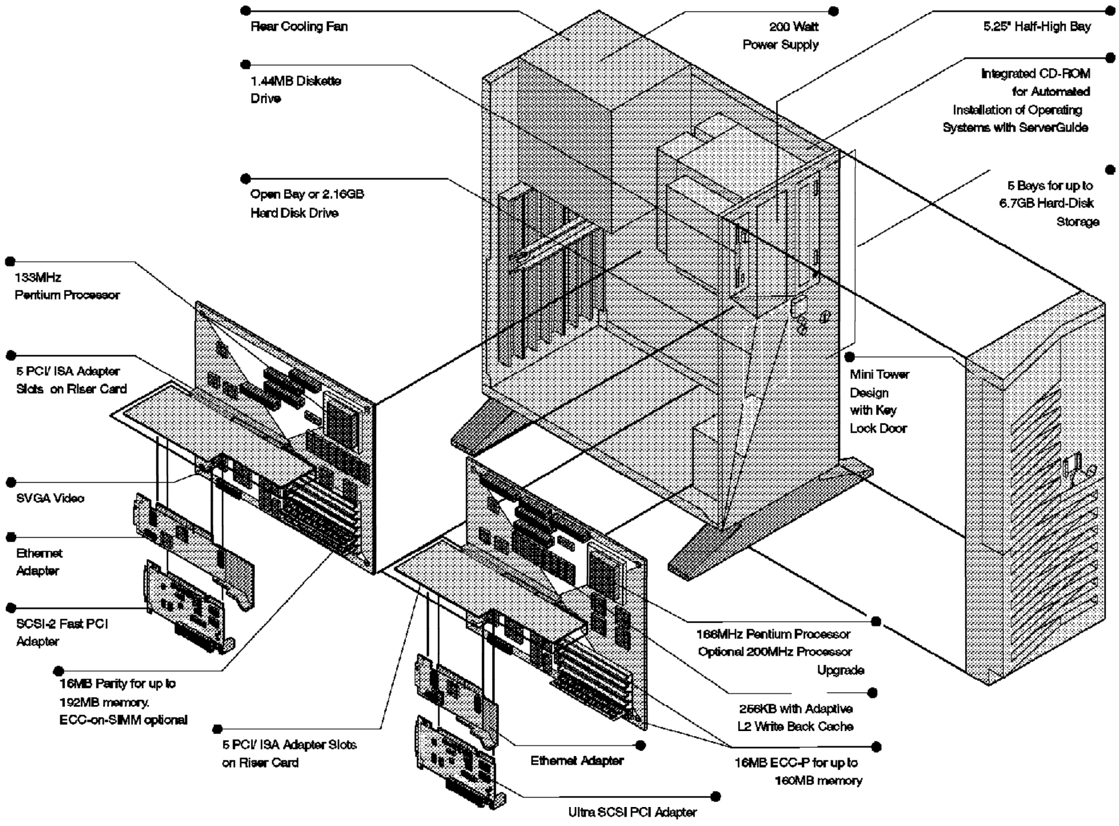


Figure 35. Exploded View of the IBM PC Server 310

10.1.2 CD-ROM

An 8x SCSI CD-ROM drive is installed in the newer models of the PC Server 310. It has the following characteristics:

- Average access time of 150 ms
- Average data transfer rate of 1200 KBps
- Burst rate of 5.0 MBps
- Buffer size of 256 KB

A quad-speed SCSI CD-ROM drive is standard on the older models. It has the following specifications:

- Average access time: 200 ms
- Average data transfer rate: 600 KBps
- Maximum transfer in burst mode: 5 MBps
- Buffer size: 256 KB

10.1.3 Ethernet Adapter

An Ethernet adapter is standard with all models of the PC Server 310 except for Model 0E0. Model 0EV uses an IBM EtherJet Adapter, part #72H2613, and all other models use IBM LAN Adapter for Ethernet, part #60G0605. Both adapters support 10Base-T, 10Base-2 and AUI media.

Table 8. Server 310 Memory Options

Size	Parity SIMMs (xDx models)	ECC-on-SIMM (xDx models)	ECC DIMMs (xEx models)	ECC-P SIMMs (Tin/Lead) (xEx models)
16 MB	92G7302 (2 x 8 MB kit)	none	92G7338	92G7310 (2 x 8 MB kit)
32 MB	92G7304 (2 x 16 MB kit)	11H0657 (2 x 16 MB kit)	92G7339	92G7312 (2 x 16 MB kit)
64 MB	92G7306 (2 x 32 MB kit)	none	none	92G7317 (2 x 32 MB kit)

Notes:

1. Memory SIMMs must be installed in matched pairs.
2. SIMM kits for the Server 310 contain two SIMMs (a matched pair).
3. EOS SIMMs cannot be mixed with Parity SIMMs.
4. ECC tin/lead SIMMs should only be used on 0E0 and 0EV models.
5. Gold and tin/lead SIMMs should not be mixed.

10.1.4 Expansion Slots

Five slots are standard on all models as shown in Table 9:

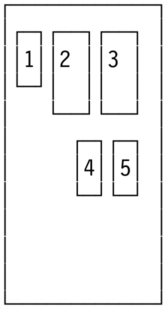
Table 9. Slots in the IBM PC Server 310.

(PCI = 32-bit PCI slot; ISA = 16-bit ISA slot)

Slot	Models 0D0, 0DT, and NDT	Models 0E0 and 0EV
1	Full-size PCI or ISA - Contains PCI SCSI-2 Fast adapter	Full-size PCI or ISA - Free
2	Full-size PCI or ISA - Free	Full-size PCI or ISA - Free
3	Full-size ISA - Free	Full-size PCI or ISA - contains PCI Ultra-Wide SCSI adapter
4	Full-size ISA - Free	Full-size ISA - Free
5	Half-size ISA	Half-size ISA - contains Ethernet adapter

Expansion Bays:

Table 10. Server 310 Expansion Bays

	Each bay has the following specifications: 1. 3.5″ slimline (1″) bay, accessible from the front of the server - in use by 1.44 MB diskette drive 2. 5.25″ half high (1.6″) bay, available, accessible from the front 3. 5.25″ half high (1.6″) bay, accessible from the front - in use by the standard quad-speed CD-ROM drive 4. 3.5″ slimline (1″) bay, not front accessible - in use by the hard disk in those models where a hard disk is standard 5. 3.5″ slimline (1″) bay, available but not accessible from the front

10.1.5 Graphics Subsystem

A PCI SVGA graphics controller is imbedded on the planar board:

- Models 0E0 and 0EV use S3 Trio64V+ while all other models use the S3 Vision868 chip set.
- 1 MB video RAM is standard and it is upgradable to 2 MB. Models 0E0 and 0EV use part #92G7336, while all other models use part #92G7443.
- When 2 MB is installed, the graphics data path is 64 bits wide. This increases graphics performance by up to 80 percent.

10.1.6 SCSI Interface

A PCI Ultra-Wide SCSI adapter is installed in 0E0 and 0EV models of the PC Server 310. For more information on this adapter, please see 17.6, “Ultra Wide SCSI PCI Adapter” on page 167.

All other models of the PC Server 310 use a PCI SCSI-2 Fast adapter. For more information on this adapter, please see 17.1, “SCSI-2 Fast PCI Adapter” on page 163.

10.1.7 SCSI-2 Fast Hard Disk

Models of the PC Server 310 are available that contain either a 1270 MB or 2160 MB hard disk. Other models are *open bay* models. That is, they do not contain hard drives as standard features.

10.1.8 Enhanced IDE Interface

All models except for 0E0 and 0EV include two EIDE connectors on the planar board for connection of up to four EIDE drives, each up to 1 GB.

Note

No IDE cable is supplied with the Server 310 although it may be ordered separately as part number #06H3610.

10.1.9 Universal Serial Bus

On Models 0E0 and 0EV, the planar board includes two Universal Serial Bus connectors. For more information on Universal Serial Bus, please see 3.3.7, “USB (Universal Serial Bus)” on page 32.

10.2 Model Summary

Table 11 shows the features of the 166 MHz Server 310s.

Table 11. Server 310 Pentium 166 MHz Models

Model	**8639-0E0**	**8639-0EV**
Processor (Int/Ext)	Pentium 166/66 MHz	
Bus Architecture	PCI/ISA	
Memory	32 MB 60 ns parity DIMMs; 160 MB Maximum	
L2 Cache	256 KB 12 ns write-back, upgradable to 512 KB	
Disk Controller	PCI Ultra-Wide SCSI adapter	
Total I/O Slots	2 ISA and 3 PCI/ISA Combo	
Available I/O Slots	2 ISA, 2 PCI/ISA Combo	1 ISA, 2 PCI/ISA Combo
Standard Hard Disk	None	2.16 GB UltraSCSI 5400 RPM
Total Bays	2 x 5.25″ half-high and 3 x 3.5″ slimline	
Available Bays	1 x 5.25″ half-high 2 x 3.5″ slimline	1 x 5.25″ half-high 1 x 3.5″ slimline
CD-ROM	4x	
Graphics	SVGA 1 MB PCI Local Bus imbedded on planar; S3 Trio64V+ chip set	
Ethernet Adapter	none	ISA 10Base-2, 10Base-T, AUI
Note: When used to describe hard drive storage, 1 GB equals 1000 million bytes.		

Table 12 shows the features of the 133 MHz Server 310s.

Table 12. Server 310 Pentium 133 MHz Models		
Model	**8639-0D0**	**8639-0DT**
Processor (Int/Ext)	P133/66 MHz	
Bus Architecture	PCI/ISA	
Memory	16 MB 70 ns parity standard (2x8 MB); 192 MB Maximum	
L2 Cache	256 KB 12 ns write-back	
Disk Controller	PCI SCSI-2 Fast	
Total I/O Slots	3 ISA and 2 PCI/ISA Combo	
Available I/O Slots	2 ISA and 1 PCI/ISA Combo	
Standard Hard Disk	None	1.27 GB 11 ms
Total Bays	2 x 5.25″ half-high and 3 x 3.5″ slimline	
Available Bays	2 x 3.5″ slimline 1 x 5.25″ half-high	1 x 3.5″ slimline 1 x 5.25″ half-high
CD-ROM	4x	
Graphics	SVGA 1 MB PCI Local Bus imbedded on planar; S3 Vision868 chip set	
Ethernet Adapter	ISA Ethernet with 10Base-2, 10Base-T, AUI connectors	
Note: When used to describe hard drive storage, 1 GB equals 1000 million bytes.		

Chapter 11. IBM PC Server 320

The PC Server 320 is positioned as a high-end workgroup or entry-level enterprise server that can serve a large number of clients for file and print applications or can act as a low-to-medium-sized database server.

A potential user of the Server 320 may have the following requirements:

- CPU capacity of up to two Pentium processors
- Disk throughput requiring SCSI-2 Fast/Wide performance and drives with high transfer rates
- Protection of investment by using existing EISA or Micro Channel adapters with the ability to migrate to advanced PCI adapters
- An upgrade path that includes ECC memory and RAID disk arrays for increased data security
- Competitively priced hardware for maximum value

The models of the PC Server 320 have the following overall variations:

- PCI/EISA models
- PCI/Micro Channel models
- Array or non-array models

Array Models

The term *array model* refers to a system with an IBM RAID controller. All IBM array models come standard with a hot-swappable backplane and hot-swap drives. The term *RAID model* may also be used interchangeably with the term array model.

11.1 Technical Description

This section describes each of the major subsystems of the IBM PC Server 320.

11.1.1 Central Electronic Complex

The central electronic complex (CEC) is made up of the following:

- Pentium processor
- Cache
- Memory
- Integrated IDE interface

11.1.1.1 Pentium Processor

There are several processors used in the PC Server 320 line. Some of the characteristics of the processors are:

- Currently either 133 MHz or 166 MHz
- SMP compatible with Multiprocessor Specification 1.1
- Additional zero insertion force (ZIF) socket available for a second processor

The following processor upgrade options are available:

- Pentium 133 MHz, part #94G5015
- Pentium 166 MHz, part #94G6053
- Pentium 200 MHz, part #94G6496

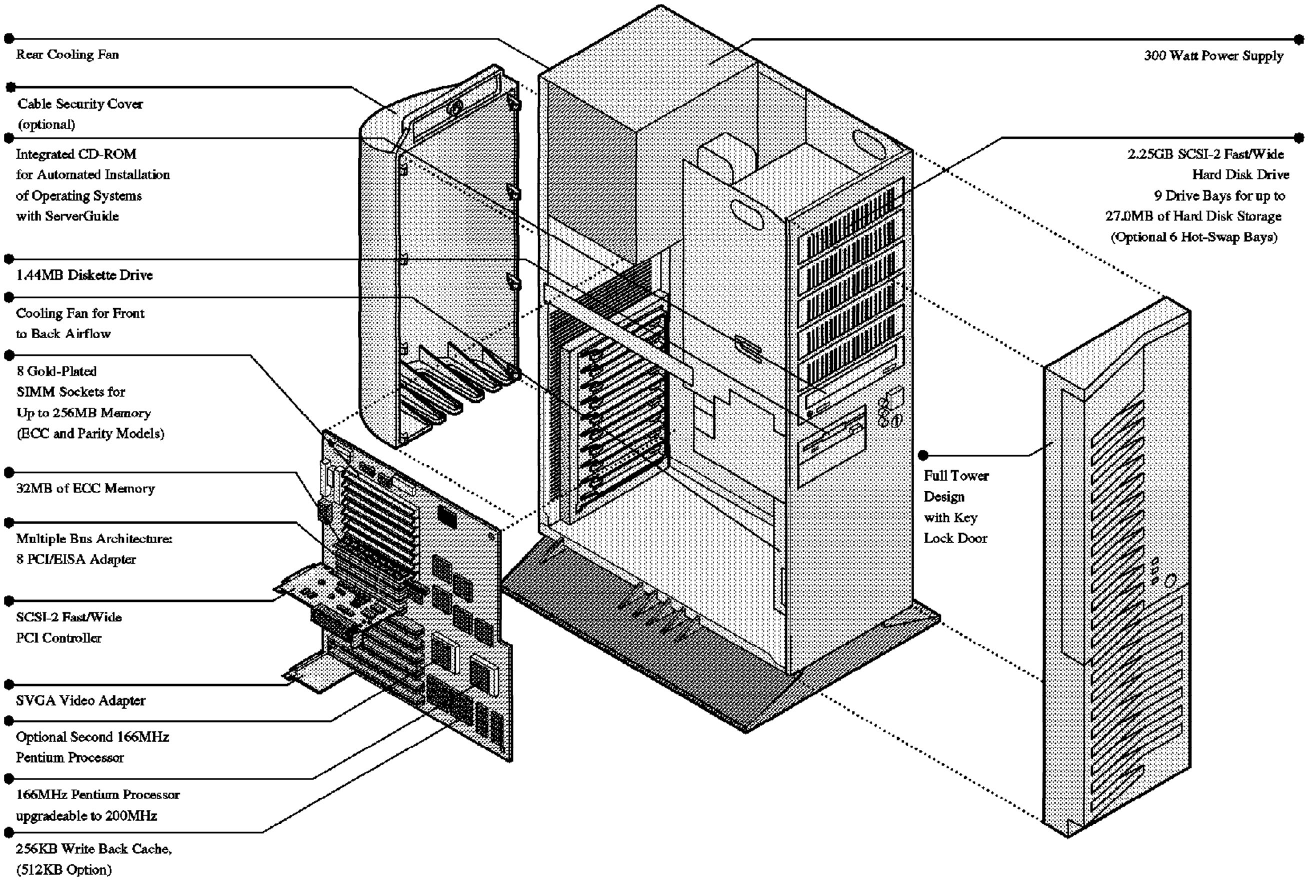


Figure 36. Covers Off Front and Side Views of the PC Server 320

Notes:

1. Pentium 133 MHz models can be upgraded to dual Pentium 133 MHz SMP machines.
2. Pentium 166 MHz models can be upgraded to dual Pentium 166 MHz SMP or to 200 MHz as single or dual Pentium SMP machines.

SMP Upgrades

When upgrading to SMP, both processors must operate at the same speed.

11.1.1.2 Cache

A 256 KB level 2 cache is standard on the PC Server 320. It has the following attributes:

- 15 ns direct-mapped
- Default is write-back; can be changed to write-through
- L2 cache is shared on SMP machines (See 4.2, "Cache in an SMP Environment" on page 37.)
- Upgradable to 512 KB using part #94G3141 on PCI/EISA models
- Upgradable to 512 KB using part #94G4462 on PCI/MCA models

Note: For performance reasons, it is recommended that the cache be upgraded to 512 KB if a second CPU is installed.

11.1.1.3 Memory

32 MB 70 ns ECC-on-SIMM is standard on all models. The memory subsystem has the following characteristics:

- Eight 72-pin, gold-plated SIMM sockets are present.
- Two sockets are used by standard 16 MB SIMMs; six are available for memory upgrades.
- A total of 256 MB is installable using eight 32 MB SIMMs.
- ECC-on-SIMM memory is standard. (Refer to 4.3.4, "ECC-on-SIMM (EOS) Memory" on page 42 for more information on this technology.)

The table below shows the IBM supported memory options for use with the Server 320. For more information concerning the usage of various memory SIMM technologies, please see 4.3.8, "General Tips on Memory Usage" on page 45.

Table 13. Server 320 Memory Options

Size	EOS SIMMs (PCI/MCA models)	EOS SIMMs (PCI/EISA models)
8 MB	None	11H0618
16 MB	None	11H0621
32 MB	11H0645 (2 x 16 MB kit)	11H0624
64 MB	11H0648 (2 x 32 MB kit)	None

Notes:

1. Memory SIMMs must be installed in matched pairs.
2. ECC-on-SIMMs cannot be mixed with Parity SIMMs.

11.1.2 Graphics Subsystem

The PCI/EISA models have an ISA SVGA graphics adapter that uses a Boca Research chip set with 1 MB of display memory. This adapter is installed in slot 8. (See Table 14 on page 114.)

The PCI/MCA models integrate the Cirrus Logic 5340 PCI SVGA chip set onto the planar board and have 1 MB of display memory standard.

11.1.3 IDE Interface

The PCI/EISA models have an enhanced IDE controller integrated onto the planar board. This provides support for up to two EIDE drives.

Note

No IDE cable is supplied with the Server 320 although it may be ordered separately as part number #06H3610.

11.1.4 SCSI-2 Controller

The SCSI-2 interface is implemented using two different controllers. The PCI SCSI-2 Fast/Wide Adapter, part #94G4673, is used for non-RAID models. Refer to 17.3, "SCSI-2 Fast/Wide PCI Adapter and Adapter II" on page 164 for more information on this adapter. The PCI SCSI-2 Fast/Wide RAID Adapter, part #94G2764, is used for RAID models. Refer to 17.5, "SCSI-2 Fast/Wide PCI-bus RAID Adapter" on page 167 for more information on this adapter.

PCI SCSI-2 Fast/Wide Adapters in the Server 320

There have been two different PCI SCSI-2 Fast/Wide adapters supplied with PC Server 320s. These are part #94G3771 and #94G4673. Part #94G3771 is known as the *PCI SCSI-2 Fast/Wide Adapter* and is now withdrawn and no longer available. Part #94G4673 is known as *SCSI-2 Fast/Wide PCI Adapter II* and all new Server 320 non-RAID PCI/EISA models are shipped with this adapter.

The different adapters should not be of concern *except* in a duplexing environment using two SCSI-2 adapters, one of the old type and one of the new type. They are incompatible in a duplexing environment as the old 94G3771 operates at 42 MHz and the new 94G4673 runs at 40 MHz.

Refer to 11.1.6, "Expansion Bays" for information on the internal SCSI cabling supplied with each of the PC Server 320 models.

11.1.5 Expansion Slots

There are eight expansion slots as shown in Table 14.

Table 14. Slots in the Server 320

Slot	PCI/MCA models	PCI/EISA models
1	Full-sized PCI	Full-sized PCI
2	Full-sized PCI	Full-sized PCI
3	Half-sized MCA	Half-sized Combo - used by the PCI SCSI-2 F/W or PCI SCSI-2 RAID
4	Half-sized MCA	Half-sized EISA
5	Full-sized MCA	Half-sized EISA
6	Full-sized MCA	Full-sized EISA
7	Full-sized MCA	Full-sized EISA
8	Full-sized MCA	Full-sized EISA - used by the graphics adapter

Note: PCI = 32-bit PCI slot; EISA = 32-bit EISA, or 16-bit ISA slot; Combo = combination PCI/EISA slot

11.1.6 Expansion Bays

The PC Server 320 has two different bay configurations as shown in Table 15 on page 115. These different configurations are explained below.

11.1.6.1 Non-Array Models

These systems have six 5.25″ half-high (1.6″) bays and three 3.5″ slimline (1″) bays.

The top drive bay of the three 3.5″ slimline bays contains the standard 1.44 MB diskette drive. The middle bay is empty and accessible from the front of the system unit. The lower bay is not front-accessible and contains the hard drive on systems equipped with a hard drive. The bottom two bays can be combined to form a half-high non-front-accessible drive bay. The IDE cable to the diskette drive is a two-drop cable allowing the connection of an IDE tape drive or second diskette drive.

The 5.25″ half-high bays are all accessible from the front of the system unit and devices such as SCSI hard drives and tape drives are suitable for these bays. The standard CD-ROM is installed in the lowest of these bays. 5.25″ to 3.5″ trays are installed in the top five bays for use when installing 3.5″ drives. Four of the 5.25″ bays can be used as 2 full-height (3.2″) bays.

Table 15. Server 320 Expansion Bays

Non Hot-Swap models		**Hot-Swap models**	
	5¼″ HH	HDD	3½″ HS SL
	5¼″ HH	HDD	3½″ HS SL
	5¼″ HH		3½″ HS SL
	5¼″ HH		3½″ HS SL
	5¼″ HH		3½″ HS SL
CD-ROM	5¼″ HH		3½″ HS SL
FDD	3½″ SL	CD-ROM	5¼″ HH
	3½″ SL	FDD	3½″ SL
HDD	3½″ SL		3½″ SL
			3½″ SL

Note: HH = half-high; HS = hot-swap; SL = slimline

Non Hot-Swap models are configured with either one 2.25 GB drive or *open bay*; that is, without any hard drives.

The top five bays can also be converted to a hot-swap configuration using the PC Server 300 Hot-Swap Bay option (part #94G2763). This gives capacity for six slimline (1″) hot-swap drives.

A 7-drop Fast/Wide cable is used to connect the standard PCI SCSI-2 Fast/Wide adapter to the CD-ROM drive and hard drive.

The CD-ROM drive is connected to the SCSI controller via a 68-to-50 pin (16-bit to 8-bit) converter, part #32G3925.

11.1.6.2 Array Models

These systems are the same as their non-array counterparts, with the following changes:

- PC Server 300 Hot Swap option installed, part #94G2763.
- PCI SCSI-2 Fast/Wide RAID Adapter, part #94G2764, is installed instead of the Fast/Wide adapter. (See 17.5, "SCSI-2 Fast/Wide PCI-bus RAID Adapter" on page 167.)

> **Note**
>
> The IBM PC ServeRAID Adapter is also available as an option for the PC Server 320 as part #70G8489. (See 17.10, "ServeRAID SCSI Adapter" on page 169.)

PC Server 320 Hot-Swap Chassis: The PC Server 320 hot-swap option (referred to here as the *hot-swap chassis*) installs in the top five drive bays of the Server 320 and provides the facility to install six slimline (1″) hot-swappable drives, or three half-high (1.6″) hot-swappable drives.

The hot-swap chassis uses a *backplane* to provide the SCSI-2 Fast/Wide interface between the adapter and the six drives. The backplane allows for removal of drives while the system is operating (using a RAID controller and certain RAID levels such as RAID-1 and RAID-5).

The hard drives used in the hot-swap models have special hot-swap trays attached to them containing the necessary electronics to provide the correct interface to the backplane. Drives can be ordered with the hot-swap trays pre-installed, or the trays can be ordered separately. The latter method would be used if upgrading a non-hot-swap model to a hot-swap model. The following hot-swap options are available for the Server 320:

- 1.12 GB SCSI-2 Fast/Wide Hot-Swap drive, part #94G2649
- 2.25 GB SCSI-2 Fast/Wide Hot-Swap drive, part #94G2650
- 4.51 GB SCSI-2 Fast/Wide Hot-swap drive, part #94G2651
- SCSI-2 Hot-Swap tray for Fast/Wide drive, part #70G9741
- SCSI-2 Hot-Swap tray for Fast drive, part #70G9851

The hot-swap chassis is generally used in conjunction with a RAID controller to provide data-redundancy as well as provide the ability to replace drives without powering off the server. It is, however, possible to use the hot-swap chassis *without* the use of the RAID functionality. In fact, some older models of the PC Server 320 were equipped this way.

Hot-swapping of hard disk drives in non-array PC Server 320 models is currently supported without system reboot when using NetWare 3.12 and 4.1 and Windows NT 3.51 network operating systems.

A 2-drop SCSI-2 Fast/Wide internal cable is provided standard with all hot-swap models of the PC Server 320. One connector is for the CD-ROM drive and the other is to connect to the backplane of the hot-swap chassis.

Array models are configured with either two 2.25 GB drives or an *open bay*, that is, without any hard drives.

Note: When the hot-swap chassis is installed, the top five 5.25″ bays are converted to hot-swap bays. With the standard CD-ROM, this means there are no spare 5.25″ bays for devices such as internal tape drives. If a tape drive is required, you can connect an external tape drive via the external SCSI port.

In addition to the hot-swap bays, the array models also have three additional 3.5″ bays in the lower section of the machine. Please see the discussion on non-array models in 11.1.6.1, “Non-Array Models” on page 115 for information on these bays.

11.1.6.3 CD-ROM

A quad-speed SCSI drive is installed in all models of the PC Server 320.

11.2 Model Summary

The tables below provide a summary of all current IBM PC Server 320 models.

Note: IBM does not market all models in all countries.

11.2.1 PC Server 320 166 MHz Models

The following table shows the 166 MHz models of the Server 320.

Table 16 (Page 1 of 2). Server 320 PCI/EISA Pentium 166 MHz Models

Model	**8640-EE0**	**8640-EEV**	**8640-EE1**	**8640-EES**
Processor (Int/Ext)	Pentium 166/66 MHz, upgradable to 200/66 MHz			
SMP Capable	Dual SMP Pentium 166 or 200 MHz			
Bus Architecture	PCI/EISA			
Memory	32 MB 70 ns ECC-on-SIMM (2 x 16 MB); 256 MB maximum			
L2 Cache	256 KB write-back; expandable to 512 KB			
Disk Controller	PCI SCSI-2 F/W	PCI SCSI-2 F/W	PCI SCSI-2 F/W RAID Controller	PCI SCSI-2 F/W Raid Controller
Total I/O Slots	2 PCI, 5 EISA, and 1 Shared			
Available I/O Slots	1 PCI, 4 EISA, and 1 Shared			
Standard Hard Disk	None	2.25 GB 8 ms	None	2 x 2.25 GB 8 ms
Hot-swap chassis	Optional	Optional	Standard	Standard
Total Bays	6 x 5¼″ half-high 3 x 3½″ slimline	6 x 5¼″ half-high 3 x 3½″ slimline	1 x 5¼″ half-high 3 x 3½″ slimline 6 x 3½″ hot-swap	1 x 5¼″ half-high 3 x 3½″ slimline 6 x 3½″ hot-swap
Available Bays	5 x 5¼″ half-high 2 x 3½″ slimline	5 x 5¼″ half-high 1 x 3½″ slimline	0 x 5¼″ half-high 2 x 3½″ slimline 6 x 3½″ hot-swap	0 x 5¼″ half-high 2 x 3½″ slimline 4 x 3½″ hot-swap
CD-ROM	4x			

Table 16 (Page 2 of 2). Server 320 PCI/EISA Pentium 166 MHz Models				
Model	**8640-EE0**	**8640-EEV**	**8640-EE1**	**8640-EES**
Graphics	ISA adapter with SVGA and 1 MB RAM			

Note:

1. When used to describe hard drive storage, 1 GB equals 1000 million bytes.
2. When upgrading to SMP, both processors must operate at the same speed.
3. 512 KB of L2 cache is recommended in SMP configurations.

11.2.2 PC Server 320 133 MHz Models

The following table shows the 133 MHz models of the Server 320.

Table 17. Server 320 PCI/EISA and PCI/MCA Pentium 133 MHz Models

Model	**8640-2D0**	**8640-3D0**	**8640-MD2**
Processor (Int/Ext)	Pentium 133/66 MHz		
SMP Capable	Dual SMP Pentium 133 MHz		
Bus Architecture	PCI/EISA	PCI/EISA	PCI/MCA
Memory	32 MB 70 ns ECC-on-SIMM (2x16 MB); 256 MB maximum		
L2 Cache	256 KB write-back; expandable to 512 KB		
Disk Controller	PCI SCSI-2 F/W	PCI SCSI-2 F/W	Integrated PCI SCSI-2 F/W
Total I/O Slots	2 PCI 5 EISA 1 Shared	2 PCI 5 EISA 1 Shared	2 PCI 6 MCA
Available I/O Slots	1 PCI 4 EISA 1 Shared	1 PCI 4 EISA 1 Shared	2 PCI 6 MCA
Standard Hard Disk	None		
Hot-swap chassis	Optional	Standard	Standard
Total Bays	6 x 5¼″ half-high 3 x 3½″ slimline	1 x 5¼″ half-high 3 x 3½″ slimline 6 x 3½″ hot-swap	1 x 5¼″ half-high 2 x 3½″ slimline 6 x 3½″ hot-swap
Available Bays	5 x 5¼″ half-high 2 x 3½″ slimline	2 x 3½″ slimline 6 x 3½″ hot-swap	1 x 3½″ slimline 6 x 3½″ hot-swap
CD-ROM	4x		
Graphics	ISA adapter with SVGA and 1 MB RAM		

Note:

1. When used to describe hard drive storage, 1 GB equals 1000 million bytes.
2. When upgrading to SMP, both processors must operate at the same speed.
3. 512 KB of L2 cache is recommended in SMP configurations.

11.2.3 Additional Server 320 Models

In addition to the PC Server 320 models previously discussed, there are also three additional models:

Model 8640-IS2: This system is pre-loaded with IBM Internet Connection Secure Server for OS/2, Internet Connection Secure WebExplorer for OS/2, and OS/2 Warp Server.

Model 8640-NTN: This system is pre-loaded with Netscape Commerce Server for Windows NT, Netscape Navigator, and Microsoft Windows NT Server Edition.

Model 8640-SUN: This system is pre-loaded with Netscape Commerce Server for Solaris, Netscape Navigator, Sunsoft Solaris Server x86 Edition, Internet Gateway for Solaris, and Solaris System Administrator′s AnswerBook.

All models are equipped with 133 MHz processors, 32 MB memory, a 2.25 GB UltraSCSI hard disk (up to 7200 RPM), and a quad-speed CD-ROM drive.

Chapter 12. IBM PC Server 325

The PC Server 325 is recommended for customers in large-to-medium-sized businesses with interconnected departmental files and/or databases or Internet servers. In addition, these systems are ideal for branch office file/print, e-mail, or database servers. The PC Server 325 Mini Tower model is shown in Figure 37.

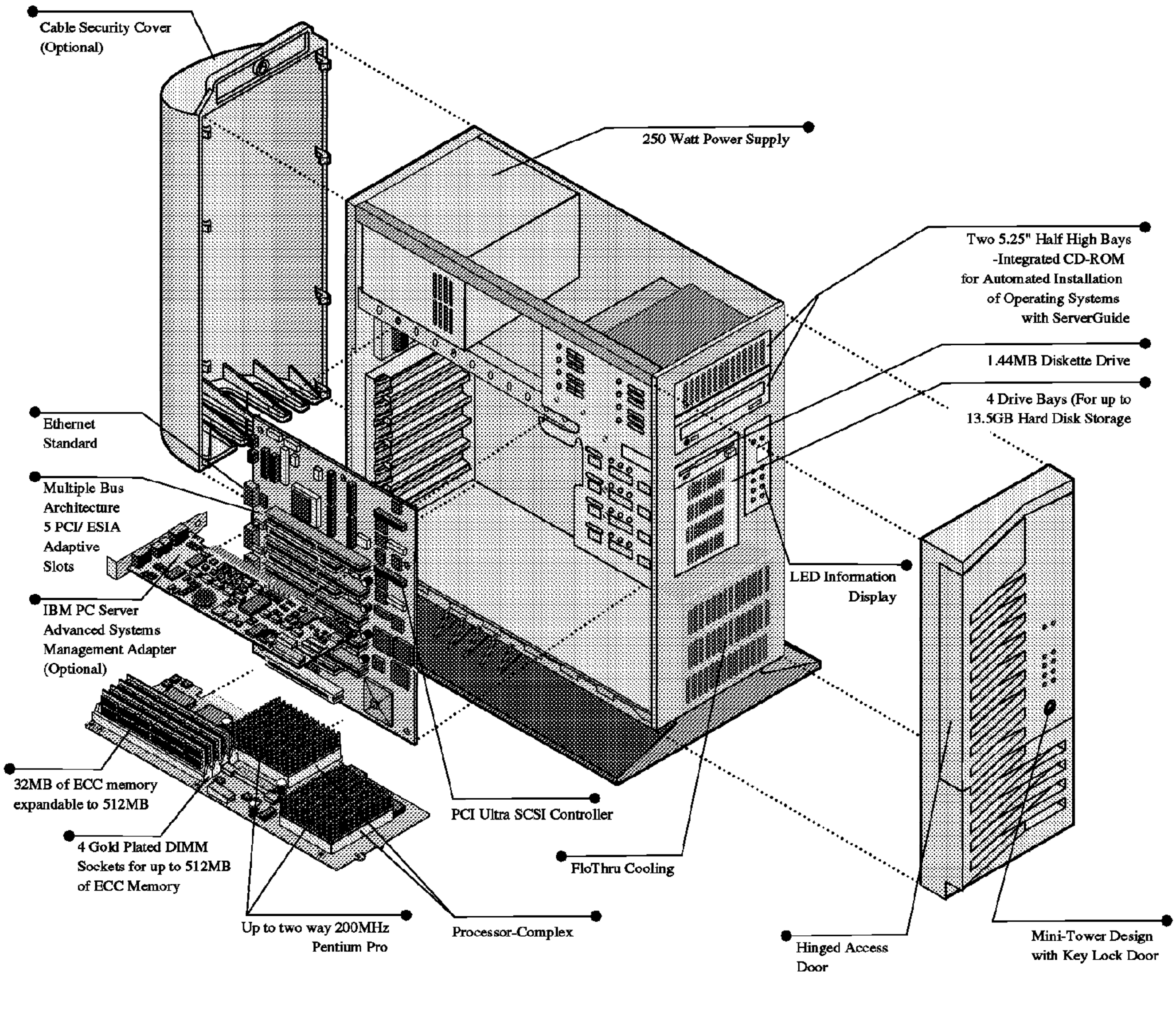


Figure 37. Exploded View of the IBM PC Server 325 Mini Tower Model

A potential user of the Server 325 may have the following requirements:

- Pentium Pro technology in an upgradable processor complex
- SMP capability
- EDO ECC memory for data security
- Medium disk storage or external hot-swap requirements

- Disk throughput requiring UltraSCSI performance and drives with high transfer rates
- Five available PCI slots
- Mini tower or 19″ rack drawer model

12.1 Technical Description

The PC Server 325 has the following features:

- 180 or 200 MHz Pentium Pro (256 KB L2 cache)
- 32 MB EDO ECC 60 ns DIMMs (up to 1 GB)
- Six slots available (1 dedicated PCI only, 4 PCI/EISA Combo, 1 dedicated EISA only)
- Two 132 MBps PCI buses with secondary EISA bus
- 8x CD-ROM
- Ultra-Wide SCSI controller on planar board
- Full duplex PC Ethernet controller on planar board (10Base--T port using RJ-45 or an AUI port)
- IR port supporting 4 Mbps and 1.15 Mbps
- Two high-speed serial/asynchronous ports (NS16550 compatible)
- One high-speed 2 MBps bidirectional parallel port supporting devices using ECP, EPP or SPP
- 10 status LEDs for critical function monitoring (Ethernet, power, SCSI drive, primary and secondary processor, and security activities)
- Keyboard
- Mouse (except for Rack Model RS0)
- ServerGuide
- TME 10 NetFinity

The PC Server 325 comes in both a mini tower and a 19″ rack drawer model. It is also possible to convert a mini tower system into a rack drawer model by ordering the PC Server 325 Rack Upgrade Option. (See 12.1.11, “Related Options” on page 129 for more information on options.)

12.1.1 Central Electronic Complex

The central electronic complex (CEC) is made up of the following:

- Processor subsystem
- Memory
- System board

12.1.1.1 Processor Subsystem

The Pentium Pro processor is part of a *processor complex*. The processor complex is a special adapter that plugs into a dedicated slot on the planar board and provides the following functions:

- Central processor - Pentium Pro 180 or 200 MHz
- ECC memory controller
- DMA controller

The PC Server 325s are equipped with a 180 or 200 MHz Pentium Pro with 256 KB L2 cache. A second identical Pentium Pro processor may be added to take advantage of the SMP capability. The 180 and 200 MHz Pentium Pro processors operate at 180 and 200 MHz for internal operations and handles external operations from the bus and system memory at 60 and 66 MHz respectively.

The following processor upgrades are available:

- PC Server SMP 166 MHz/512 KB Processor Upgrade, part #94G4908
- PC Server SMP 180 MHz/256 KB Processor Upgrade, part #94G7082
- PC Server SMP 200 MHz/256 KB Processor Upgrade, part #94G6175
- PC Server SMP 200 MHz/512 KB Processor Upgrade, part #94G6463

When a second processor is installed, the second processor becomes the primary processor.

These processor option packages contain a heat sink, clip, grease, a voltage reduction module (VRM), and publications. The VRM module is necessary to step down the 3.3V power supply to the 2.9V required by the Pentium Pro. An extra fan (also supplied) attaches to the top of the heat sink of the processor. If this fan is not used the processor will overheat.

SMP Upgrades

When upgrading to SMP, both processors must operate at the same speed and have the same cache size.

Figure 38 shows the processor complex in the PC server 325.

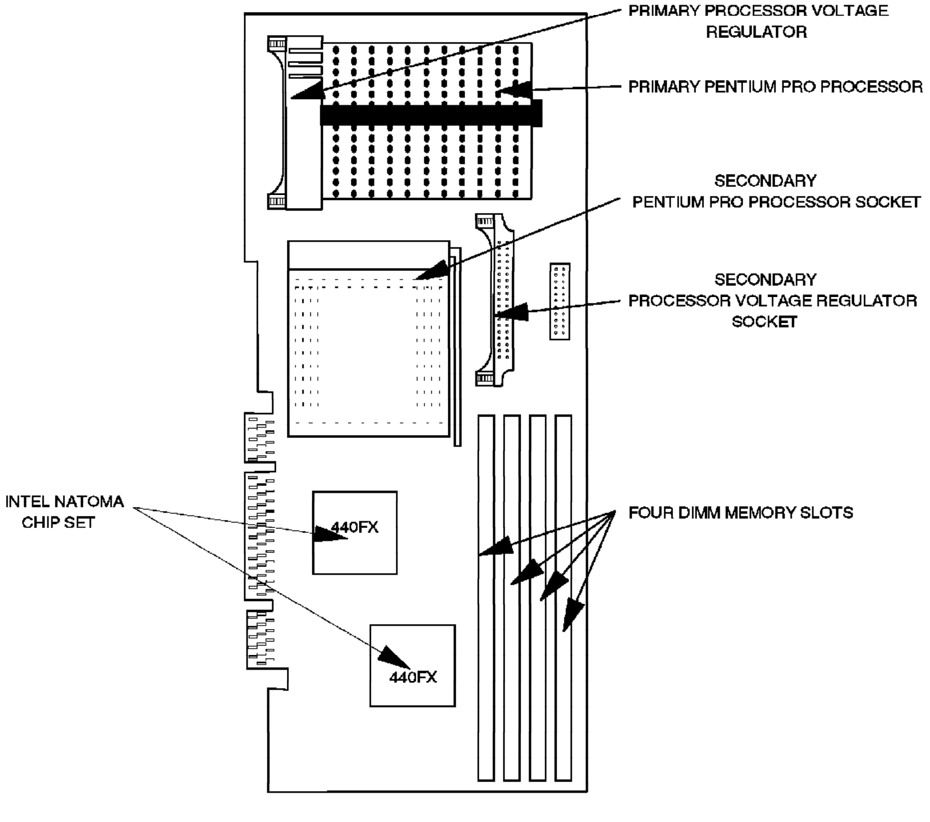


Figure 38. IBM PC Server 325 Processor Board

12.1.1.2 Memory

32 MB of 60 ns EDO ECC memory is standard on all Server 325 models. ECC (*error correcting code* or *error checking and correcting*) memory can detect and correct single-bit errors, detect double-bit errors and detect some triple-bit errors. Refer to Chapter 4, “Memory and Cache Technology” on page 33 for more information on ECC and EDO.

One 32 MB EDO ECC DIMM is installed as standard on all models. With current memory technology, a total of 1 GB of EDO ECC memory is installable using four 256 MB 60 ns DIMMs.

The following memory options are available:

- PC Server 32 MB DIMM, Part #94G6473
- PC Server 64 MB DIMM, Part #94G6474
- PC Server 128 MB DIMM, Part #94G6475
- PC Server 256 MB DIMM, Part #94G7079

All supported memory DIMMs can be installed in any combination. Figure 38 on page 123 shows the location of the DIMM slots.

12.1.2 Bus Architecture

The PC Server 325 contains three bus architectures:

- PCI - Peripheral Component Interconnect which supports data transfer rates of up to 132 MBps and runs at 33 MHz clock
- EISA - Extended Industry Standard Architecture supports data transfer rates of up to 33 MBps and runs at 8.33 MHz clock speeds
- ISA - Industry Standard Architecture which supports data transfer rates of up to 8.33 MBps and runs at 8.33 MHz clock speeds

3.3, “Bus Architectures” on page 24 contains detailed information on each of these bus architectures.

Figure 39 on page 125 shows the architecture of the PC Server 325.

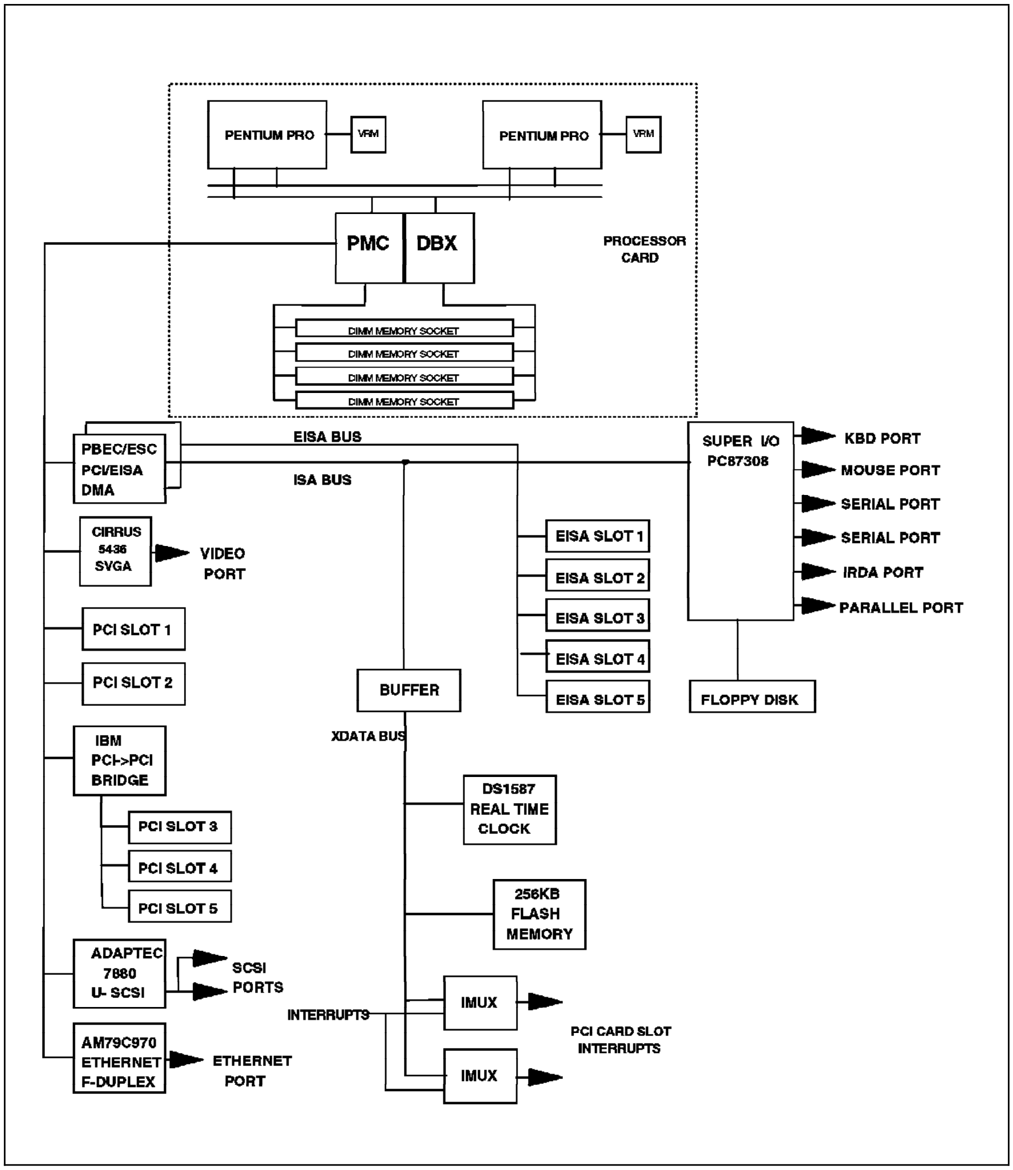


Figure 39. IBM PC Server 325 System Architecture

12.1.3 UltraSCSI Controller

The PC Server 325 has an UltraSCSI controller integrated on the planar board. (See 17.6, "Ultra Wide SCSI PCI Adapter" on page 167 for more information on UltraSCSI.) This controller is an Adaptec 7880 UltraSCSI controller, which provides throughput of up to 40 MBps. It supports up to 15 devices and has both an internal and external connector.

The UltraSCSI controller supports the high performance 7200 RPM SCSI-2 F/W drives (1.12 GB, 2.25 GB and 4.51 GB) or the economical 5400 RPM hard disk drives.

Refer to 12.1.7, "Tower Model Cabling" on page 127 for information on the internal SCSI cabling supplied with each of the PC Server 325 models.

12.1.4 Expansion Slots

Six slots are standard on all models of PC Server 325. As shown in Figure 40, four of these are combination PCI/EISA slots and can accommodate either type of board. The very top slot is a dedicated EISA slot, while the very lowest is a dedicated PCI slot.

Full-length cards are supported in all slots except the top dedicated EISA slot and the one beneath it, which is a PCI/EISA combination slot.

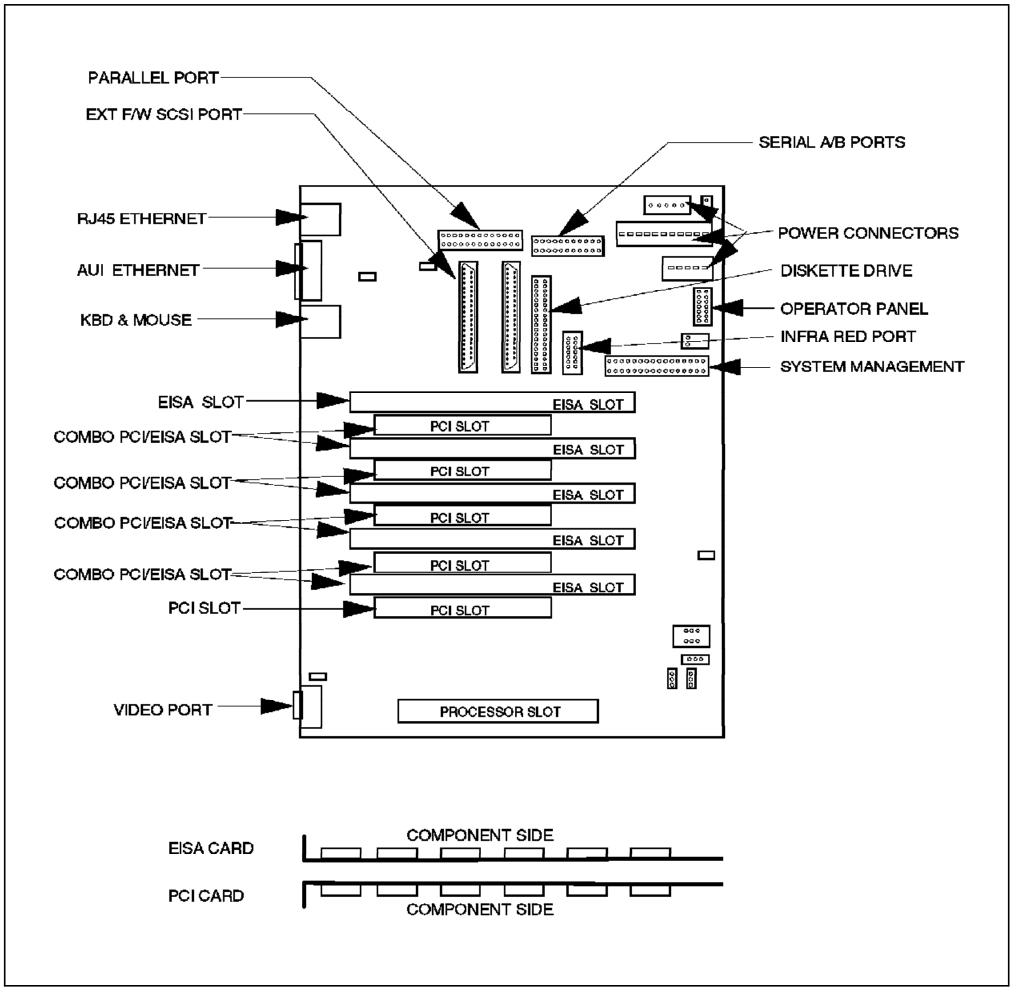


Figure 40. IBM PC Server 325 System Board

12.1.5 Graphics Subsystem

The Server 325 uses the Cirrus Logic 5436 PCI SVGA chip set integrated on the planar board and comes standard with 1 MB of VRAM. Some of the characteristics of the graphics subsystem are:

- Integrated RAMDAC with 262,144 color palette
- Support for up to 24-bit per pixel color
- True color VGA graphical user interface accelerator
- Bit block transfer
- 16-bit interface

- Compatible with VGA color, analog monochrome monitors, and multiple frequency monitors

12.1.6 Expansion Bays

The Server 325 has a total of seven front-accessible bays:

- Two 5.25″ half-high bays (one with CD-ROM installed)
- Five 3.5″ slimline bays (one with diskette drive installed and another bay is used for a 2.25 GB SCSI-2 F/W 7200 RPM F/W hard drive in models where a hard disk is standard)

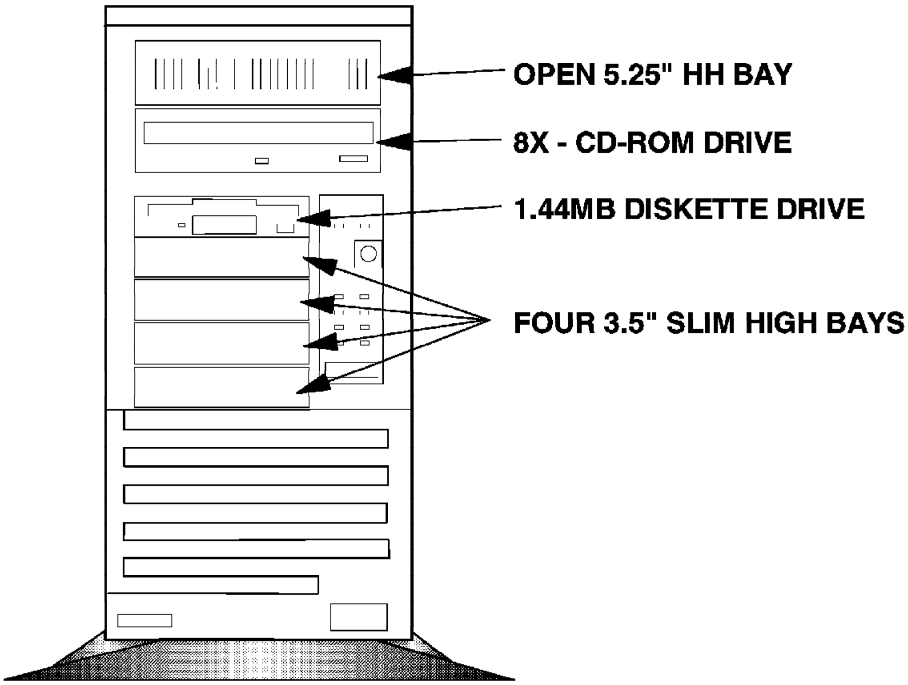


Figure 41. IBM PC Server 325 Expansion Bays

12.1.7 Tower Model Cabling

The tower models of the PC Server 325 are cabled internally with a 7-drop 16-bit wide SCSI cable. One end of this cable is attached to the internal 16-bit connector of the 7880 UltraSCSI F/W controller. On open bay models, the last drop of the 7-drop cable is attached to the standard CD-ROM through a 68-to-50-pin converter. When installing an SCSI-2 F/W hard drive, the CD-ROM should be disconnected from the last drop and reconnected to the next to the last drop. The hard drive should be installed in the lowest 3.5″ bay and connected to the last drop on the SCSI cable to allow proper termination for the SCSI bus. In the standard drive model, the 2.25 GB SCSI-2 F/W hard disk drive and CD-ROM are pre-configured in this same manner. If connecting narrow devices to this cable, additional 68-to-50-pin SCSI converters (part #32G3925) must be ordered.

Hot-Swap Capability

The PC Server 325 does not support hot-swap devices internal to the system. The PC ServeRAID Adapter may be ordered to add an external hot-swap and RAID capability. (See Chapter 18, “Server Enclosures” on page 177 for more information on external enclosures.)

12.1.8 CD-ROM

An 8x SCSI CD-ROM drive is installed in all models of the PC Server 325. It has the following characteristics:

- Average access time of 150 ms
- Average data transfer rate of 1200 KBps
- Burst rate of 5.0 MBps
- Buffer size of 256 KB

12.1.9 Operator Panel

The following figure describes the LEDs on the PC Server 325 Operator Panel:

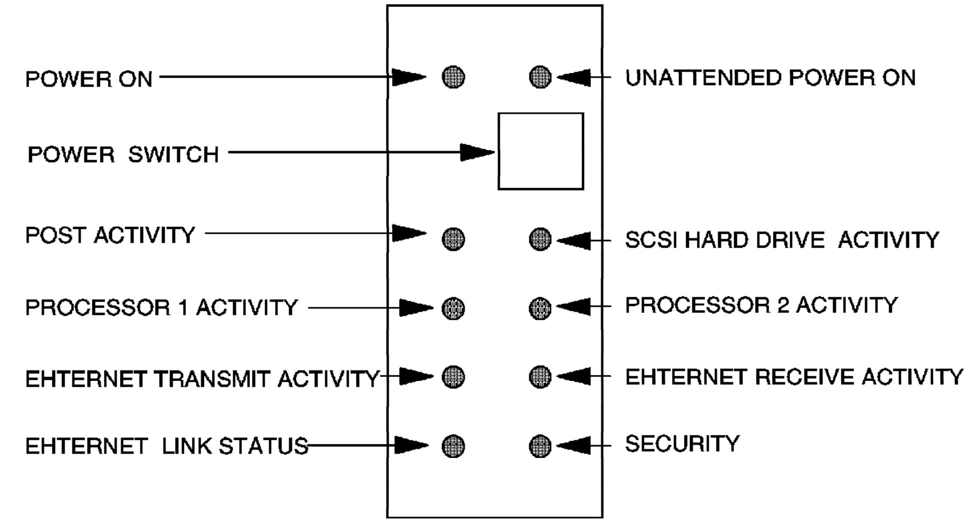


Figure 42. IBM PC Server 325 Operator Panel

12.1.10 Security

The security and auditability features of the PC Server 325 models are:

- *Power on* and *privileged access password* controls who has access to the data and server setup program on the server.
- *Set keyboard password* is used to lock the keyboard without turning off the system.

- A *mechanical lock* allows the user to lock the system′s front cover door to prevent unauthorized personnel access to diskette drives or removable media.
- *Selectable boot sequence* prevents unauthorized installation of software or removal of data from the diskette drive.
- *Tie-down* capability is achievable by attaching a common U bolt attached to the back frame.
- Operation without a keyboard and display are functions that are supported once the server is configured with the appropriate network operating system. This minimizes the risk of unauthorized persons tampering with the system software and configuration.
- A PC Server 325 Security Cover is an option available to cover retaining screws and to prevent unauthorized access to the server′s external ports.
- A PC Server 325 Security Cable Kit can detect tampering or removal of the server′s covers. This option, when enabled, can also alert the system administrator once this detection occurs.

Refer to 12.1.11, "Related Options" for information on security option part numbers.

Note: The electronic eject 2.88 MB diskette drive is not supported by the PC Server 325 models.

12.1.11 Related Options

The following related options are announced for PC Server 325:

- 94G5695 PC Server 325 Rack Upgrade
- 94G6398 PC Server 325 Security Cable Kit
- 94G6471 PC Server 325 Security Cover
- 94G5570 Advanced System Management Adapter

The PC Server 325 Rack Upgrade Option provides the necessary hardware and instructions to convert a PC Server 325 mini tower model for installation into a rack enclosure. This option contains rack mount slides, a modified cover that opens from the top, a new bezel and door, and a cable management arm.

The PC Server 325 Security Cable Kit and PC Server 325 Security Cover are designed to provide a higher level of security by detecting unauthorized removal of, or tampering with, server covers even with the power off. They also restrict unauthorized access to external ports. These two options are used only for the mini tower model.

The 325 Security Cable kit requires the Advanced System Management card, part #94G5570, to enable the privileged access password.

12.1.12 Tower Model versus Rack Model

The PC Server 325 rack model has rack slides, a modified cover which opens from the top, a specially designed bezel and door, and a cable management arm. It is designed to fit into an industry-standard 19″ rack enclosure. A rack option is available for upgrading tower models to rack models (See 12.1.11, "Related Options" for information on converting a tower model into a rack model.) The rack model does not ship with a mouse or keyboard.

PC Server 325 tower models are intended for use as floor-standing systems and are tested and designed to operate in a vertical position. The PC Server 325 rack model and tower models which have been upgraded the with PC Server 325 Rack Upgrade option are designed and tested to operate in a horizontal position.

12.2 Model Summary

Four models of the Server 325 are available. These are shown in Table 18.

Table 18. Server 325 Models

Model	8639-EJ0	8639-ES0	8639-ESV	8639-RS0
Processor (Speed in MHz/ Cache in KB)	Pentium Pro 180/256	Pentium Pro 200/256	Pentium Pro 200/256	Pentium Pro 200/256
SMP Capable	Yes			
Bus Architecture	PCI/EISA			
Memory	32 MB 60 ns EDO ECC standard as 1 DIMM; 1 GB EDO ECC maximum			
Disk Controller	Integrated UltraSCSI F/W controller on planar board (supports 40 MBps and up to 15 devices)			
Total I/O Slots	1 PCI, 4 shared, and 1 EISA			
Available I/O Slots	1 PCI, 4 shared, and 1 EISA			
Standard Hard Disk	None	None	1 x 2.25 GB 7200 RPM	None
Hot-Swap Chassis	None; use external storage enclosure if required			
Total Bays	2 x 5.25″ half-high and 5 x 3.5″ slimline			
Available Bays	1 x 5.25″ half-high 4 x 3.5″ slimline	1 x 5.25″ half-high 4 x 3.5″ slimline	1 x 5.25″ half-high 3 x 3.5″ slimline	1 x 5.25″ half-high 4 x 3.5″ slimline
LAN Port	Integrated full-duplex 10 Mbps Ethernet on planar board			
CD-ROM	8x			
Graphics	SVGA with 1 MB RAM on planar board; Cirrus Logic 5436 chip set			
Rack model	No, upgradable	No, upgradable	No, upgradable	Yes
Mouse and Keyboard Standard	Yes	Yes	Yes	No

Note:

1. When used to describe hard drive storage, 1 GB equals 1000 million bytes.
2. When upgrading to SMP, both processors must operate at the same speed.

Chapter 13. IBM PC Server 330

The PC Server 330 is positioned as a high-end workgroup or entry-level enterprise server that can serve a large number of clients for file and print applications or can act as a small-to-medium-sized database server. The PC Server 330 is shown in Figure 43.

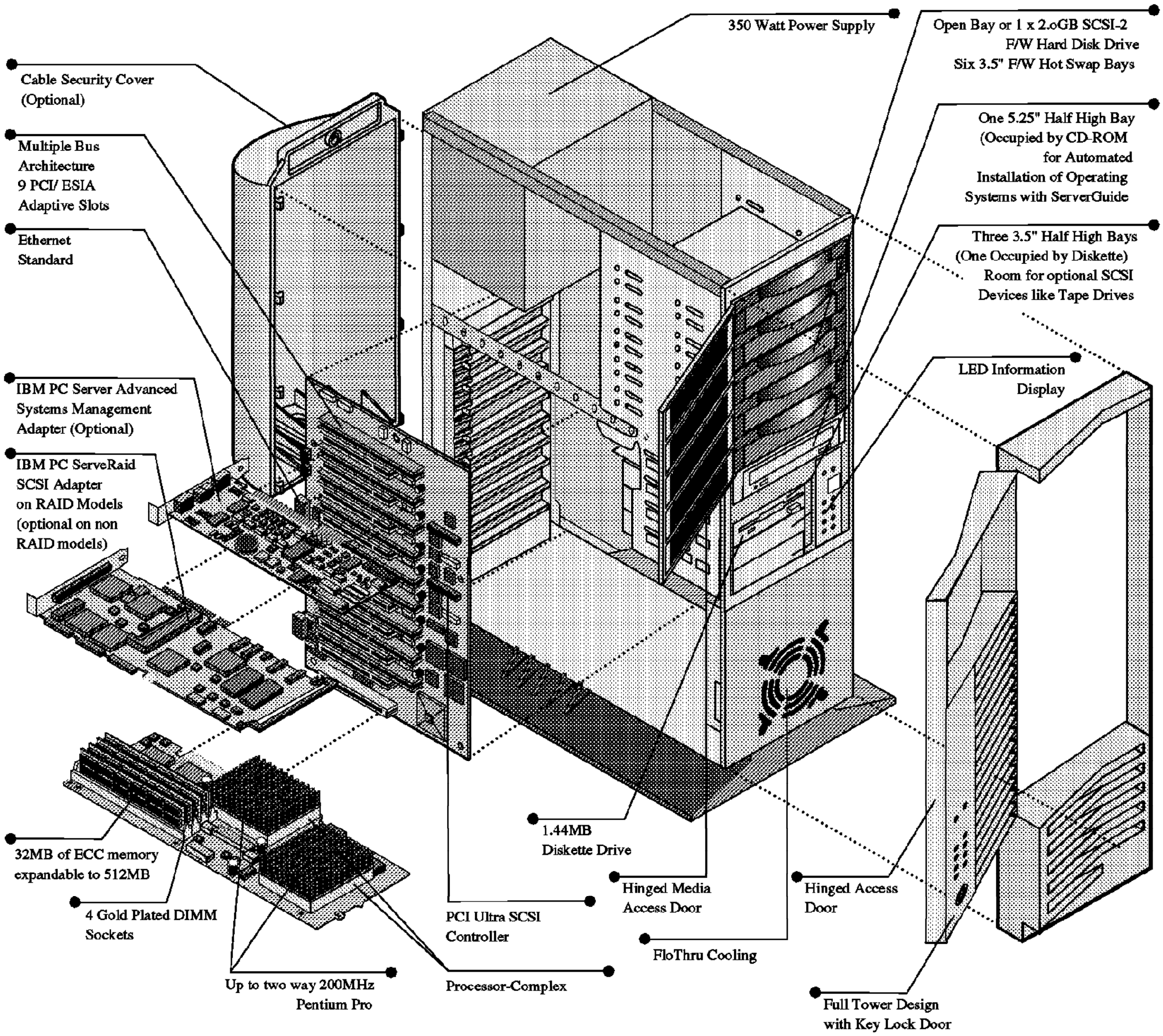


Figure 43. Exploded View of the IBM PC Server 330

A potential user of the Server 330 may have the following requirements:

- Pentium Pro technology in an upgradable processor complex
- SMP capability

- ECC memory, RAID and disk hot-swap capabilities for data security and high availability
- Large disk storage
- Disk throughput requiring UltraSCSI performance and drives with high transfer rates
- Eight to nine available I/O slots

13.1 Technical Description

The IBM PC Server 330 is a powerful full-tower system with a Pentium Pro 200 MHz processor with SMP capability. An integrated I2C bus provides Vital Product Data (VPD) and a connector for an optional IBM PC Advanced Systems Management Card for monitoring temperature and other server functions. The IBM PC Server 330 has additional LED lights located on the front panel that provide visual information about SCSI, Ethernet, power, POST, primary and secondary processor, and security activities.

13.1.1 Central Electronic Complex

The central electronic complex (CEC) is made up of the following:

- Processor subsystem
- Memory
- System board

13.1.1.1 Processor Subsystem

The Pentium Pro 200 MHz processor is part of a *processor complex*. The processor complex is a special adapter that inserts into the planar board and provides the following functions:

- Central processor - Pentium Pro
- ECC memory controller
- DMA controller

The processor complex used in the PC Server 330 is the same as that used in the PC Server 325 with the exception of the CPU itself. The two machines have different CPU options available.

The following processor upgrades are available for the PC Server 330:

- 166 MHz processor (512 KB cache) Part #94G4908
- 200 MHz processor (256 KB cache) Part #94G6175
- 200 MHZ processor (512 KB cache) Part #94G6463

When a second processor is installed, the second processor becomes the primary processor.

These processor option packages contain a heat sink, clip, grease, a voltage reduction module (VRM), and publications. The VRM module is necessary to step down the 3.3V power supply to the 2.9V required by the Pentium Pro. An extra fan (also supplied) attaches to the top of the heat sink of the processor. If this fan is not used the processor will overheat.

SMP Upgrades

When upgrading to SMP, both processors must operate at the same speed and have the same cache size.

Please see 12.1.1.1, “Processor Subsystem” on page 122 for more information about the processor complex used in the PC Server 325s and 330s.

13.1.1.2 Memory

32 MB of 60 ns EDO memory is standard on all Server 330 models and installed as one 32 MB DIMM. Since the PC Server 330 uses the same processor complex as the 325, all of the memory options for the 330 are the same as for the 325. Please see 12.1.1.2, “Memory” on page 124 for more information on the memory options for the Server 330.

13.1.2 Bus Architecture

The PC Server 330 contains the same bus architectures as the IBM PC Server 325. Refer to 12.1.2, “Bus Architecture” on page 124 for more information.

13.1.3 SCSI Controller

There are two varieties of SCSI controller used in the Server 330.

1. Non-RAID array models: PCI UltraSCSI integrated on the planar board. The integrated SCSI controller is on all Server 330 models. (Refer to 17.6, “Ultra Wide SCSI PCI Adapter” on page 167 for more information on UltraSCSI.)
2. RAID array models: PCI ServeRAID SCSI Adapter, part #70G8489. (See 17.10, “ServeRAID SCSI Adapter” on page 169 for more information on this adapter.)

Refer to 13.1.6, “Expansion Bays” for information on the internal SCSI cabling supplied with each of the PC Server 330 models.

13.1.4 Expansion Slots

Nine slots are standard on all models as shown in Figure 44 on page 134.

13.1.5 Graphics Subsystem

The Server 330 uses the Cirrus Logic 5436 PCI SVGA chip set integrated on the planar board and comes standard with 1 MB of VRAM. Some of the characteristics of the graphics subsystem are:

- Integrated RAMDAC with 262,144 color palette
- Support for up to 24-bit per pixel color
- True color VGA graphical user interface accelerator
- Bit block transfer
- 16-bit interface
- Compatible with VGA color, analog monochrome monitors, and multiple frequency monitors

13.1.6 Expansion Bays

The Server 330 has a total of 9 bays in non-hot-swap models and 10 bays in hot-swap models, of which 6 are hot-swappable.

The 330 uses the PC Server HS Backplane III (part #76H2670). This new backplane supersedes the PC Server HS Backplane III (part #70G9855). A single drop SCSI cable connects the hot-swap backplane to the ServeRAID SCSI adapter.

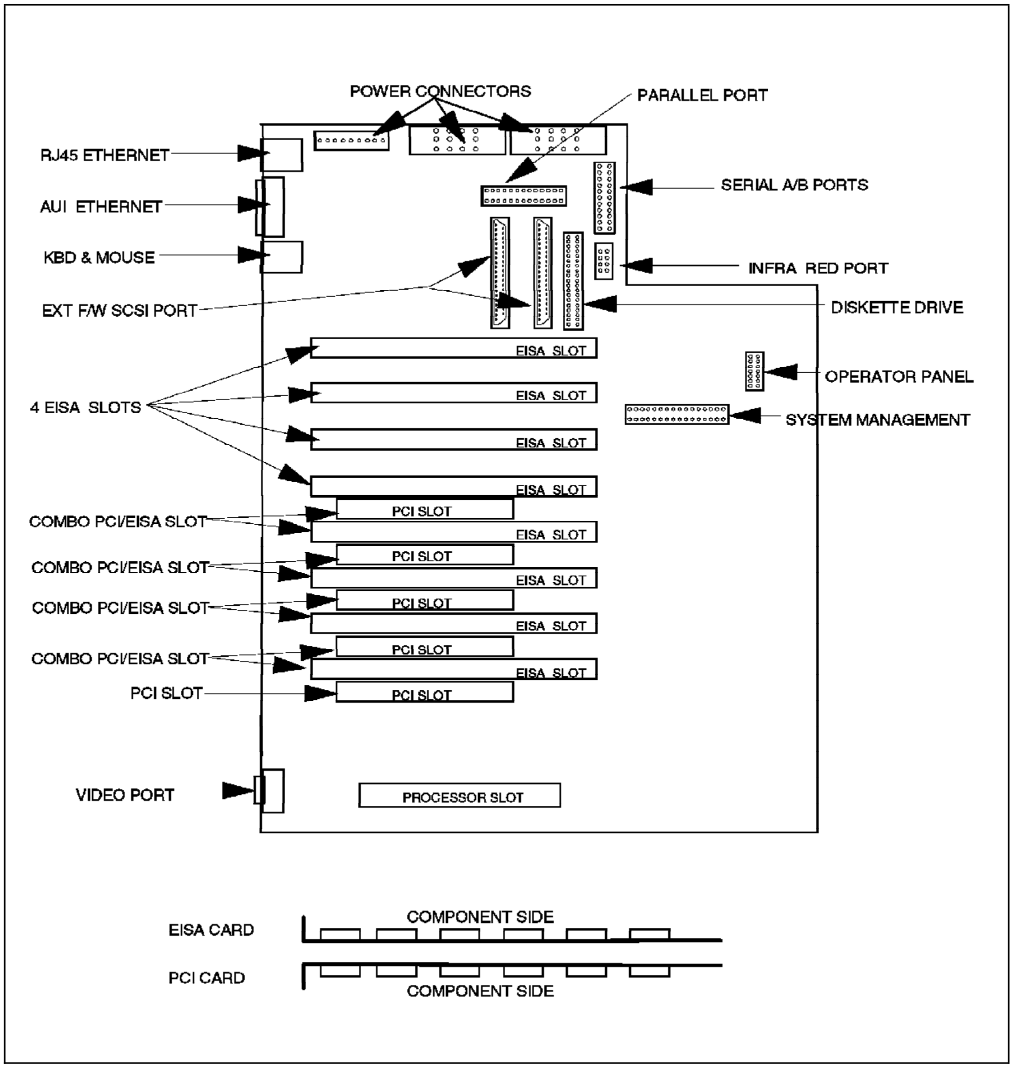


Figure 44. View of the IBM PC Server 330 System Board

Table 19. Server 330 Expansion Bays. (HH = half-high; HS = hot-swap; SL = slimline)

Non-hot-swap models		Hot-swap models	
	5¼″ HH	HDD	3½″ HS SL
	5¼″ HH	HDD	3½″ HS SL
	5¼″ HH		3½″ HS SL
	5¼″ HH		3½″ HS SL
	5¼″ HH		3½″ HS SL
CD-ROM	5¼″ HH		3½″ HS SL
FDD	3½″ SL	CD-ROM	5¼″ HH
	3½″ SL	FDD	3½″ SL
HDD	3½″ SL		3½″ SL
			3½″ SL

Hot-Swap and Non-RAID Array Systems

Hot-swapping of hard disk drives in non-array PC Server models is currently supported without system reboot when using NetWare 3.12 and 4.1 and Windows NT 3.51 network operating systems.

In other operating environments, if you aren't using a RAID controller, then it is best to shut down the server before replacing a drive. Even though the hardware would support a hot-swap, the operating system would need the capability to recognize that a device has been added or replaced.

13.1.6.1 CD-ROM

An 8x SCSI CD-ROM drive is installed in all models of the PC Server 330. It has the following characteristics:

- Average access time of 150 ms
- Average data transfer rate of 1200 KBps
- Burst rate of 5.0 MBps
- Buffer size of 256 KB

13.1.6.2 PC Server 330 Operator Panel

Figure 15 describes the functions of the PC Server 330 Operator Panel.

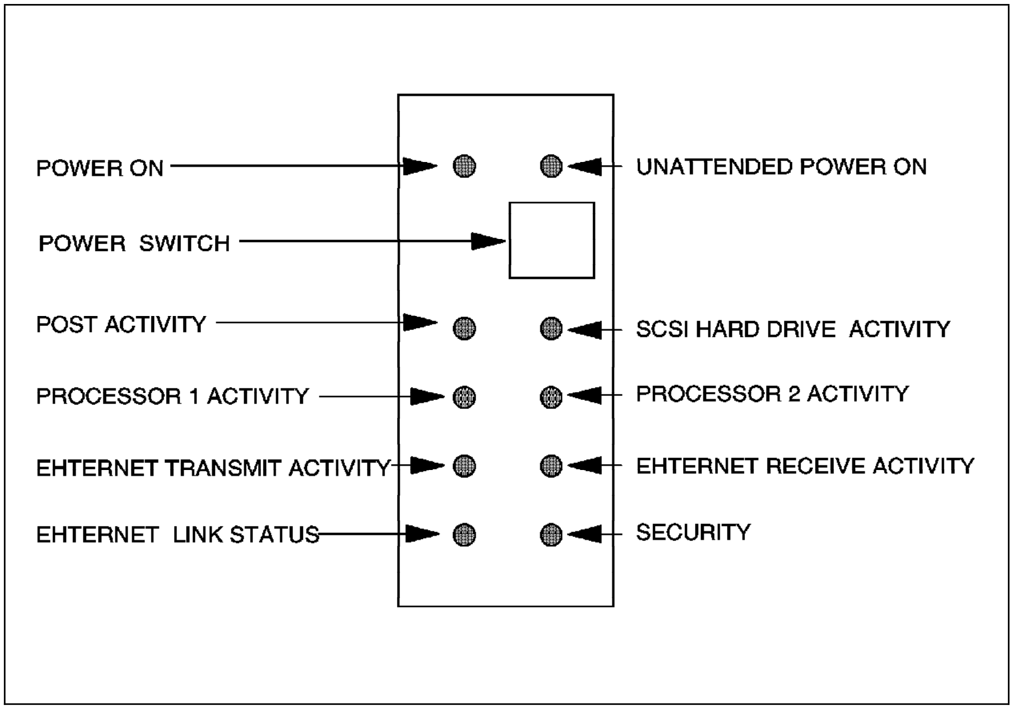


Figure 45. View of the IBM PC Server 330 Operator Panel

13.1.7 Security

The security and auditability features of the PC Server 330 models are as follows:

- Power on and privileged access password

  Functions provide control of who has access to the data and server setup program on the server.

- Set keyboard password

  Is used to lock the keyboard without turning off the system.

- A mechanical lock

  To allow the user to lock the system front cover door to prevent unauthorized personnel access to diskette drives or removable media.

- Selectable boot sequence

  Can be used to prevent unauthorized installation of software or removal of data from the diskette drive.

- Tie-down capability

  Is achievable by utilizing a common U bolt attached to the back frame.

- Operation without a keyboard and display

  Functions are supported once the server is configured with the appropriate network operating system. This minimizes the risk of unauthorized persons tampering with the system software and configuration.

- A PC Server 300 Security Cover

  Is optionally available to prevent unauthorized access to server external ports and cover retaining screws.

- A PC Server 330 Security Cable Kit

  Can detect tampering or removal of the server covers. This option, when enabled, can also alert the system administrator once this detection occurs.

- Vital Product Data (VPD) Information

13.2 Model Summary

Four models of the Server 330 are available. These are shown in the table below.

Note: IBM does not market all models in all countries.

Table 20. Server 330 200 MHz Models

Model	**8640-ES0**	**8640-ES2**	**8640-EM2**	**8640-ESS**
Processor (Speed in MHz/ Cache in KB)	Pentium Pro 200/256	Pentium Pro 200/256	Pentium Pro 200/512	Pentium Pro 200/256
SMP Capable	Yes			
Bus Architecture	PCI/EISA			
Memory	32 MB 60 ns EDO ECC standard as 1 DIMM; 1 GB EDO ECC maximum			
Disk Controller	Integrated PCI F/W Controller	Integrated PCI F/W Controller	Integrated PCI F/W Controller	PCI 3 channel F/W RAID Adapter
Total I/O Slots	4 PCI, 4 shared, and 1 EISA			
Available I/O Slots	4 PCI 4 shared 1 EISA	4 PCI 4 shared 1 EISA	4 PCI 4 shared 1 EISA	4 PCI 4 shared
Standard Hard disk	None	None	None	2 x 2.25 GB 7200 RPM
Hot-Swap Chassis	No	Standard	Standard	Standard
Total Bays	6 x 5.25″ half-high 3 x 3.5″ slimline	6 x 3.25″ hot-swap 1 x 5.25″ half-high 3 x 3.5″ slimline	6 x 3.25″ hot-swap 1 x 5.25″ half-high 3 x 3.5″ slimline	6 x 3.25″ hot-swap 1 x 5.25″ half-high 3 x 3.5″ slimline
Available Bays	5 x 5.25″ half-high 2 x 3.5″ slimline	6 x 3.25″ hot-swap 2 x 3.5″ slimline	6 x 3.25″ hot-swap 2 x 3.5″ slimline	4 x 3.25″ hot-swap 2 x 3.5″ slimline
CD-ROM	8x			
Graphics	SVGA with 1 MB RAM integrated on planar board; upgradable to 2 MB; Cirrus Logic 5436 chip set			

Note:

1. When used to describe hard drive storage, 1 GB equals 1000 million bytes.
2. When upgrading to SMP, both processors must operate at the same speed.

Chapter 14. IBM PC Server 520

The PC Server 520 offers the best of Server 320 and the storage capacity of the Server 720. It offers the same dual-SMP architecture as the Server 320 and offers the large storage capacity of the Server 720 with 18 hot-swap bays, and also the choice of PCI/EISA or PCI/MCA.

A potential user of the Server 520 may have the following requirements:

- CPU capacity of up to two Pentium processors
- Disk throughput requiring UltraSCSI performance and drives with high transfer rates
- Large disk storage requirements with 22 bays, 18 of which are hot-swap capable
- ECC memory, RAID and disk hot-swap capabilities for data security
- Protection of investment by using existing EISA or Micro Channel adapters, and have the ability to migrate to advanced PCI adapters
- An upgrade path that includes RAID disk arrays for increased data security
- Competitively priced hardware for maximum value

The Server 520 is shown in Figure 46 on page 140.

14.1 Technical Description

The Server 520 has the following overall specifications:

- Electronic complex of the Server 320
- Mechanical complex of the Server 720

Refer to Chapter 11, “IBM PC Server 320” on page 111 for information on the electronic complex of the Server 320. The Server 520 uses the hot-swap complex of the Server 720. Refer to 14.1.3, “Expansion Bays” on page 141 for further information on this.

The following are additional points of interest:

14.1.1 Memory

The Server 520 uses ECC-on-SIMM (EOS) memory as standard, providing full error checking and correcting capabilities. Refer to 4.3.4, “ECC-on-SIMM (EOS) Memory” on page 42 for more information on ECC-on-SIMM technology.

64 MB of 70 ns EOS memory is supplied as standard on PCI/EISA machines using two 32 MB SIMMs while 32 MB of 70 ns EOS memory is standard on PCI/MCA machines with a total of 256 MB possible using eight 32 MB SIMMs.

Note: Table 21 on page 141 shows the IBM supported memory options for use with the Server 520. For more information concerning the usage of various memory SIMM technologies, please see 4.3.8, “General Tips on Memory Usage” on page 45.

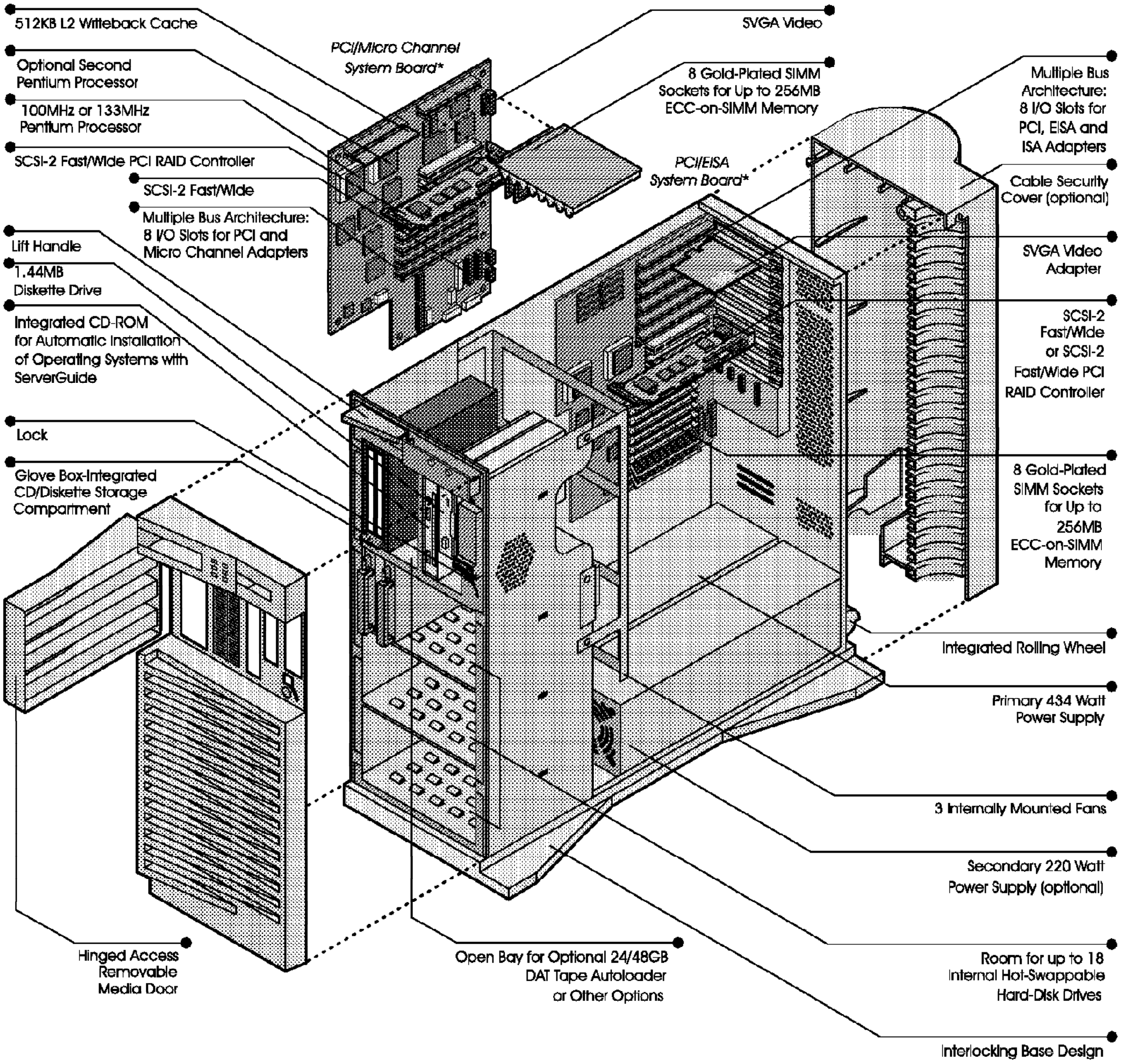


Figure 46. Exploded View of the IBM PC Server 520. The PC Server 520 ships with either a PCI/EISA or PCI/MCA system board.

Table 21. Server 520 Memory Options		
Size	**PCI/MCA Models**	**PCI/EISA Models**
32 MB	11H0645 (2x 16 MB kit)	11H0633 (2x 16 MB kit)
64 MB	11H0648 (2x 32 MB kit)	11H0636 (2x 32 MB kit)

Notes:

1. ECC-on-SIMM is standard on the Server 520
2. Memory SIMMs must be installed in matched pairs.
3. A SIMM kit contains a pair of matched SIMMs.
4. ECC-on-SIMMs cannot be mixed with Parity SIMMs.

14.1.2 Diskette Drive

The Server 520 has a 1.44 MB diskette drive installed as standard in bay A2 (refer to Figure 47 on page 142).

14.1.3 Expansion Bays

The Server 520 has a total of 22 bays, including 18 hot-swap capable as shown in Figure 47 on page 142.

Drive bays A1 and A2 hold the standard CD-ROM and 2.88 MB diskette drive respectively. These bays are both 5.25″ slimline (1″) bays.

Drive bays B1 and B2 are both 5.25″ half-high (1.6″) drive bays and do not contain any drives as standard. Bay B1 does contain a plastic removable storage compartment for CD-ROMs and diskettes as required. The separator between the B1 and B2 is removable and can form one full-height (3.6″) drive bay suitable for devices such as the 24/48 GB DAT Autoloader tape drive. (Refer to 19.2, “24/48 GB Internal Tape Autoloader” on page 189 for information on this option.)

Banks C, D and E all can contain up to six fast/wide hot swappable devices:

- Six slimline drives
- Three half-high
- Four slimline and one half-high
- Two slimline and two half-high

Bank C is already configured with a *hot-swap backplane* to allow connection of hot-swap drives into the system. The backplane has six special hot-swap connectors that allow hot-swap drives to be inserted and removed while the system is running.

Hot-swap and Non-RAID Array Systems

Hot-swapping of hard disk drives in non-array PC Server 520 models is currently supported without system reboot when using NetWare 3.12 and 4.1 and Windows NT 3.51 network operating systems.

In other operating environments, if you aren′t using a RAID controller, then it is best to shut down the server before replacing a drive. Even though the hardware would support a hot swap, the operating system would need the capability to recognize that a device has been added or replaced.

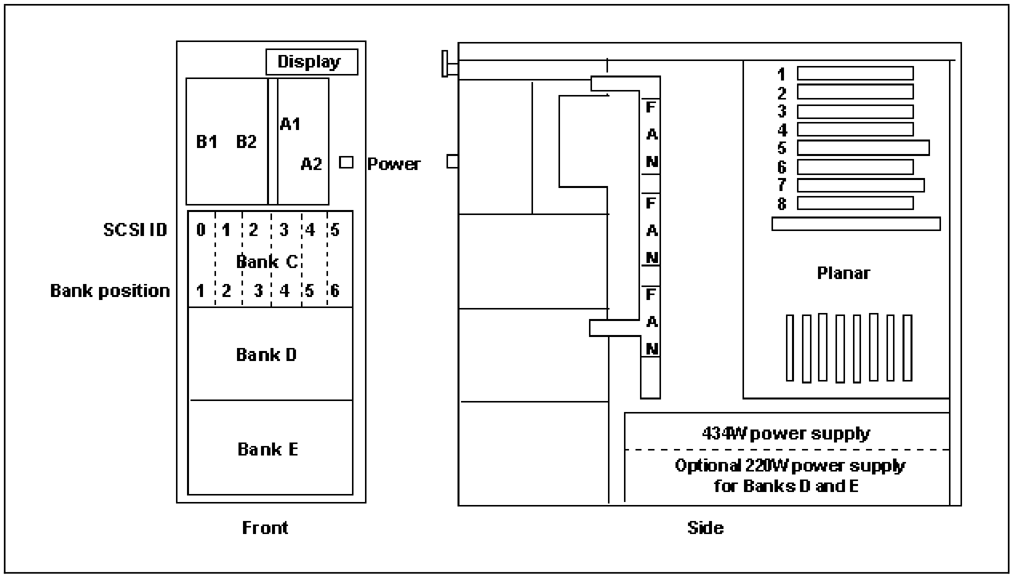


Figure 47. Side View of Server 520. This diagram shows the layout of the bays in the Server 520 system unit and the layout of the adapter slots on the planar board. Note there are four fans to provide sufficient circulation of air throughout the complex, three as shown and one in the power supply.

A 1-drop cable connects the hot-swap backplane to the SCSI-2 adapter. Attached to the backplane in Bank C is a three-drop cable connecting the CD-ROM in bay A1, a spare 68-pin connector for devices in bay B and a third connector normally terminated for a second device in B.

Neither Bank D nor Bank E are usable as standard. To use either bank, the Server 520 220 W Additional Power Supply (part #70G9739) must be installed, and a backplane must be installed for each bank using the Server 520 Hot-swap Backplane. A 1-drop cable is supplied with each backplane and is connected either to another SCSI-2 Fast/Wide adapter (or the second channel of the Streaming RAID adapter for RAID models). The backplane is actively terminated, so it will terminate itself if no other termination is present on the SCSI channel.

Note: The additional power supply is sufficient to provide power to both Banks D and E, although a hot-swap backplane is required for each bank.

Other points to note about the SCSI configuration of the Server 520 (and the Server 720 and the 3518 Enterprise Expansion Enclosure):

- 16-bit devices should be connected on the SCSI channel *before* any 8-bit devices.
- SCSI IDs are automatically assigned by position within a bank. Backplane connectors are assigned IDs 0 to 5, left to right. SCSI IDs on the drives themselves are ignored.
- There is a jumper on each backplane that allows the SCSI ID range to be either 0-5 or 8-13.
- The standard CD-ROM is assigned ID 6 at the factory.
- The SCSI controller (host) is assigned ID 7.
- Installing a device in Bank B will use an ID that is assigned to a position in Bank C. In this case, the corresponding position for that ID in Bank C must

not be used. It is recommended that a low ID be used for a device in Bank B.

- The CD-ROM is normally connected to the SCSI channel which is also driving the backplane in Bank C. If required, it can be reconnected to the channel driving the backplane in either Bank D or E, if installed.

  There are two hot-swap backplanes available for the Server 520 and Server 720. Each backplane requires the right sort of hot-swap drive tray to be used. Hot-Swap Backplane II can only use hot-swap II drive trays and Hot-Swap Backplane III can only use hot-swap III drive trays.

  - Hot-Swap Backplane II - part #94G4605
  - Hot-Swap Backplane III - part #76H2670
  - Hot-Swap Drive Tray II - part #70G9741
  - Hot-Swap Drive Tray III - part #70G9860
  - PCI to Hot-Swap Backplane II cable - part #94G3988
  - PCI to Hot-Swap Backplane III cable - part #70G9876

14.1.4 Cabling for Additional Backplanes

The Server 520 133 MHz models use the same backplanes as the Server 720. The 166 MHz models use the Hot-Swap Backplane III. However, the Server 520 uses a *PCI* RAID adapter and the additional hot-swap II backplanes have, as standard, a 1-drop cable suitable only for the *Micro Channel* RAID adapter. As a result, when installing the second and third backplanes into the Server 520, a PCI controller-to-backplane cable (part #70G9749) is also required for each backplane to replace the existing 1-drop cable.

14.1.4.1 Backplane Compatibility

Backplane II and Backplane III are not compatible with each other. You need different trays for the disk drives. To move F/W disks from a Backplane II system to a Backplane III system you have to install the drives on an SCSI-2 Fast/Wide Hot-Swap Drive Tray III.

14.2 Model Summary

The following tables summarize the different models of the PC Server 520 line.

14.2.1 PC Server 520 PCI/EISA Pentium 166 MHz Models

Table 22 shows the four Pentium 166MHz models with PCI/EISA architecture. One is a non-RAID models and three have a ServeRAID controller installed.

Table 22. Server 520 PCI/EISA Pentium 166MHz Models

Model	**8641-EE0**	**8641-EE1**	**8641-EEE**	**8641-EEL**
Processor (Int/Ext)	P166/66 MHz			
SMP Capable	Dual P166 MHz			
Bus Architecture	PCI/EISA			
Memory	64 MB 70 ns ECC-on-SIMM standard (2x32 MB); 256 MB maximum			
L2 Cache	512 KB write-back			
Disk Controller	PCI SCSI-2 Fast/Wide	PCI ServeRAID Adapter		
Total I/O Slots	2 PCI, 5 EISA, and 1 Shared			
Available I/O Slots	2 PCI and 4 EISA			
Standard Disk	None	None	4x2.25 GB 8 ms	6x2.25 GB 8 ms
Hot-swap chassis	Standard			
Total Bays	18 x 3½″ HS, 2 x 5¼″ HH, and 2 x 5¼″ SL			
Available Bays	18 x 3½″ HS 2 x 5¼″ HH	17 x 3½″ HS 2 x 5¼″ HH	18 x 3½″ HS 2 x 5¼″ HH	16 x 3½″ HS 2 x 5¼″ HH
CD-ROM	4x			
Graphics	SVGA with 1 MB RAM via an ISA adapter			

Note: When used to describe hard drive storage, 1 GB equals 1000 million bytes.

(HH = half-high; SL = slimline; HS = slimline hot swap)

14.2.2 PC Server 520 Pentium 133 MHz Models

Table 23 shows the two Pentium 133 MHz models. All models have PCI/MCA architecture and have a RAID controller as a standard feature.

Table 23. Server 520 PCI/MCA Pentium 133 MHz Models

Model	**8641-MD2**	**8641-MDG**
Processor (Int/Ext)	P133/66 MHz	
SMP Capable	Dual P133 MHz	
Bus Architecture	PCI/MCA	
Memory	32 MB 70 ns ECC-on-SIMM standard (2x16 MB); 256 MB maximum	
L2 Cache	256 KB write-back; expandable to 512 KB (512 KB recommended for two processors)	
Disk Controller	PCI SCSI-2 Fast/Wide Dual Channel RAID	
Total I/O Slots	2 PCI and 6 MCA	
Available I/O Slots	1 PCI and 5 MCA	
Standard Disk	None	4x2.25 GB 8ms
Hot-swap chassis	Standard	
Total Bays	18 x 3½″ HS, 2 x 5¼″ HH, and 2 x 5¼″ SL	
Available Bays	18 x 3½″ HS 2 x 5¼″ HH	14 x 3½″ HS 2 x 5¼″ HH
CD-ROM	4x	
SVGA Graphics	1 MB RAM; imbedded on planar; CL 5430	

Note: When used to describe hard drive storage, 1 GB equals 1000 million bytes.

(HH = half-high; SL = slimline; HS = slimline hot swap)

Chapter 15. IBM PC Server 704

The PC Server 704 is a four-way Pentium Pro system offering up to 54 GB of hot-swappable storage. It offers PCI and EISA architectures for support of industry standards.

A potential user of the Server 704 may have the following requirements:

- High CPU capacity of up to four Pentium-Pro processors for compute-intensive applications
- Disk throughput requiring SCSI-2 Fast/Wide performance and drives with high transfer rates
- Large disk storage requirements
- ECC memory, RAID, disk hot-swap and power redundancy capabilities for data security
- Optional PC Server High Availability Solution for automatic fail over for clustering environments
- Protection of investment by using existing EISA adapters, and the ability to migrate to advanced PCI adapters

15.1 Technical Description

The following sections describe each of the major subsystems.

15.1.1 Central Electronic Complex

The central electronic complex (CEC) is made up of:

- Processor
- Cache
- Memory
- Graphics

15.1.1.1 Processor

The Server 704 is SMP capable with one to four Pentium Pro 166 MHz or Pentium Pro 200 MHz processors:

- Each system ships with two processor cards.
- Each processor card contains two Zero Insertion Force (ZIF) sockets for Pentium Pro processors.
- The dual processor model (Model 7AX) contains one Pentium Pro on each processor card.
- One, two, three or four processors can be installed.
- The Pentium Pro processors operate at 66 MHz on the I/O bus.
- If multiple processors are installed they must operate at the same speed.

The following processor options are available:

- Pentium Pro 166 MHz processor upgrade, part #94G5881
- Pentium Pro 200 MHz processor upgrade, part #94G6678

The processor upgrades are designed to be installed in a vacant ZIF socket of a PC Server 704 Processor Card.

The PC Server 704 200 MHz Processor Card (part #94G6681) is designed to support the 200 MHz Pentium Pro processors. This option may be required if

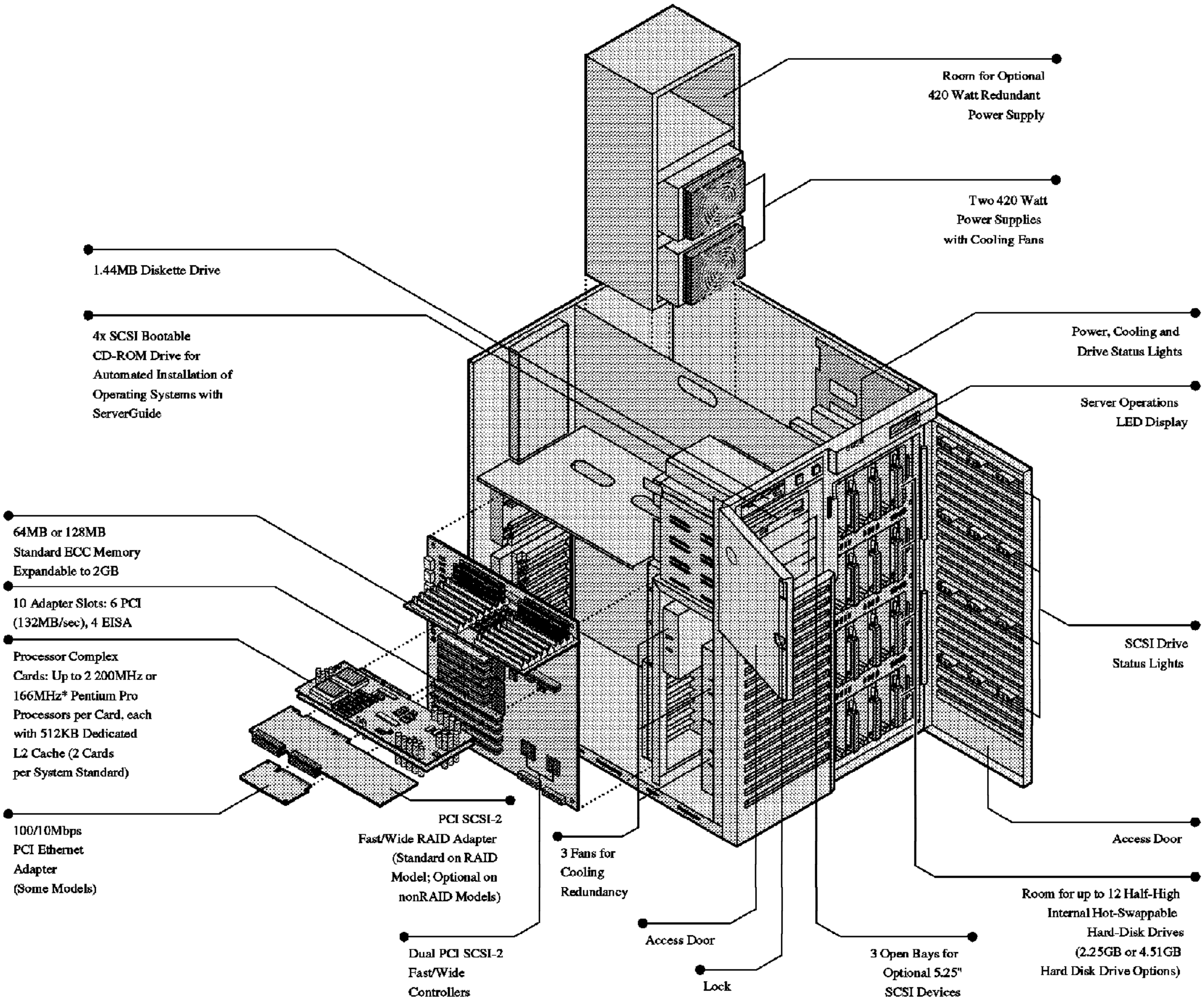


Figure 48. Exploded View of the Server 704 200 MHz

you want to upgrade 166 MHz 704 machines, as the earlier processor boards do not support 200 MHz chips. If at least one CPU is upgraded to 200 MHz, then all processor cards must support 200 MHz processors. You need to check the suffix of the safety label on the real panel. If the suffix is -003 or lower, then you need to replace the existing processor cards with the PC Server 704 200 MHz Processor Card.

15.1.1.2 Cache

The Pentium Pro processor has an integrated 512 KB L2 write-back cache as well as an integrated 16 KB L1 cache. No additional cache is provided on the planar nor on the processor cards.

15.1.1.3 Memory

The PC Server 704 provides ECC memory for added security. The non-RAID model (166 MHz) of the Server 704 has 64 MB standard using four 16 MB 60 ns parity SIMMs. The RAID model (166 MHz) has 128 MB standard using eight 16 MB 60 ns parity SIMMs. The 200 MHz model has 128 MB standard using four 32 MB 60 ns parity SIMMs. The following characteristics apply to PC Server 704 memory:

- Memory is installed on a memory card attached to the planar.
- There are a total of 16 SIMM sockets on the memory card.
- The memory supported is 60 ns parity SIMMs.
- 16 MB, 32 MB, 64 MB and 128 MB SIMMs are available.
- The Server 704 has an advanced ECC memory controller that provides the ECC function to the parity memory SIMMs. This technology is known as *ECC-P*. Refer to 4.3.3, “Error Correcting Code-Parity Memory (ECC-P)” on page 41 for more information on this technology.
- 16 MB and 32 MB SIMMs can coexist in the same server. However, 64 MB and 128 MB SIMMs cannot coexist with either 16 MB or 32 MB SIMMs.
- The PC Server 704 systems have two memory banks each with eight SIMM sockets. All memory SIMMs installed in each bank must be of the same speed and size and must be installed in matched sets of 4, 8, or 16.
- To reach the maximum 2 GB of memory, all sockets have to be with 128 MB SIMMs.
- SIMMs must be installed in matched fours, eights or sixteen. With matched fours, the Server 704 uses two-way interleaving. With matched eights, the system operates with four-way interleaved memory. Maximum performance is obtained with four-way interleaved memory. Refer to 4.3.5, “Memory Interleaving” on page 43 for more information.
- When installing SIMMs, refer to the server documentation to ensure they are installed in the correct SIMM sockets. The layout of the Server 704 card is different from that of the Server 720.

The following memory options are available:

- 16 MB 60 ns Parity: 94G5877
- 32 MB 60 ns Parity: 94G5878
- 64 MB 60 ns Parity: 94G5879
- 128 MB 60 ns Parity: 94G6882

15.1.1.4 Graphics Subsystem

The Server 704 uses a Cirrus Logic GD5424 with 0.5 MB video memory. This controller is integrated on the planar.

15.1.2 Ethernet

The Server 704 (166 MHz) ships with a 100/10 Mbps PCI Ethernet adapter. It includes a single RJ-45 connector for attachment to 10Base-T and 100Base-TX networks. For 100 Mbps operation, Category 5 cabling must be used. For the 200 MHz models, a LAN adapter may be ordered as an option.

15.1.3 SCSI Controller

The PC Server 704 has two RISC-based Adaptec AIC-7880 compatible UltraSCSI Fast/Wide controllers integrated on the planar. Connectors are provided from these two controllers to the two standard hot-swap backplanes.

The RAID model also has a PCI dual-channel SCSI-2 Fast/Wide RAID adapter that supports RAID levels 0, 1 and 5 with up to seven devices per channel. The RAID adapter is based on the i960 RISC processor and has 4 MB of on-board cache. Cabling is supplied to attach the two channels of the adapter to each of the two hot-swap backplanes. The RAID controller also has one external Fast/Wide port with an *Industry Standard* connector (as defined in "SCSI-2 Fast/Wide External Connectors" on page 175).

Note: Only two ports of the RAID controller can be used at one time.

Should you need to upgrade the non-RAID model to RAID capability, order the following parts:

- PC Server 704 SCSI-2 F/W PCI RAID Adapter: 94G5884
- 4-drop SCSI Cable: 94G5883

The 4-drop cable should be ordered when converting to RAID to connect the standard CD-ROM and other non-RAID devices installed in the system to the integrated SCSI-2 F/W controller.

15.1.4 Expansion Slots

The Server 704 uses the following two industry standard bus architectures:

1. PCI - Supports data transfers of of up to 132 MBps and runs at 33 MHz
2. EISA - Supports 33 MBps data transfers and operates at 8.33 MHz

The system uses a Phoenix BIOS, flash upgradable via diskette.

The Server 704 has 10 slots:

- Six PCI 32-bit bus master slots
- Four EISA 32-bit bus master slots

The RAID model of the Server 704 has an SCSI-2 Fast/Wide RAID adapter installed in one of the PCI slots.

15.1.5 Expansion Bays

The Server 704 has 17 drive bays for internal data storage, as shown in Figure 48 on page 148:

- Two hot-swap banks each with six 3.5″ half-high hot-swap drive bays
- One 5.25″ half-high bay containing the standard quad-speed CD-ROM
- One 3.5″ slimline bay for the standard 1.44 MB diskette drive
- Three 5.25″ half-high bays, available and accessible from the front of the system

Note: The hot-swap drives in the Server 704 are half-high bays. This is in contrast to other hot-swap bays in the server range that use slimline drives.

Two of the three 5.25″ half-high bays can be converted into a single full-height for options such as the 24/48 GB Internal Tape Autoloader. (Refer to 19.2, "24/48 GB Internal Tape Autoloader" on page 189 for information on the 24/48 GB Autoloader.)

Note: It is not recommended that hard drives be installed in these 5.25″ half-high bays due to electromagnetic interference (EMI) generated by these devices and the electromagnetic discharge (ESD) susceptibility that could damage the drives.

The non-RAID model (166 MHz) of the Server 704 has two 2.14 GB SCSI-2 Fast/Wide hard drives as standard. One 2.14 GB drive is installed in each backplane. A narrow (that is, 8-bit) 3-drop SCSI cable is daisy-chained from the top hot-swap backplane to the standard CD-ROM and one or two other devices installed in the open bays.

The RAID model has four 2.14 GB SCSI-2 Fast/Wide drives as standard. Two drives are installed in each backplane. A narrow 4-drop SCSI cable is connected to one of the two integrated SCSI-2 F/W controllers to connect the standard CD-ROM and up to three other devices installed in the open bays.

The 200 MHz model has no drives as standard.

Additional hot-swap drives can be ordered using part numbers:

- PC Server 704 2.14 GB SCSI F/W HDD: 94G5882
- PC Server 704 4.51 GB SCSI F/W HDD: 94G6679

15.1.5.1 CD-ROM

A quad-speed SCSI CD-ROM is installed in the Server 704. It is attached to the integrated SCSI-2 Fast/Wide controller via an SCSI-2 Fast (8-bit) cable.

15.1.6 Redundant Power Supply

The Server 704 has two 420 Watt power supplies standard with an optional additional redundant power supply. This option allows the Server 704 to operate without interruption if one of the two standard power supplies fails. The part number of this redundant power supply varies from country to country and is of the form SPSCxxx.

15.2 Model Summary

Table 24 shows the three specific models that are available.

Table 24 (Page 1 of 2). Server 704 Models

Model	**8650-4BW**	**8650-7AX**	**8650-4M0**
Processor (Int/Ext)	Pentium Pro (166/66)	Pentium Pro (166/66)	Pentium Pro (200/66)
Standard Processors	1	2	1
SMP Capable	One to four Pentium Pro processors of the same speed		
Memory	64 MB (4x16 MB)	128 MB (8x16 MB)	128 MB (4x32 MB)
L2 Cache	512 KB L2 write-back cache integrated on processor chip		
Disk Controller (SCSI-2 F/W)	PCI Dual Channel integrated	PCI Dual Channel RAID adapter	PCI Dual Channel integrated
Total I/O Slots	6 PCI and 4 EISA		
Available I/O Slots	5 PCI 4 EISA	4 PCI 4 EISA	6 PCI 4 EISA
Standard Hard Drive	2 x 2.14 GB 9 ms 7200 RPM	4 x 2.14 GB 9 ms 7200 RPM	NONE

Table 24 (Page 2 of 2). Server 704 Models			
Model	**8650-4BW**	**8650-7AX**	**8650-4M0**
Hot-Swap Chassis	Standard		
Total Bays	12 x 3½″ hot-swap 4 x 5¼″ half-high 1 x 3½″ slimline		
Available Bays	10 x 3½″ hot-swap 3 x 5¼″ half-high	8 x 3½″ hot-swap 3 x 5¼″ half-high	12 x 3½″ hot-swap 3 x 5¼″ half-high
CD-ROM	4x		
Graphics	SVGA with 0.5 MB Video RAM		
Ethernet Adapter	100/10 Mbps PCI Ethernet adapter with RJ-45 connector		Optional
Redundant Power Supply	Optional		
Note: To reach the 2 GB maximum RAM capacity, the existing SIMMs must be replaced by 128 MB SIMMs.			

Chapter 16. IBM PC Server 720

The PC Server 720 is a very high-capacity enterprise-wide server designed to provide very high levels of performance and capacity.

A potential user of the Server 720 may have the following requirements:

- Six-way processing power using CPUs up to Pentium 166 MHz
- Fast memory subsystem with up to 1 GB of RAM
- ECC, RAID and hot-swap capabilities for data security
- Optional PC Server High Availability Solution for automatic fail over for clustering environments
- Large disk storage requirements within the server cabinet as well as in external enclosures
- PCI and Micro Channel slots for investment protection

The PC Server 720 is ideal as a large application/database server or very large file/print server. The Server 720 is shown in Figure 49.

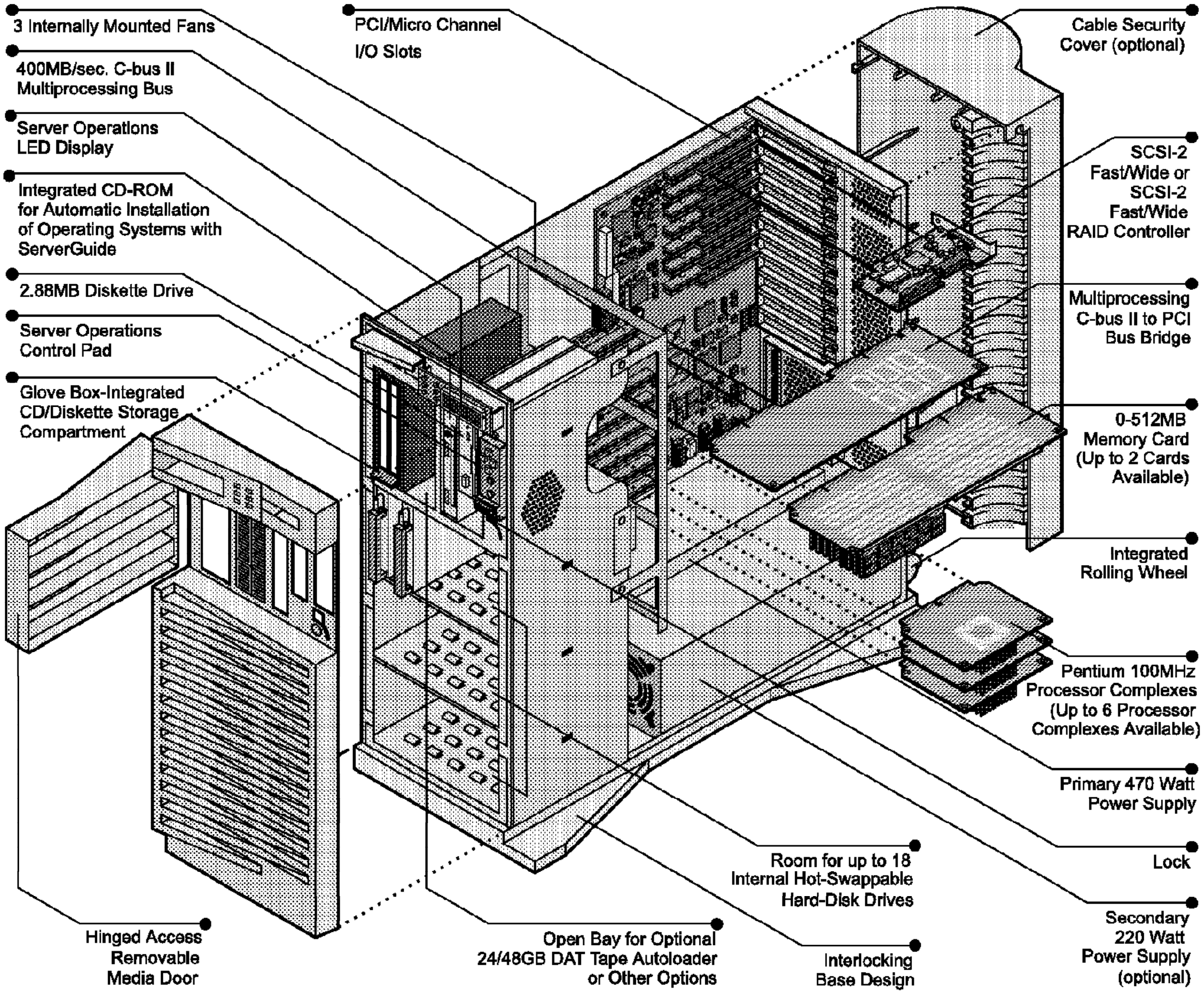


Figure 49. Exploded View of the IBM PC Server 720

One of the key factors that affect the performance of PC servers is *balance*; that is, all subsystems are designed and tuned to ensure there are no bottlenecks. The memory, disk and network subsystems all match the power of the system′s processor subsystems. This is most critical in symmetrical multiprocessor systems. If the processor subsystem is significantly more powerful than the other subsystems, then bottlenecks develop, limiting the extent to which the benefit of having more than one processor can be realized.

16.1 Technical Description

After a summary of the architecture of the Server 720, each of the following subsystems is described:

- Multiprocessor bus
- Processor subsystem
- Memory subsystem
- PCI and Micro Channel subsystem
- Service subsystem (I•C Bus)
- Storage subsystem
- Graphics subsystem

16.1.1 Architecture

Figure 50 on page 155 shows the architecture of the PC Server 720. The central interface of the entire system is the C-Bus II multiprocessor bus. This provides the necessary bandwidth for six processor complexes, two memory cards each capable of holding 512 MB of 60 ns ECC memory, and a bridge to interface to the primary PCI bus. The main I/O interface is via the primary PCI bus. With the combination of the primary and secondary PCI buses, a total of seven PCI adapter slots are available. Also attached to the primary PCI are the Micro Channel bus providing seven MCA slots, an I/O controller providing port functions, and the I•C bus providing systems management and service control facilities to the system.

16.1.2 Multiprocessor Bus

The multiprocessor bus is based on Corollary′s C-Bus II architecture and provides the high-speed connection (400 MBps) necessary for the processor complexes to communicate with each other, to main memory and to the PCI bridge. The bus has eight connections as follows (refer to Figure 51 on page 159):

- Slot 0 is for the bridge between the multiprocessor bus and the primary PCI bus.
- Slot 1 is dedicated to a memory card.
- Slot 2 supports either a memory card or a processor card.
- Slots 3-7 are reserved for processor cards.

The design of the Server 720 is such that the multiprocessor bus is the major line for data transfer between the major components of the system. With its 50 MHz clock cycle and 64-bit width, it is able to transfer 8 bytes per cycle or 400 MBps. The bus is able to operate at this high speed due to its use of new switching technology, which overcomes the limitations of existing bus designs that trade-off speed against the number of expansion slots.

To enhance reliability, the bus itself uses ECC to guard against errors on the bus that might be caused by factors such as electrostatic discharges.

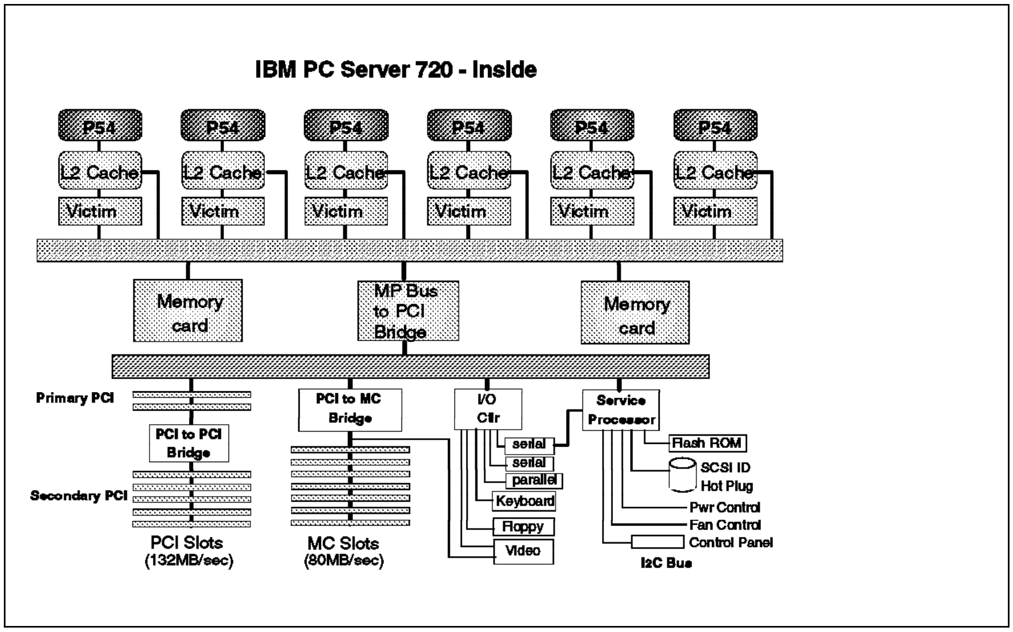


Figure 50. The Architecture of the IBM PC Server 720

An important feature of the bus design is that it runs independently to the speed of the processors. This protects a customer′s investment by allowing processor speeds to be mixed. As an example, a Server 720 with one Pentium 100 MHz processor complex can be upgraded to have additional processors such as the Pentium 133 MHz and Pentium 166 MHz complexes. This feature is known as *processor agility*. This can be achieved without removing the existing processor.

Note: The design of the multiprocessor bus provides for one dedicated memory card slot, five dedicated processor card slots and one shared slot that can contain either a memory card or a processor card. Therefore, if six processors are required, then only one memory card can be installed. Conversely, if two memory cards are installed, then only five processors can be used. With current memory technology, one memory card can be fully populated with a total of 512 MB or 60 ns ECC RAM.

The multiprocessor bus-to-PCI bridge is a special adapter that physically plugs into two sockets, one on the multiprocessor bus and one on the PCI bus. The circuitry on this card controls the movement of data between the processor complexes, memory cards and the I/O adapters on the Micro Channel and PCI buses. It implements a split bus design so that during slower I/O operations the multiprocessor bus can be made available for processor to memory transfers without having to wait on the I/O transfer.

16.1.3 Processor Subsystem

Up to six processor complexes can be installed onto the multiprocessor bus on the Server 720. Each processor complex has the following functions:

- An Intel Pentium processor. Currently announced are 100 MHz, 133 MHz, and 166 MHz options.
- 512 KB of write-back L2 cache.
- A 192 byte L3 (victim) cache.
- Bus interface and control logic.

To ensure highest performance, the processor and cache are designed to operate with zero wait states. In addition, all processor cards are fully symmetrical for main memory access, I/O, and interrupts. This means that no single processor can be a bottleneck, for example, through having to process all the hardware interrupts in the system.

In an SMP system, some data might be held within multiple processor caches. If this data is modified by one processor, then it needs to be updated in the other caches before another processor operates on out-of-date information. This is called cache coherency and is handled by the *MESI* (Modified, Exclusive, Shared or Invalid) protocol. For an explanation of MESI, refer to 4.2, "Cache in an SMP Environment" on page 37.

Each of the PC Server 720 processor cards has three levels of private cache. Lower level caches are typically very fast but also very expensive. In general, the closer to the processor, the faster and more expensive the cache will be.

The first, or L1 cache, is built into the Pentium processor. This is th first place the processor will look when it needs its next instruction or piece of data to operate on. If this data is not contained in L1, the system next searches the second level, or L2 cache. If the information is found in either of these locations, the processor can continue without incurring wait states. However, if the information is still not available from the L1 cache or the L2 cache, the processor will need to obtain it from main memory. Depending on the speed of the memory subsystem, it may take a few clock cycles for this transaction to complete.

When the information is retrieved from main memory, it is stored in a free cache location. If there are no free locations, then the new data replaces the least frequently used information currently stored in the cache.

Since the Server 720 uses performance-boosting write-back cache, this least frequently used data may have already been modified by the CPU and must, therefore be evicted first or written back to memory. In most systems, the processor would have to wait for that data to be written out before it could bring in a new line of cache.

The Server 720 takes the unique approach of adding a small six-line L3 cache to each processor card. The 720′s L3 or *victim* cache allows that data to be quickly cleared from L2 to L3 making room for new information and allowing modified data to be written to main memory later during free bus cycles. In addition, this information is still available to be quickly accessed if needed in future cycles.

This approach improves the overall cache hit ratio with little additional cost while saving significant clock cycles during the write back or eviction process. Access to data in any of these three caches can be accomplished with zero wait states so the processor continues work on its applications without wasting time.

It may seem more effective to simply make the L2 cache much larger to increase the cache-hit ratio directly. However, since the cache memory tends to be expensive, the size of the L2 cache is carefully chosen to provide a high cache-hit rate at reasonable cost. As it takes time to search the cache for the required information, it is also possible to have a cache that is too large and too slow to respond to each request.

The carefully balanced cache design of the 720, however, provides for a high percentage of zero-wait states processing resulting in excellent multiprocessor performance. Refer to 4.1, "Cache" on page 34 for more information on cache.

Currently, the following processor complexes are announced:

- 100 MHz Processor Option (part 94G2724) - For all supported operating systems other than NetWare 4.1 SMP
- 100 MHz Processor Option II (part #94G6054) - For NetWare 4.1 SMP only
- 133 MHz Processor Option (part #94G6055) - For all supported operating systems other than NetWare 4.1 SMP
- 133 MHz Processor Option II (part #94G6056) - For NetWare 4.1 SMP only
- 166 MHz Processor Option II (part #94G6057) - For all supported operating systems

Note: Of a maximum of six processors in the 720, the number of earlier 166 MHz standard processor boards is limited to two. This only applies to processor boards delivered as standard processors. This limit of two 166 MHz standard processors is due to a limitation of the power supply. These 166MHz boards draw current from the 5 V power supply and not the 3.3 V supply as do the other processor boards. All 166 MHz optional processors boards (P/N 94G6057) are 3.3 V. They may be added without a restriction.

16.1.4 Memory Subsystem

To ensure data is fed to the processor complexes in a timely fashion, the memory subsystem is four-way interleaved. In sequential data requests, four-way interleaving can greatly improve system performance by retrieving not only the data in the requested memory location, but also data in the next three data locations. Refer to 4.3.5, "Memory Interleaving" on page 43 for more information. To achieve four-way interleaving, memory SIMMs must be installed in the Server 720 in sets of eights.

Up to two memory cards can be installed on the C-Bus II multiprocessor bus. Each memory card has two banks of eight SIMM sockets, supporting 72-pin 60 ns ECC SIMMs. The memory modules used are industry standard 60 ns ECC SIMMs and are 36-bits wide (32 bits of data plus four bits to implement ECC functions).

SIMMs are currently available in 8, 16, or 32 MB sizes. Each memory card can support 512 MB of memory (using 16 32 MB modules). Part numbers supported are:

- 8 MB 60 ns ECC: 94G2939
- 16 MB 60 ns ECC: 94G2940
- 32 MB 60 ns ECC: 94G2941

Table 25 on page 158 shows the memory configurations that are supported in the Server 720.

Table 25. Allowable Memory Configurations of the Server 720. These are the only configurations available. Each bank contains eight SIMMs of identical size and all SIMMs are 60 ns ECC SIMMs.

Memory Card One		Memory Card Two		Total (MB)
Bank One	**Bank Two**	**Bank One**	**Bank Two**	
8 MB				**64 MB**
8 MB	8 MB			**128 MB**
8 MB	8 MB	8 MB		**192 MB**
8 MB	8 MB	8 MB	8 MB	**256 MB**
16 MB				**128 MB**
16 MB	16 MB			**256 MB**
16 MB	16 MB	8 MB		**320 MB**
16 MB	16 MB	8 MB	8 MB	**384 MB**
16 MB	16 MB	16 MB		**384 MB**
16 MB	16 MB	16 MB	16 MB	**512 MB**
32 MB				**256 MB**
32 MB	16 MB			**384 MB**
32 MB	32 MB			**512 MB**
32 MB	32 MB	8 MB		**576 MB**
32 MB	32 MB	8 MB	8 MB	**640 MB**
32 MB	32 MB	16 MB		**640 MB**
32 MB	32 MB	16 MB	16 MB	**768 MB**
32 MB	32 MB	32 MB		**768 MB**
32 MB	32 MB	32 MB	16 MB	**896 MB**
32 MB	32 MB	32 MB	32 MB	**1024 MB**

Notes:

1. Memory SIMMs must be installed in sets of eight.
2. 8 MB SIMMs cannot coexist with 16 MB or 32 MB SIMMs on the same memory card. The 8 MB SIMMs are based on 4 Mb (megabit) technology, whereas the 16 MB and 32 MB SIMMs are based on 16 Mb technology. Therefore, a memory card cannot contain 8 MB SIMMs with either 16 MB or 32 MB SIMMs.
3. Higher capacity SIMMs must always be populated first. For example, 32 MB SIMMs in bank one and 16 MB SIMMs in bank two are allowed. The reverse is not allowed. The same holds true between memory cards also. For example, 32 MB SIMMs in card one and 16 MB SIMMs in card two is allowed. The reverse are not allowed.

16.1.5 PCI/MCA I/O Subsystem

The PC Server 720 has capacity to use seven adapters. These adapters can be either PCI or Micro Channel. Physically, the PCI card slots are paired with the Micro Channel card slots on the system board. (Refer to Figure 51 on page 159.) Within each pair only one PCI or one Micro Channel adapter can be installed. Each pair of slots also has an LED associated with it to indicate any active adapters.

The seven PCI slots in the PC Server 720 conform to the PCI Local Bus Specification, Revision 2.0, which provides a burst data transfer rate of up to 132 MBps. All of the PCI cards installed can be bus masters. The PCI specification has a limit of 10 *PCI loads* that can be attached to the PCI bus. In the Server 720 only two PCI slots are available on the primary PCI bus. To enable more adapter slots, a PCI-to-PCI bridge circuit connects an additional five PCI slots on

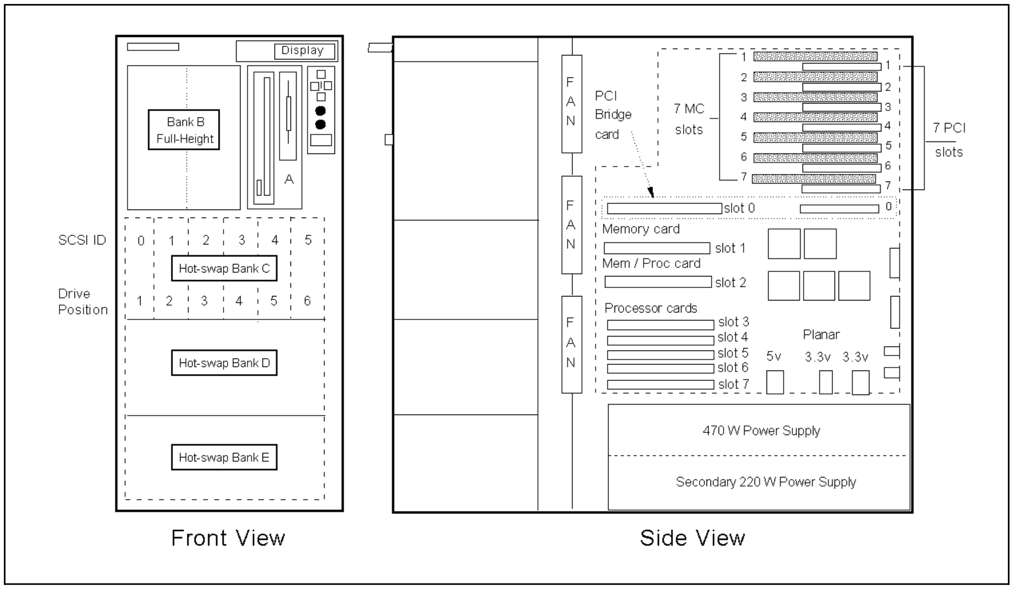


Figure 51. Side View of Server 720

the secondary PCI bus to the primary PCI bus. As per Figure 51 on page 159, PCI slots 6 and 7 are on the primary PCI bus. Slots 1 to 5 are on the secondary PCI bus. There is no performance degradation using the secondary PCI bus other than a 1-clock cycle delay during initialization. Slot five also has the video bus extension.

For non-RAID models of the Server 720, slot one contains the PCI SCSI-2 Fast/Wide Adapter, part #70G8498. Refer to 17.3, "SCSI-2 Fast/Wide PCI Adapter and Adapter II" on page 164 for information about this adapter.

Seven 32-bit adapter slots accept Micro Channel adapters and support up to 80 MBps data streaming on the bus. All slots accept type 5 Micro Channel cards.

For RAID models of the Server 720, slot one contains the SCSI-2 Fast/Wide PCI-bus RAID Adapter, part #94G2764. Refer to 3.3.5, "PCI Local Bus" on page 27 for information about this adapter.

Non-RAID models may be upgraded to RAID by ordering the optional PC ServeRaid Adapter, part #70G8489. Refer to 17.10, "ServeRAID SCSI Adapter" on page 169 for information about this adapter.

16.1.6 Graphics Subsystem

The SVGA graphics controller is integrated on the planar. It uses the Cirrus Logic 5429 chip set with 1 MB of video RAM. Earlier versions of the PC Server 720 used the Cirrus Logic 5428 chip set with 0.5 MB of video RAM.

16.1.7 Service Subsystem

The I•C (Inter-Integrated Circuit) bus is a serial bus that is used in the PC Server 720 to monitor and control several peripheral functions within the machine. The bus is driven through a VLSI module on the system board. A service processor on the system board (based on the Intel 83C552 micro-controller) controls and monitors these functions within the machine through the I•C bus. These include:

- The front panel LCD
- The front panel power controls

- Detecting the setting of the SCSI ID jumper on the hot-swap backplanes
- Reading the voltage and temperature sensors on the planar (these can be viewed by the operator using NetFinity)
- Controlling the three fans mounted behind the drive bays

16.1.8 Storage Subsystem

The Server 720 uses the same chassis as the Server 520 offering the following characteristics: (Refer to Figure 51 on page 159.)

- 2.88 MB 3.5″ diskette drive in bay A2
- CD-ROM in bay A1
- Two 5.25″ half-high (1.6″) bays convertible to one full-height bay (suitable for the 24/48 GB Autoloader; refer to 19.2, “24/48 GB Internal Tape Autoloader” on page 189)
- 18 hot-swap capable slimline drive bays - banks C, D and E.

Refer to 14.1.3, “Expansion Bays” on page 141 for a complete description of these bays.

Note: The backplanes used in the Server 720 are Hot-Swap Backplane II (part #94G4605). The compatible cable is the SCSI PCI Option Card to Backplane Fast/Wide Cable (part #94G3988 or SC7201Y).

16.2 Model Summary

Table 26 shows the two currently offered models of the PC Server 720.

Table 26. PC Server 720 Standard Models

Model	**8642-0E1**	**8642-2E1**
Processor (Int/Ext)	P166/66 MHz	
Standard Processors	1	2
SMP Capable	One to six processor complexes; processors do not have to be identical; current processor complexes available: Pentium 100, 133, 166 MHz	
Memory	128 MB 60 ns ECC standard (8x8 MB); four-way interleaved; 1 GB maximum using 32x32 MB SIMMs	
L2 Cache	write-back, direct-mapped with supporting victim cache; 64-bit path to processor	
L2 Cache capacity	512 KB per processor	
L3 Cache	192 bytes per processor *victim cache*	
Disk Controller	PCI SCSI-2 F/W	Streaming RAID
Total I/O Slots	7 MCA/PCI	
Available I/O Slots	6 MCA/PCI	
Standard Hard Disk	None	2 x 2.25 GB 7.5 ms
Hot-Swap Chassis	Standard	
Total Bays	18 x 3½″ hot(*) 2 x 5¼″ slimline 2 x 5¼″ half-high	
Available Bays	18 x 3½″ hot-swap 2 x 5¼″ half-high	16 x 3½″ hot-swap 2 x 5¼″ half-high
CD-ROM	2x	
Graphics	SVGA with 1 MB imbedded on planar	

Note: When used to describe hard drive storage, 1 GB equals 1000 million bytes.

(*) The first six bays are hot-swap capable as standard; the remaining 12 bays require a backplane for each bank and the additional power supply to be hot-swap capable.

Chapter 17. Storage Controllers

This chapter describes the major storage controllers used in IBM servers. The following adapters are discussed:

- IBM SCSI-2 Fast PCI Adapter
- IBM SCSI-2 Fast/Wide Adapter/A
- IBM SCSI-2 Fast/Wide PCI Adapter and Adapter II
- IBM SCSI-2 Fast/Wide Streaming RAID Adapter/A
- IBM SCSI-2 Fast/Wide PCI-bus RAID Adapter
- IBM UltraSCSI PCI Adapter
- PC Server 704 SCSI-2 F/W PCI RAID Adapter
- PC Server 320 Embedded SCSI controller
- PC Server 330 Embedded SCSI controller
- PC ServeRAID Adapter
- IBM SSA Adapter

A section is also included on the different SCSI connectors used on IBM controllers.

17.1 SCSI-2 Fast PCI Adapter

The IBM SCSI-2 Fast PCI Adapter is a high-throughput 32-bit PCI device capable of data transfer rates of up to 132 MBps. It performs DMA transfers to reduce the CPU overhead by transferring data into the system memory directly. It can operate in synchronous and asynchronous modes with a peak data transfer rate of 10 MBps. It is based on the Adaptec AHA-2940U controller and supports advanced SCSI features such as multi-threaded I/O, scatter/gather, and tagged command queuing (TCQ).

The adapter has one internal port (50-pin connector) and one external port (50-pin connector). It can support up to seven SCSI or SCSI-2 Fast devices (for example, three internally and four externally). A 50-pin 4-drop internal SCSI cable is standard with all models for connection to the standard CD-ROM (and hard drive where a hard drive is standard). It ships with an ASPI manager.

The IBM SCSI-2 Fast PCI Adapter ships with some models of the PC Server 310. It is not orderable as a separate option.

17.2 SCSI-2 Fast/Wide Adapter/A

The IBM SCSI-2 Fast/Wide Adapter/A (part #70G8498) is a high-performance Micro Channel bus master adapter capable of streaming data at 40 MBps. It has dual SCSI-II fast and wide channels (one internal and one external). It supports devices using either asynchronous, synchronous, or fast synchronous SCSI data transfer rates, with a peak data transfer rate of 20 MBps. It also supports standard 8-bit SCSI devices.

The adapter has a dedicated 80C186 local processor on board, which allows it to implement advanced features such as TCQ. It has a dual bus design that prevents access to internal DASD from the external port. This also allows the maximum cable length to be calculated individually for each bus for additional capability externally.

The adapter is shown in Figure 52 on page 164.

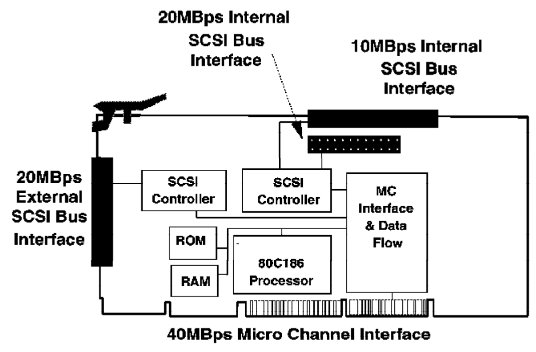


Figure 52. SCSI-2 Fast/Wide Adapter/A

Up to 15 devices can be connected to the adapter using a combination of internal and external ports. The internal channel supports up to seven SCSI devices and the external channel supports from 1 to 14 devices.

There are three ports on the adapter:

1. Internal 68-pin Fast/Wide port
2. Internal 50-pin SCSI-2 Fast port
3. External 68-pin Fast/Wide port with an IBM SCSI-2 P connector (see 17.12, “Choosing SCSI Cables” on page 175)

Notes:

1. Only two ports can be used at any one time.
2. The external port supports attachment of both 8-bit and 16-bit devices.
3. Multiple adapters are supported in the same machine with the limit being the number of Micro Channel slots.

The IBM SCSI-2 Fast/Wide Adapter/A is standard in PC Server 500 non-RAID models (and older systems, including the PS/2 Server 95).

17.3 SCSI-2 Fast/Wide PCI Adapter and Adapter II

The IBM SCSI-2 Fast/Wide PCI Adapter II (part #94G4673) is a high-performance 32-bit PCI adapter capable of data transfer rates of up to 132 MBps. This adapter is based on the Adaptec AHA-2940W SCSI controller. It can operate in synchronous and asynchronous modes with a peak data transfer rate of 20 MBps. It has an on-board RISC processor and supports advanced SCSI features such as Multi-threaded I/O, Scatter/Gather, and Tagged Command Queuing (TCQ).

This PCI bus master adapter is shown in Figure 53 on page 165.

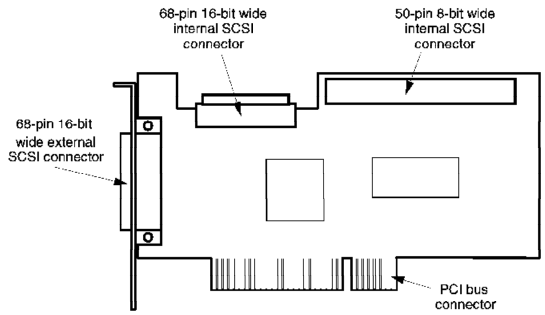


Figure 53. PCI SCSI-2 Fast/Wide Adapter

Note: Early models of the PC Server 320 non-RAID array PCI/EISA models used a different SCSI-2 F/W controller, part #94G3771. This adapter was known as *PCI SCSI-2 Fast/Wide Adapter* and is now withdrawn and no longer available.

The old 94G3771 and the new 94G4673 had the same specifications although the old adapter operated at 42MHz whereas the new adapter operates at 40 MHz. One implication of this difference is that you cannot use the different adapters together in a duplex configuration (that is, a software based disk mirror configuration).

Like the SCSI-2 Fast/Wide Adapter/A, this adapter offers connection of up to 15 devices using a combination of internal and external ports. The internal channel supports up to seven SCSI devices and the external channel supports 1 to 14 devices. The internal and external channels are independent to secure internally connected devices from being accessed externally.

The adapter has three ports:

1. Internal 68-pin Fast/Wide port
2. Internal 50-pin SCSI-2 Fast port
3. External 68-pin Fast/Wide port with a High Density 68-pin connector (see 17.12, “Choosing SCSI Cables” on page 175)

Notes:

1. Only two ports can be used at any one time.
2. Up to three PCI SCSI-2 Fast/Wide adapters can be used in a system.

The IBM SCSI-2 Fast/Wide PCI Adapter II is installed in the PC Server 320 PCI/EISA models, Server 520 PCI/EISA models and the Server 720 non-RAID models.

17.4 SCSI-2 Fast/Wide Streaming RAID Adapter/A

The IBM SCSI-2 Fast/Wide Streaming-RAID Adapter/A (part #70G9263) is a high-performance SCSI-II fast and wide 32-bit Micro Channel adapter with a built-in RAID controller. It is a high-throughput adapter that supports 80 MBps Data Streaming on the Micro Channel. The dual SCSI-II channels support asynchronous, synchronous, or fast synchronous SCSI modes with a peak data transfer rate of 20 MBps. It supports RAID-0, RAID-1 enhanced and RAID-5 configurations. (Refer to 5.5, “RAID Technology” on page 62 for information about RAID levels.)

The adapter is an i960 RISC processor-based design. It has 4 MB of 60 ns cache, configurable as write-through or write-back. It supports a wide range of IBM options, including the Fast/Wide hot-swap disk drives.

80 MBps Data Streaming

In order to use the 80 MBps Micro Channel Data Streaming feature, the BIOS level of the adapter must be at level 2.19 or higher. This BIOS update can be obtained from a number of sources, including the Web and the PC Company BBS. Refer to Appendix A, “Sources of Drivers and Information” on page 203 for details.

The adapter has two independent SCSI channels that can be configured with both internal and external array configurations. A maximum of 14 devices may be attached, seven on each individual channel. The adapter supports both 16-bit (wide) and 8-bit SCSI devices.

The following connectors are provided on the adapter:

1. One internal Fast/Wide connector - channel 1
2. One internal Fast/Wide connector - channel 2
3. One external Fast/Wide (*IBM SCSI-2 P*) connector - channel 2 (see 17.12, “Choosing SCSI Cables” on page 175)

Note: Channel 2 cannot support both internal and external devices simultaneously.

The array controller on this adapter has the following characteristics:

- A maximum of four arrays can be defined per adapter.
- An array can span both channels of the adapter, using either both internal ports or one internal port and one external port.
- Up to eight physical drives can be defined in an array.
- An array cannot span two adapters.
- Up to seven physical drives can be physically connected to each channel.
- An array can have one or more logical drives.
- Each logical drive within an array can be a different RAID level (for example, RAID-1 and RAID-5).
- A maximum of eight logical drives can be defined per adapter, irrespective of the number of arrays defined.

When the Streaming RAID Adapter is purchased as an option, a side card, or daughter card, is provided that allows the connection of both Fast/Wide channels externally. This daughter card connects with an internal cable to the second internal Fast/Wide port and occupies the next Micro Channel slot providing an external SCSI connector.

Note: If the boot device is connected to the Streaming RAID adapter, make sure that the INT 13 support of the Streaming RAID adapter is enabled. INT 13 support can be enabled/disabled from the Reference Diskette by selecting:

1. Set configuration
2. Change configuration

Up to four Streaming RAID adapters can be installed in a server.

17.5 SCSI-2 Fast/Wide PCI-bus RAID Adapter

This high-performance adapter (part #94G2764) has the same capabilities as the Streaming RAID Adapter/A except for the bus interface, which is a 132 MBps PCI bus. Refer to 17.4, "SCSI-2 Fast/Wide Streaming RAID Adapter/A" on page 165 for more information on the characteristics of the Streaming RAID adapter. Figure 54 shows the layout of the RAID PCI adapter and the position of the i960 RISC processor and the 4 MB cache.

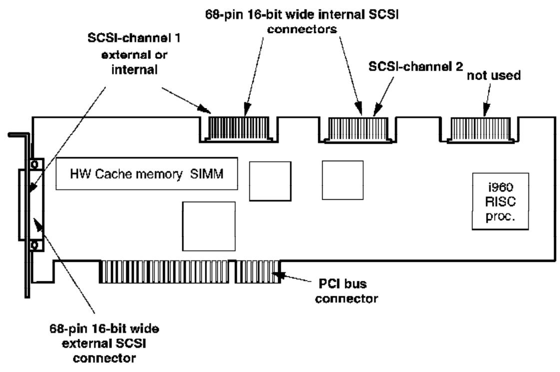


Figure 54. PCI SCSI-2 Fast/Wide Adapter

Up to four RAID adapters can be installed in a server.

17.6 Ultra Wide SCSI PCI Adapter

The IBM Ultra Wide SCSI PCI Adapter (part #76H5407) is a high-performance 32-bit PCI adapter capable of data transfer rates of up to 132 MBps. This adapter features a RISC processor that reduces the time it takes to process SCSI commands and reduces the CPU overhead. The adapter features data transfer rates of up to 40 MBps and advanced SCSI features such as:

- Multi-threaded I/O
- Scatter/gather
- Tagged command queuing (TCQ)
- Disconnect/reconnect
- Synchronous and asynchronous Ultra-SCSI

The adapter has three ports:

1. Internal 68-pin UltraSCSI port
2. Internal 50-pin SCSI-2 Fast port
3. External 68-pin UltraSCSI port with a high density 68-pin connector (see 17.12, "Choosing SCSI Cables" on page 175)

Notes:

1. Only two ports can be used at any one time.
2. Up to three PCI SCSI-2 Fast/Wide adapters can be used in a system.

This adapter is standard on some models of the PC Server 310.

17.7 PC Server 704 SCSI-2 F/W PCI RAID Adapter

This adapter (part #94G5884) is a high-performance SCSI-2 32-bit PCI bus master adapter with a built-in RAID controller that supports RAID levels 0, 1, and 5. It is based on the Mylex DAC960 RAID controller.

It has two channels and can support up to seven SCSI-2 Fast/Wide hot-swap drives per channel. It also has a 4 MB cache with battery backup capability.

The adapter has two internal and two external connectors (miniature high-density 68-pin), any two of which can be used simultaneously. All are SCSI-2 Fast/Wide connections (see 17.12, "Choosing SCSI Cables" on page 175).

Up to four RAID adapters can be installed in a server.

17.8 PC Server 320 Embedded SCSI Controller

The PC Server 320 PCI/MCA models have an embedded SCSI controller. This SCSI-2 Fast/Wide PCI controller supports asynchronous, synchronous, or fast synchronous SCSI modes with a peak data transfer rate of 20 MBps. It contains an integrated RISC processor that reduces the time it takes to process SCSI commands and reduces the CPU overhead. It supports advanced SCSI features such as:

- Multi-thread I/O
- Scatter/gather
- Tagged queuing
- Disconnect/reconnect

The interface characteristics of this controller are as follows:

- SCSI-2 Fast/Wide Interface
- Single channel
- Supports up to 15 SCSI-2 F/W devices
- One internal SCSI-2 Fast 50-pin connector
- One internal SCSI-2 Fast/Wide 68-pin connector
- One external SCSI-2 Fast/Wide 68-pin connector
- The external connector is attached to the rear of the system unit using one of the connector knockouts. This connector is a High Density 68-pin connector as defined in "SCSI-2 Fast/Wide External Connectors" on page 175.

17.9 PC Server 330 Embedded SCSI Controller

The PC Server 330 models have an embedded SCSI controller. This UltraSCSI Fast/Wide PCI controller supports asynchronous, synchronous, or fast synchronous SCSI modes with a peak data transfer rate of 40 MBps. It contains an integrated RISC processor that reduces the time it takes to process SCSI commands and reduces the CPU overhead. It supports advanced SCSI features such as:

- Multi-thread I/O
- Scatter/gather
- Tagged queueing
- Disconnect/reconnect

The interface characteristics of this controller are as follows:

- UltraSCSI Fast/Wide Interface
- Supports up to 15 SCSI-2 F/W devices
- One internal UltraSCSI Fast/Wide 68-pin connector
- One external UltraSCSI Fast/Wide 68-pin connector
- The external connector is attached to the rear of the system unit using one of the connector knockouts

17.10 ServeRAID SCSI Adapter

The IBM ServeRAID Adapter (part #70G8489) is a high-performance UltraSCSI 32-bit PCI bus adapter with a built-in RAID controller. The PCI bus interface has a 132 MBps peak transfer rate. The UltraSCSI interface supports a 40 MBps peak transfer rate.

The ServeRAID adapter incorporates many powerful features such as:

- High performance UltraSCSI that doubles the bandwidth over that of the SCSI-2 F/W interface
- Three UltraSCSI RAID channels
- Logical drive migration that allows you to make online capacity increases as well as change the RAID level of the logical drive
- New methods of storing RAID configuration information
- GUI management tools that allow you to easily manage your arrays

A RISC-based processor, together with microcode, serves as a dedicated RAID controller. Figure 55 shows the layout of the ServeRAID adapter.

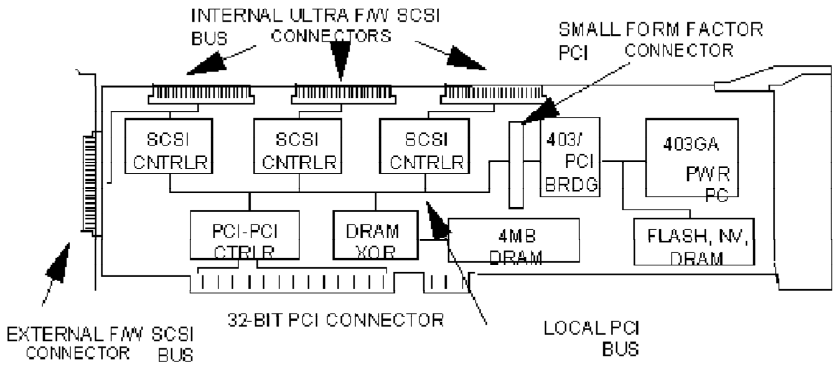


Figure 55. PCI ServeRAID Adapter

The ServeRAID adapter has a 4 MB write-through cache. It has three SCSI channels that can support up to 15 devices on each channel. These may also be used either internally or externally.

Interleave depth (stripe size) can be set to 8 KB, 16 KB, 32 KB or 64 KB. Applications such as multimedia and video will benefit from 32 KB or 64 KB interleave depth. Commercial applications that require many small transactions will do better with an 8 KB interleave depth.

The ServeRAID adapter supports RAID levels 0, 1, 5 and multiple hot spare drives. The following RAID configurations are supported.

- Up to 8 Arrays per adapter
- Up to 8 Logical drives per adapter
- Up to 16 disks per logical drive if stripe size is 8 KB or 16 KB
- Up to 8 disks per logical drive if stripe size is 32 KB or 64 KB
- An Array may span more that one channel
- One logical drive can span up to 16 drives attached to different channels

The ServeRAID stores critical RAID configuration data in three different places:

- FLASH EEPROM
- NVRAM
- Reserved area of hard disks

Up to four ServeRAID adapters can be installed in a server.

17.10.1.1 RAID Administration

The ServeRAID adapter ships with administration and monitoring tools that make the job of RAID administration much easier than in the past. For example, to change the RAID configuration with previous RAID adapters, you had to back up your data, delete the existing array, add or remove disks, define the new array and logical drives, and then restore your data.

The ServeRAID adapter allows you to add or remove disks from your array quickly and easily by simply adding or removing disks from the server, starting the online RAID administration tool and defining a new array. You can also change from RAID-0 to RAID-5 or from RAID-5 to RAID-0. (Refer to 5.5, “RAID Technology” on page 62 for information about RAID levels.)

These tools allow you to administer your RAID systems both locally and remotely through a client/server architecture that runs on top of TCP/IP. Locally, you can view the configuration or monitor the ServeRAID adapter on the following NOS environments:

- OS/2 Warp Server
- Windows NT
- Novell NetWare

Note: This is the currently supported list. More environments may be added in the future.

From your server, you can:

- Select the adapter that you want to monitor.
- View information about the states of your hard disks as well as change the device state.
- Replace a defunct drive.

In the client/server environment, server programs are currently available for the following NOS environments:

- OS/2 Warp Server
- Windows NT

- Novell NetWare
- SCO OpenServer

Note: The server programs run where the ServeRAID adapter is installed.

The client code runs on one of the following platforms:

- Windows NT Server (Versions 3.51 and 4.0)
- Windows NT Workstation (Versions 3.51 and 4.0)
- Windows 95

From the client, you can do many of the same functions that you can from the ServeRAID configuration program. These functions include:

- View the RAID configuration
- Create an array
- Create a logical drive
- Delete an array
- Replace a defunct drive
- Add hard disks to an existing array
- Change the RAID level of an array

However, the real power of these capabilities is that these tasks can be done to your server *while it is online*. For example, to add a new hard disk to your server, all you need to do is plug it in, start the online RAID utility and select **Detect New Drives** from the pull-down menu. The new drive that you installed will be shown in the RDY device state and you can then perform the desired function.

Also, if, previously, you defined a hot-spare (HSP) device, it had to be defined to a specific array. Now, a drive that is in the ready (RDY) state can be defined as a hot-spare for any array or even multiple arrays, even on different channels of the adapter.

17.10.1.2 Disk Migration

ServeRAID allows you to migrate hard disks from one machine to another quite easily. Because the critical array information is now stored in three different places, disk migration is straightforward. All you need to do is remove the disks from one server and then place them in another server.

Previously, all critical array information was held on the RAID adapter and the drive order (SCSI IDs) was critical for the array to be recognized. With the ServeRAID, it is not even necessary to maintain the same order when you migrate drives. The ServeRAID adapter will get the configuration information from the disks themselves.

17.10.1.3 Logical Drive Migration

The ServeRAID adapter supports the migration of logical drives from RAID level 0 to level 5 and from level 5 to level 0. This functionality can be quite useful in several situations. For example, some PC Server RAID models ship with only two hard disks in the standard configuration. Two drives allow you to create a RAID 0 array, but not, however, a RAID 5 array. By adding one disk to such a system you could then implement a RAID 5 array.

However, in the past, if you already had data on the RAID 0 logical drive, then migrating to a RAID 5 array was rather complicated. It involved a backing up of the data, a deleting the logical drive and the array, creating a new array at RAID level 5, creating a new logical drive, formatting from the NOS, and a restorating of your data.

ServeRAID makes this a much simpler process. With the ServeRAID adapter, you can migrate from RAID-0 to RAID-5 *while the system is running* and the ServeRAID adapter re-stripes the data across all the drives using the RAID-5 algorithm.

Notes:

1. In order to perform logical drive migration, all logical drives on the array to be migrated must be at the same RAID level.
2. The total number of all logical drives on the ServeRAID adapter must be less than seven in order to perform logical drive migration.
3. The logical drive on which this operation is being performed is not accessible during this operation. However, the system is still operational.

Figure 56 depicts a migration from RAID 0 to RAID 5. It should be noted that the usable storage capacity did not increase when the new drive was added due to the fact that RAID level 5 always consumes the equivalent of one disk in order to store the parity information. This is a standard function of RAID level 5 and not any limitation on the ServeRAID adapter.

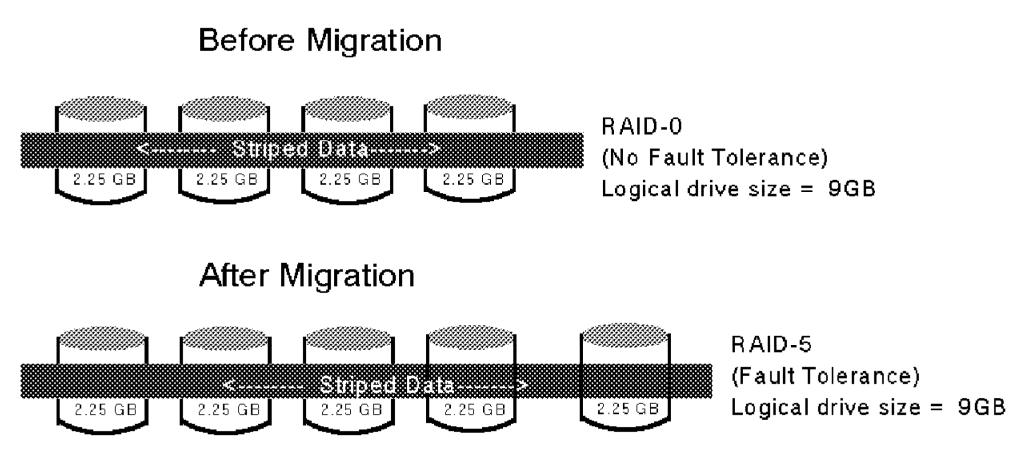


Figure 56. Migration from RAID-0 to RAID-5

You can also migrate from RAID-5 to RAID-0. In this case, the ServeRAID simply re-stripes data across the former number of drives minus one. You can then use the drive that has just been freed up for other uses in the machine.

Figure 57 on page 173 depicts a migration from RAID 5 to RAID 0.

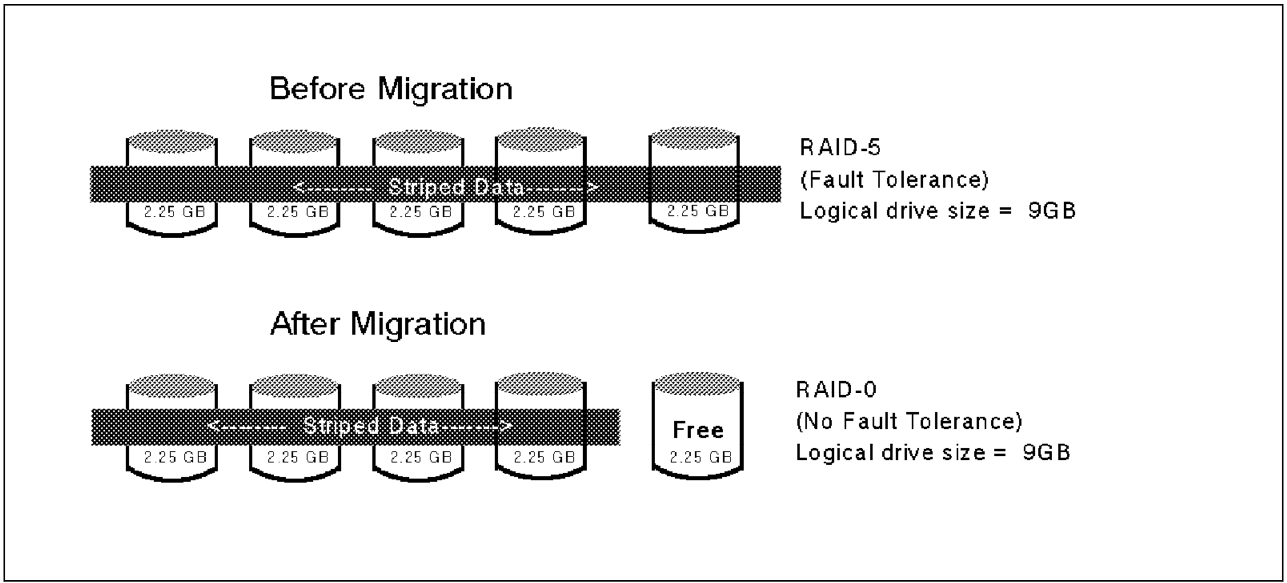


Figure 57. Migration from RAID-5 to RAID-0

Again, in order to do this migration, RAID levels must be the same on all logical drives in the affected array.

One potential ServeRAID function that has been demonstrated at trade shows and is being prepared as a future enhancement is the ability to add a hard disk to an existing RAID level 5 array. This allows you to increase the capacity of your RAID-5 array. While the original logical drive does not increase in capacity, space for additional logical drives on that same array is created. You can then use the NOS to span volumes across the multiple logical drives and start using the new disk space immediately.

Note: Depending on which operating system you are using you may still be required to restart the server to see the new disk space.

Figure 58 on page 174 depicts adding one drive to an array. Actually, up to three drives can be added at once.

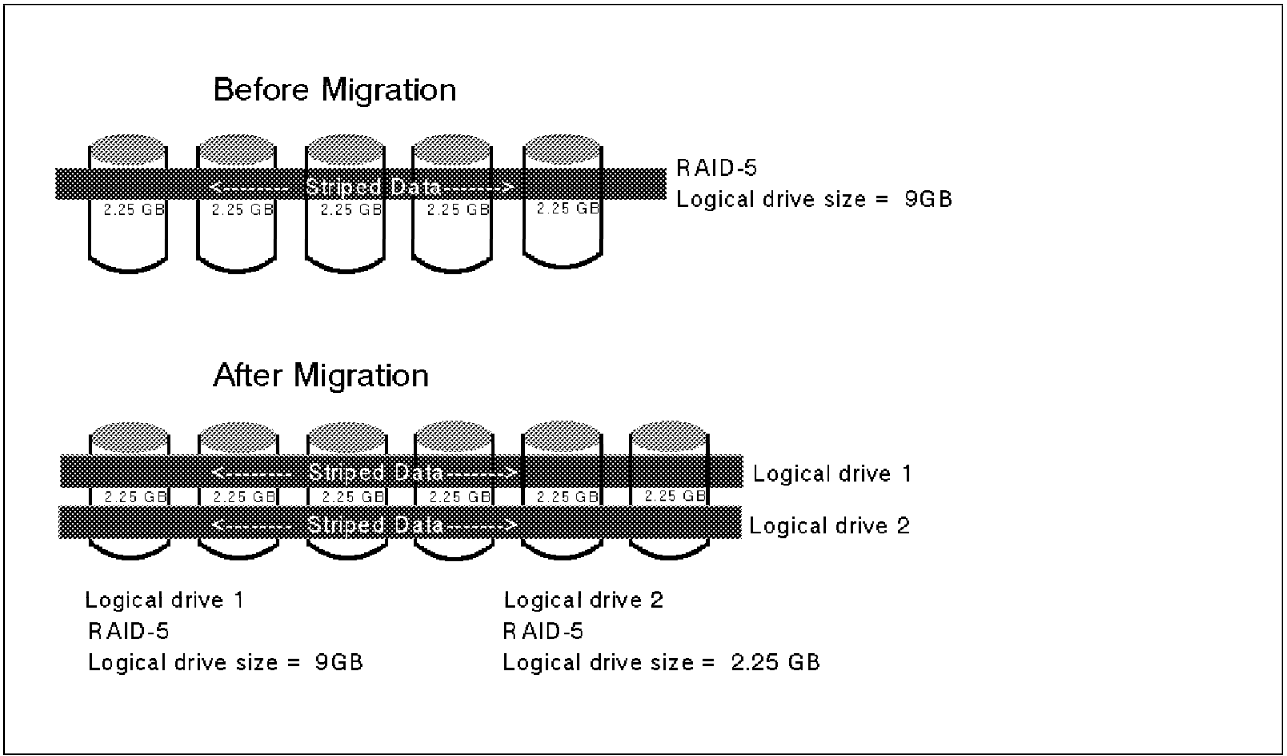


Figure 58. Addition of One Disk to a RAID-5 Array

17.11 SSA RAID Adapter

The IBM SSA RAID Adapter is a high-performance, high connectivity, enhanced-availability adapter that conforms to the industry-standard Serial Storage Architecture (SSA).

The SSA RAID Adapter supports:

- High performance storage attachment using the SSA interface
- Flexible attachment of multiple storage subsystems at distances of up to 25 meters
- Two SSA loops attaching up to 48 SSA drives in each loop
- RAID-0, 1, 5, and non-RAID operation supporting up to 32 RAID arrays
- Global hot-spare disk support

This multi-type RAID adapter can attach up to 96 SSA disk drives to PC Servers that have a PCI bus. Four serial ports, forming two SSA loops, can be configured in up to 32 arrays. Each array can be RAID-0, 1, 5, or non-RAID. Any combination of supported IBM SSA subsystems can be attached, provided no more than 48 disk drives are configured in one loop. Each SSA loop has a bandwidth of 80 MBps full-duplex using point-to-point copper cables up to 25 meters long.

Performance Specifications: The performance of the adapter takes full advantage of the SSA links. A single adapter can support up to 200 overlapped physical I/O requests concurrently.

It can execute up to 3,000 non-RAID I/O operations per second for a workload consisting of 4 KB block transfers with a 70/30 read/write ratio. When used with

RAID-5 arrays, the adapter will execute up to 1,000 I/O operations with the same workload.

The adapter can sustain up to 60 MBps when executing 128 KB read operations, 40 MBps for 128 KB writes in a non-RAID operation, and 35 MBps read and 7 MBps write for RAID-5 arrays.

Note

The performance numbers quoted are the hardware capabilities of the adapter. Actual performance will depend on a number of factors, including the operating system, file system, and the system processor.

Limitations:

- A RAID-1 array can only have two components per array
- The maximum number of drives per loop is 48
- The maximum number of drives per adapter is 96
- Each SSA loop must be connected to a valid pair of connectors on an SSA adapter card (connectors A1 and A2, or connectors B1 and B2)

The SSA adapter is currently supported in the IBM PC Server 320, 520 and 720 families. One adapter can be installed into the server 320 and 520 while two adapters can be installed in the Server 720.

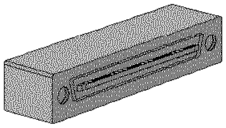
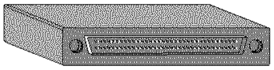
17.12 Choosing SCSI Cables

There are three types of SCSI-2 Fast/Wide external connectors currently being used in IBM servers. One type is a Centronics-style connector and is used only on Micro Channel adapters and system boards. The second type is used on systems equipped with PCI SCSI adapters and is a high-density D-shell style of connector. The third type is used on the PC Server 704 and is a miniaturized version of the connector used on the PCI adapters.

There are a variety of names that are used to describe each of these connectors. The connector used on Micro Channel devices is officially named the *IBM SCSI-2 P* connector but is sometimes referred to as a *Burndy* connector. The connector used on PCI adapters is an ANSI standard (actually defined in the SCSI-III specification) and is referred to as the *high density 68-pin* connector or sometimes as the *Honda* connector since Honda Corporation is one of its manufacturers. The connector used on the PC Server 704 is referred to as the *miniature high-density 68-pin* connector.

Also, there are two types of fasteners used on PCI-based adapter connectors. One uses screws and the other uses clips to attach the cable to the connector.

Table 27 on page 176 shows the IBM SCSI-2 P and the High Density 68-pin connector.

Table 27. SCSI-2 Fast/Wide External Connectors	
IBM SCSI-2 P Connector	**High Density 68-pin Connector**
	

It is important to select the correct cable for use with each of these connectors. Also, be sure to always use high-quality SCSI cables. Using inexpensive cables that are of poor quality is a sure way to invite problems on your system.

Chapter 18. Server Enclosures

This chapter describes options from IBM that make it easier for you to house your server systems and components. The following enclosures are discussed:

- 9306 PC Server Rack Enclosure
- 3517 SCSI Multi-Storage Enclosure
- 3518 PC Server Enterprise Expansion Enclosure
- 3510 SCSI Storage Enclosure
- 3527 SSA Entry Storage Subsystem

18.1 9306 PC Server Rack Enclosures

The IBM PC Server Rack product line consists of the IBM 9306 PC Server Rack Enclosure, available in two pre-defined, pre-assembled models and seven accessory options that allow a customer to customize the rack solution for their selected PC Servers. The IBM PC Server Rack Enclosure is designed to provide efficient space management for housing multiple PC Servers. It includes the capability to share a single monitor, keyboard, and pointing device across up to eight server systems. A 200-240 volt, 50/60 Hz, 16 amp power distribution unit is standard. The server system units and the console devices are not provided as part of this product.

There are two 24″ models of the 9306 Rack Enclosure:

1. 9306-4QS

   The 9306 Model 4QS is a 24-inch primary rack cabinet containing two dual-server slide tray shelves capable of supporting four PC Server 300, 320, 500, 520, 720, or PS/2 Server 85 or 95 system units. It includes side covers and console support accessories, consisting of a monitor compartment, keyboard slide tray, mouse pad, and multi-port server selector switch unit with control keypad.

2. 9306-4QX

   The 9306 Model 4QX is a 24-inch expansion rack cabinet containing two dual-server slide tray shelves capable of supporting four PC Server 300, 320, 500, 520, 720, or PS/2 Server 85 or 95 system units. It is designed for sharing the primary rack′s (model 4QS) console across additional server system units. The expansion rack attaches to the primary rack housing the console.

The 9306 is shown in Figure 59 on page 178.

Each model of the 9306 can hold up to four servers, with each server mounted on a sliding tray. There are two sliding trays in each of the two compartments, top and bottom, as shown in Figure 59 on page 178. The servers are attached to the trays via a locking plate to ensure stability. The locking plates replace the server′s existing pedestal. Currently, there are locking plates for the following servers:

- PS/2 Server 85, 95 and 95 Array (part #94G4995)
- PC Server 300, 320, 325 and 330 (part #94G4996)
- PC Server 500, 520 and 720 (part #94G4997)
- 3518 PC Server Enterprise Expansion Enclosure (part #94G4997)

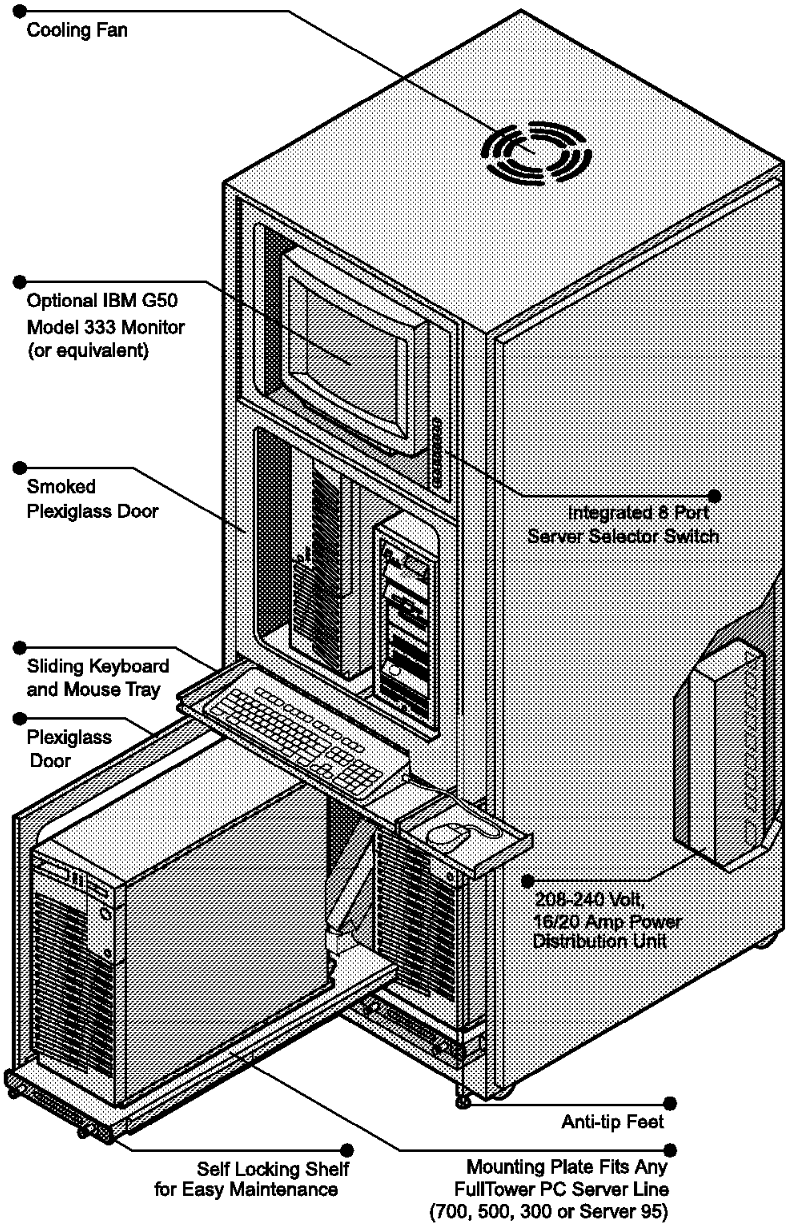


Figure 59. Exploded View of the IBM 9306 PC Server Rack Enclosure

Notes:

1. The 3518 is only supported with IBM PC Server SCSI-2 F/W Repeater and the required rack cable. (See 18.4, "PC Server SCSI-2 F/W Repeater" on page 185 for more information.)
2. The Server 310 and 704 are not supported in the 9306 Rack Enclosures.

The console in the primary cabinet (9306-4QS) can support up to eight servers by providing a multi-port server selector to share the monitor, keyboard and mouse among them. The expansion cabinet (9306-4QX) is designed to share the primary rack's console rather than using its own monitor, keyboard and mouse.

The necessary cabling for connecting the expansion cabinet to the primary cabinet is included with the expansion cabinet. One or two expansion cabinets can be used with each primary cabinet, although the maximum number of servers supported by the server selector is eight. The expansion cabinets are placed on either side of the primary cabinet.

The server selector is controlled by either a manual keypad located next to the monitor compartment or by a key sequence using the common keyboard. Illuminated buttons on the keypad indicate which server is currently using the console.

Notes:

1. The 9306 Rack Enclosures do not normally support ISO SVGA. If ISO-capable SVGA is required, the Enhanced Video Cable option (part #94G5458) should be ordered. This 15-foot coaxial video cable is recommended when using SVGA video modes.
2. The PS/2 Server 95, Server 500 and Server 720 may need a BIOS upgrade to function properly with the server selector:
   - Server 95: BIOS Revision 08 or later
   - Server 500: BIOS Revision 08 or later
   - Server 720: BIOS Revision 06 or later
3. The servers and the console devices (monitor, keyboard and mouse) are not provided as part of the rack enclosure.

The 9306 Rack Enclosures require a 200-240 V 50/60 HZ power supply and uses up to 20 amps on a single phase circuit. All powered devices (servers and monitors) must be switched to the high voltage range to support 200-240 V. If you have a PS/2 Server 95 or a PC Server 500, then the power supply is auto-sensing and there is no need to set any switches.

Note

Remember that the 220 V requirement for system components applies to the monitor as well.

The monitor used as the console must be less than the following dimensions. A special monitor stand is provided to be used with the IBM G41 and G50 monitors (which are supported in the 9306), in place of the standard tilt/swivel stand.

- Maximum height: 363 mm (14.3 in)
- Maximum width: 500 mm (19.7 in)
- Maximum depth: 424 mm (16.7 in)
- Maximum weight: 34 Kg (75 lb)

The 9306 supports the standard keyboard supplied with IBM servers.

The overall dimensions of the 9306 Rack Enclosure are:

- Height: 1937 mm (76.3 in)
- Width: 754 mm (29.7 in)
- Depth: 953 mm (37.5 in)
- Weight: 227 Kg (500 lb)

On its casters, the 9306 Rack Enclosure is capable of being rolled through a standard 80″ x 30″ door while fully loaded. As the rack is shipped almost fully assembled, you might require assistance with unloading.

18.2 3517 SCSI Multi-Storage Enclosure

The 3517 gives you the ability to have more than 22 GB of SCSI-2 Fast/Wide hot-swap storage in an external chassis. The enclosure is shown in Figure 60.

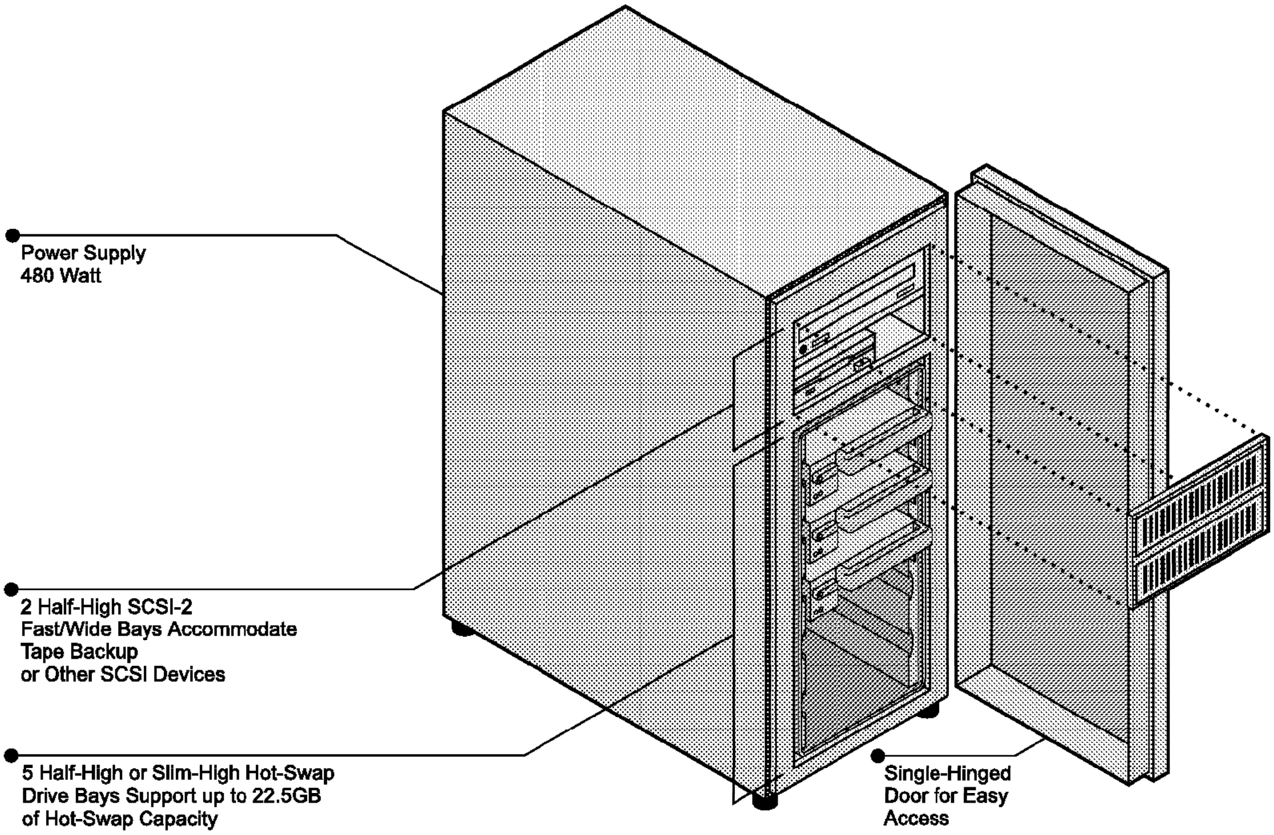


Figure 60. Exploded View of the IBM 3517 SCSI Multi-Storage Enclosure

The 3517 Multi-Storage Enclosure has the following configuration:

- Five half-high hot-swap drive bays
- Two 5.25″ half-high bays

The hot-swap drives installed in the 3517 Multi-Storage Enclosure must be SCSI-2 Fast/Wide drives with SCSI-2 Fast/Wide Hot-Swap Tray III. These trays can be ordered separately (part #70G9860) for use with compatible Fast/Wide drives, or drives can be ordered complete with the tray:

- 1.12 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9861
- 2.25 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9862
- 4.51 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9863

The 3517 enclosure connects to the SCSI-2 Fast/Wide external interface of the controlling server. Depending on the architecture of the server, the following cable is required to connect the 3517:

- For Micro Channel SCSI Adapters, such as those in the Server 500 and 720 RAID models, an external Fast/Wide card-to-option cable with a *Micro Channel* type connector is required. (Refer to “SCSI-2 Fast/Wide External Connectors” on page 183 for definitions of connector types.) Such cables include 1-meter cables, part #06H3231 or #66G5919.
- For PCI SCSI Adapters, such as those in the Server 310, 320, 520 (including PCI/MCA models) and 720 non-RAID models, an external Fast/Wide card-to-option cable with an *Industry Standard* type connector is required. (Refer to “SCSI-2 Fast/Wide External Connectors” on page 183 for definitions of connector types.) Such cables include 1-m cables, part #70G9857.

Note: The maximum external cable length supported is 1 m.

18.3 3518 PC Server Enterprise Expansion Enclosure

The 3518 Expansion Enclosure gives you the ability to have more than 40 GB of data storage externally connected to your server. The 3518 Expansion Enclosure uses the same chassis as the PC Server 720 and 520 and provides space for 18 hot-swap drives and two 5.25″ half-high drives. The enclosure is shown in Figure 61 on page 182. Refer to 14.1.3, “Expansion Bays” on page 141 for additional information on the storage bay configuration of the 3518 Expansion Enclosure.

The 3518 has a planar board attached to the side of the chassis (not shown in the diagram). This planar contains four 32-bit PCI slots, one 64-bit PCI slot and one 16-bit ISA slot. The PCI slots can be used for the optional PC Server SCSI-2 F/W Repeater cards. See 18.4, “PC Server SCSI-2 F/W Repeater” on page 185 for more information about this card.

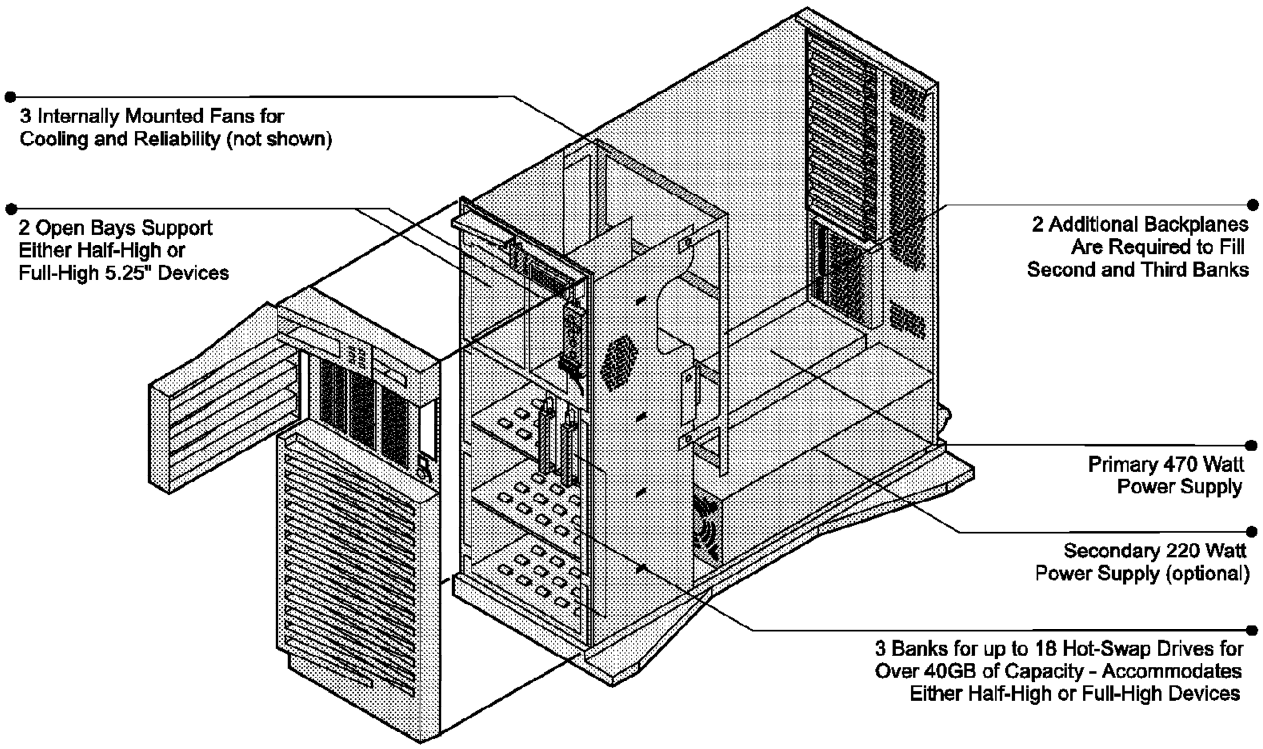


Figure 61. Exploded View of the IBM 3518 PC Server Enterprise Expansion Enclosure

18.3.1.1 Expansion Bays

The three hot-swap banks, C, D, and E, all can contain up to six Fast/Wide hot-swappable devices:

- Six slimline drives
- Three half-high
- Four slimline and one half-high
- Two slimline and two half-high

Bank C is already configured with a *hot-swap backplane* to allow connection of hot-swap drives into the system. The backplane has six special hot-swap connectors that allow hot-swap drives to be inserted and removed while the system is running.

Neither Bank D nor Bank E are usable as standard. To use either bank, the 220 W Additional Power Supply (part #70G9739) must be installed, and a backplane must be installed for each bank using the PC Server Hot-Swap Backplane III (part #70G9855). Refer to 18.3.1.2, “Cabling” on page 183 for information on cabling.

Alternatively, the PC Server 780 W Redundant Power Supply could be installed. This replaces the existing power supply and is able to support the whole system with all features. See 19.5, “PC Server 780 W Redundant Power Supply” on page 195 for more information.

Note: The additional power supply is sufficient to provide power to both Banks D and E, but a hot-swap backplane is required for each bank.

All drives installed in the 3518 enclosure must be SCSI-2 Fast/Wide drives with SCSI-2 Fast/Wide Hot-Swap Tray III. These new trays offer several advantages:

- An improved tray latch
- Drive activity and status LEDs mounted on front of tray
- Switch on front of tray to power the drive on and off
- Single 80-pin connector to back plane
- Will support Ultra SCSI devices

These trays can be ordered separately (part #70G9860) for use with compatible Fast/Wide drives, or drives can be ordered complete with the tray:

- 1.12 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9861
- 2.25 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9862
- 4.51 GB SCSI-2 Fast/Wide Hot-Swap Drive III, part #70G9863

18.3.1.2 Cabling

There is a variety of cabling options used to configure the 3518 enclosure, as shown in Table 28.

SCSI-2 Fast/Wide External Connectors

The connectors used on different IBM SCSI-2 Fast/Wide controllers are not all the same. Be sure to get the correct cable for your adapter. Refer to 17.12, “Choosing SCSI Cables” on page 175 for more information on these connectors.

Table 28 (Page 1 of 3). 3518 Enterprise Expansion Enclosure Cabling Options

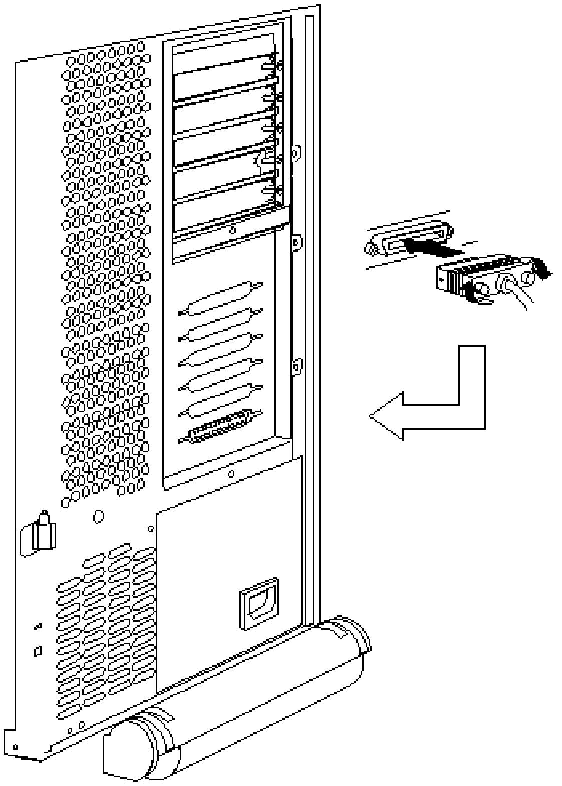
	**I. Server to Enclosure Cabling** The 3518 enclosure connects to one or more SCSI-2 Fast/Wide external interfaces of the controlling server. The connection on the enclosure is via one of the knock-outs at the rear of the unit as per the diagram to the left. Depending on the architecture of the server, the following cable is required to connect the 3518: • For Micro Channel SCSI Adapters, such as those in the Server 720 RAID array models with MCA adapter, an external Fast/Wide card-to-option cable with a *Micro Channel* type connector is required. Such cables include 1-meter cables, parts #06H3231 or #66G5919. • For PCI SCSI Adapters, such as those in the Server 310, 320, 520 (including PCI/MCA models) and 720 with PCI adapter, an external Fast/Wide card-to-option cable with an *Industry Standard* type connector is required. Such cables include 1m cable, part #70G9857. **Note:** The maximum supported length of cable is 1 m except when using the SCSI-2 F/W Repeater. See 18.4, “PC Server SCSI-2 F/W Repeater” on page 185.

Table 28 (Page 2 of 3). 3518 Enterprise Expansion Enclosure Cabling Options

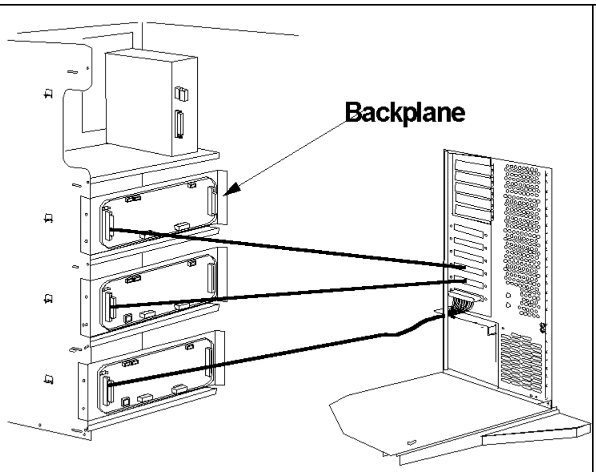
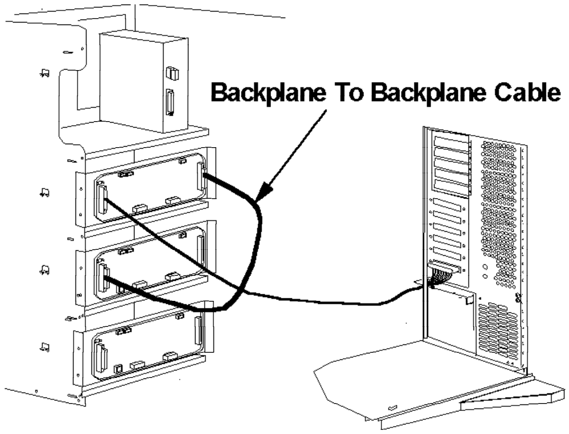
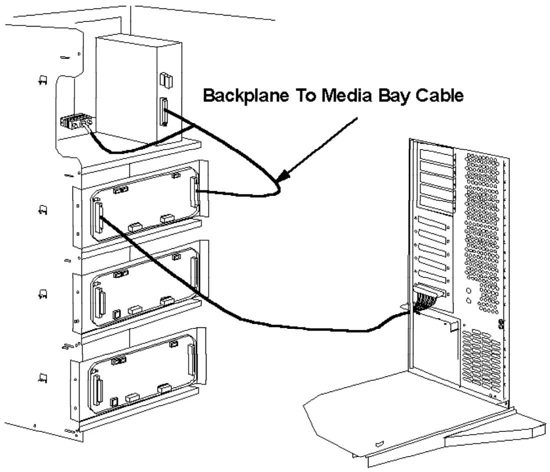
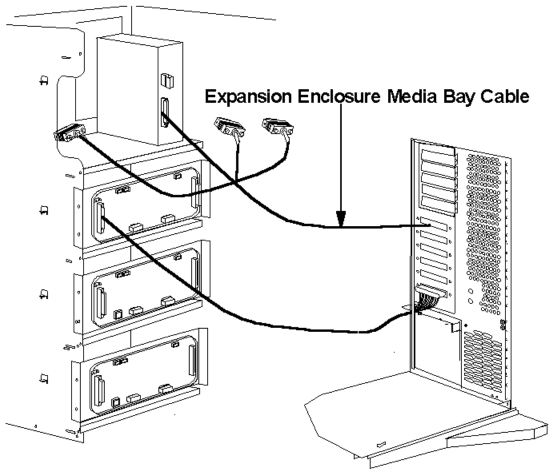
 Exploded internal view	**II. Expansion Enclosure Backplane Cable (70G9876)** This backplane cable provides the internal connection between the server subsystem and the backplanes of the enclosure. One backplane cable comes installed standard and is connected to the backplane in Bank C. For Banks D and E, if the bank is to be cabled to a separate SCSI channel on the server, a separate backplane cable is required. If the bank is to be daisy-chained, the Backplane To Backplane cable is used.
 Exploded internal view	**III. Backplane To Backplane Cable (94G4070)** The Backplane To Backplane Cable allows daisy-chaining of two backplanes on one SCSI channel. Without the SCSI-2 F/W Repeater, this cable is only supported when using half-high drives such as the 4.5 GB drive. In this scenario, a user is able to install 6 half-high drives on a single SCSI channel, with three drives installed in each bank. See 18.4, “PC Server SCSI-2 F/W Repeater” on page 185 for more information about cabling with the SCSI-2 F/W Repeater card.
 Exploded internal view	**IV. Backplane To Media Bay Cable (70G9864)** The Backplane To Media Bay Cable is used to connect SCSI devices installed in the upper two media bays (B1 and B2) to the SCSI channel of a backplane. This cable has two Fast/Wide connectors and can be used to connect two devices. The cable may be connected to backplanes installed in either the C or D banks, but not to Bank E. If the two devices in the media bays are to be installed on separate channels, two cables are required. One 16-bit to 8-bit converter is provided with the cable to enable connection of an SCSI or SCSI-2 Fast device. If a second 8-bit device requires connection, a second converter (part #32G3925) should be ordered.

Table 28 (Page 3 of 3). 3518 Enterprise Expansion Enclosure Cabling Options	
 Exploded internal view	**V. Expansion Enclosure Media Bay Cable (70G9877)** The Expansion Enclosure Media Bay Cable connects an independent SCSI channel to devices installed in the upper media bays of the enclosure. The cable has four SCSI drops, but only the use of two of them are supported. One 16-bit to 8-bit converter is provided with the cable to enable connection of an SCSI or SCSI-2 Fast device. If a second 8-bit device requires connection, a second converter (part #32G3925) should be ordered.

18.4 PC Server SCSI-2 F/W Repeater

The IBM PC Server SCSI-2 F/W Repeater (Part #94G5565) allows longer signal cables to be used between a PC Server system unit and an external expansion enclosure. It is designed for the 3518 Enterprise Expansion Enclosure and when installed, the repeater adds the following capabilities:

- Longer signal cables are allowed between the server unit and 3518.
- The 3518 can be installed in a rack enclosure.
- A larger number of 3518s may be attached to a single system unit.
- An increased number of SCSI devices are supported on a single channel.

The PC Server SCSI-2 F/W Repeater supports the SCSI-2 F/W operation at 20 MBps. The UltraSCSI mode (SCSI-III) is not supported.

The Repeater Card fits into the previously reserved PCI slots of the 3518 PC Server Enterprise Storage Enclosure. Up to four PC Server SCSI-2 F/W Repeaters are supported in one 3518. One SCSI repeater card is required per SCSI channel attached to the 3518.

18.4.1 New Capabilities

The introduction of the PC Server SCSI-2 F/W Repeater has enabled several new capabilities in using 3518s.

- Due to the required cable lengths in a 9306 rack environment, the 3518 PC Server Enterprise Enclosure was not previously supported in a 9306 PC Rack Enclosure. The PC Server SCSI-2 F/W Repeater, when installed in the 3518, allows the 3518 to operate in the rack, even with the longer cables. The PC Server 4.3 m SCSI-2 F/W Rack Cable allows the 3518 to be attached to any server located in the same or adjacent PC Server Rack.
- Without an SCSI repeater, daisy chaining of backplanes in the 3518 was limited to two backplanes and then it could only be supported if half-high

disk drives were used. By using the SCSI-2 F/W repeater along with the PC ServeRAID Adapter, more drives are supported as follows:

- You can attach up to two backplanes when using slimline Fast/Wide SCSI-2 hot-swap drives.
- You can attach up to three backplanes when using half-high Fast/Wide SCSI-2 hot-swap drives.
- There may be up to 12 slimline drives (two backplanes with six drives each) attached to one single SCSI channel or 9 half-high drives (three backplanes with three drives each).

18.4.2 Cabling

The following cables are used with the SCSI repeater card:

- PC Server 4.3 m SCSI-2 F/W Rack Cable, Part #94G5566
- PC Server 3.0 m SCSI-2 F/W Cable, Part #94G5567
- SCSI-2 Backplane to Backplane Cable, Part #94G4070
- PC Server 68-Pin SCSI-2 F/W Converter, Part #94G5569

The 4.3-meter cable is used to install a 3518 in a 9306 PC Server Rack Enclosure.

The 3.0-meter cable is used to increase the possible distance between a system unit and a stand-alone 3518.

The SCSI-2 Backplane to Backplane Cable is required for each daisy-chained backplane.

18.4.3 Compatibility

The PC Server SCSI-2 F/W Repeater can be used with the following SCSI adapters:

- IBM SCSI-2 Fast/Wide Adapter/A
- IBM SCSI-2 Fast/Wide Streaming-RAID Adapter/A
- PC Server SCSI-2 Fast/Wide PCI Adapter
- PC Server SCSI-2 Fast/Wide PCI-Bus RAID Adapter
- PC Server SCSI-2 Fast/Wide PCI Adapter II
- PC ServeRAID Adapter

See Chapter 17, “Storage Controllers” on page 163 for more information and part numbers of these adapters.

18.5 3510 SCSI Storage Enclosure

The 3510 SCSI Storage Enclosure is a one-bay external enclosure that can house a slimline (1″) and half-high (1.6″) SCSI device. Both 3.5″ and 5.25″ form-factors are supported. A standard 32 W universal power supply provides sufficient power for the device. Features include cover key lock assembly, mounting holes in the base for bolt down, external audio connection capability, and external SCSI ID selector switches.

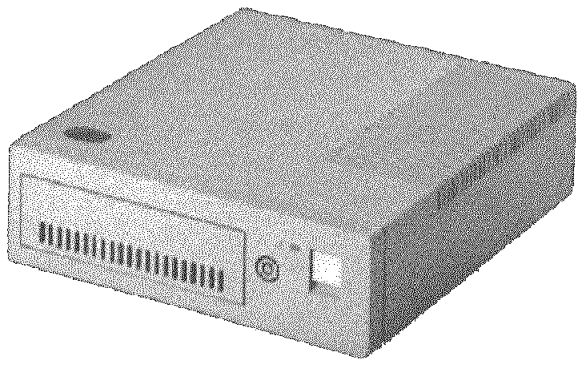


Figure 62. The IBM 3510 SCSI Storage Enclosure

The unit has two 50-pin (8-bit) external SCSI option connectors to enable the connection of SCSI or SCSI-2 Fast adapters and to daisy-chain to other external enclosures. If SCSI-2 Fast/Wide devices are to be installed in the 3510 enclosure, the SCSI-2 Upgrade Kit, part #32G3920, should be ordered.

The 3510 is non-terminating, so if the device is at the end of an SCSI chain, it requires either an 8-bit active terminator (part #32G3919) for SCSI or SCSI-2 Fast devices. If the SCSI-2 Upgrade Kit has been installed, a 16-bit active terminator (part #32G3918) is required.

Note: The installation of some hard drives may require an additional installation kit, the 3510 storage enclosure hard disk kit B, part #94G2648.

18.6 3527 SSA Entry Storage Subsystem

The IBM 3527 SSA Entry Storage Subsystem is a stand-alone storage enclosure for IBM′s Serial Storage Architecture (SSA) hard disk drives. The 3527 provides a cost-effective storage expansion option for external disk storage with all the performance, scalability, and expandability of the SSA interface architecture.

The subsystem takes up a small footprint and can be placed up to 25 meters from the host server. Expansion is easily accomplished by adding an additional subsystem in the SSA loop, up to a total of 48 disk drives per loop. With SSA architecture, you can now grow to very large storage capacities without the distance and bus limitations of SCSI storage products.

The IBM SSA Entry Storage Subsystem provides:

- Cost-effective storage enclosure for SSA devices
- Five optional, hot-swap disk drive bays for IBM SSA hard disk drives
- Up to 22.5 GB of hard disk capacity
- Lockable door
- SSA interface, providing 25 meters between adapters and subsystems

18.6.1 SSA Hard Disk Drives

There are currently two hard disk drives supported in the 3527 storage enclosure. They are the 2.25 GB and 4.51 GB Hot-Swap SSA disk drives. Table 29 shows the features of these drives.

Table 29. SSA Hard Disk Drives

Part Number	**05J6413**	**05J6414**
Interface	SSA	SSA
Capacity	2.25 MB	4.51 GB
Average Seek Time	8.0 ms	8.0 ms
Burst Transfer Rate	20.0 MBps	20.0 MBps
Sustained Transfer Rate	5.5 to 7.4 MBps	5.5 to 7.4 MBps
Drive Rotation Speed	7200 RPM	7200 RPM
Form Factor	3.5 x 1 ″	3.5 x 1.6 ″

18.6.2 3527 Options

The associated IBM options for use with the 3527 are:

- IBM SSA RAID Adapter for PC Servers, part #32H3811
- 1.0 meter SSA external cable pair, part #59H7220
- 2.5 meters SSA external cable pair, part #59H7221
- 5.0 meters SSA external cable pair, part #59H7222
- 10.0 meters SSA external cable pair, part #59H7223
- 25.0 meters SSA external cable pair, part #59H7224
- SSA Dummy Disk Drive Module, part #05J6411

Two cables are packaged with each part number. This allows the SSA loop to be implemented.

The 3527 can support a maximum of five SSA hot-swap disk drives. If less than five disk drives are used, SSA dummy disk drive modules must be inserted into the empty bays to maintain continuity of the SSA loop. A maximum of three neighboring dummy disk drive modules can be connected in a particular SSA loop. This means that the minimum configuration of a 3527 consists of two disk drives and three SSA dummy disk drive modules.

Refer to 5.3, “SSA Technology” on page 52 for more information on SSA technology.

Chapter 19. PC Server Options

IBM has a full range of options for PC Servers. This chapter describes the following five:

1. IBM 4/10 GB Internal SCSI DAT Tape Drive
2. IBM 24/48 GB Internal Tape Autoloader
3. IBM TR4 SCSI Tape Drive
4. IBM PC Server Advanced Systems Management Adapter
5. IBM PC Server 780 W Redundant Power Supply

A good place to get more information about options that can be used with IBM PC Servers is the IBM PC Server Compatibility web site. This site has information about IBM and non-IBM options (please see Appendix A, "Sources of Drivers and Information" on page 203).

19.1 4/10 GB Internal SCSI DAT Tape Drive

The IBM 4/10 GB 4 mm DAT drive has a 4 GB native capacity with hardware compression that typically enables 10 GB of data storage. The data transfer rate ranges from 400 KBps through 1600 KBps with hardware compression, with a typical transfer rate of 1000 KBps.

The drive is a half-high device (1.6″) and will fit in most half-high drive bays including a variety of externally attached storage enclosures. Installation into some systems may require the use of an attachment kit. The 4/10 GB DAT drive is available in a 5.25″ form factor, part #87G1469, or 3.5″ form factor, part #87G1470. If configuring this device for the PS/2 model 95A, use the 3.5″ drive.

The drive has an SCSI-2 Fast 50-pin interface and will require the SCSI wide-to-narrow converter for connection to an SCSI-2 Fast/Wide adapter.

Included with the drive are the following:

- NovaBack backup software (DOS, Windows, OS/2)
- Data tape
- Self-diagnostic tape
- Cleaning tape
- Installation instructions

Note: A Fast/Wide converter is not included with the drive.

19.2 24/48 GB Internal Tape Autoloader

This six DAT cartridge autoloader fits in the 5.25″ full-high (3.2″) bay on the PC Server 720 and 500 (bay B with the bracket between B1 and B2 removed) and the Server 320 non-array model (using two adjoining 5.25″ half-high bays). The autoloader is designed to fit entirely within the 5.25″ full-high bay, thus allowing the server's cover to close while the autoloader is in use. The drive can operate either horizontally or vertically. If it is mounted vertically, the control panel can be rotated to maintain usability of the controls.

This is a DDS-2 DAT drive capable of backing up 48 GB of data using compression on six cartridges. It has a typical sustained transfer rate of 1 MBps with compression.

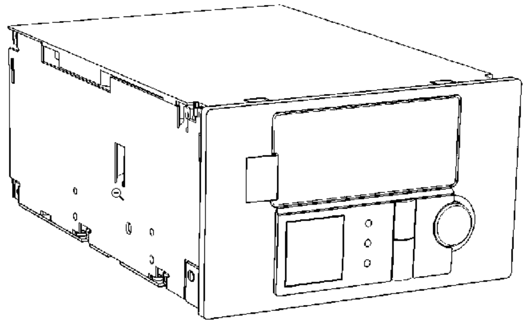


Figure 63. The IBM 24/48 GB Internal Tape Autoloader

The autoloader option kit contains:

- 24/48 GB autoloader
- SCSI wide-to-narrow (68-50 pin) converter
- Power cable
- One 6-cartridge magazine
- Five data cartridges
- One cleaning cartridge
- User's manual

Notes:

1. The autoloader is an SCSI-2 Fast device. It has a 50-pin SCSI connector and includes a converter to connect the device to an SCSI-2 Fast/Wide adapter.
2. The PC Server 500 requires BIOS level 4 or higher.
3. The autoloader will require either one SCSI ID and one or two PUNs depending on the mode of operation. See 19.2.1.2, "Operating Modes" for more information.

19.2.1.1 Tape Data Capacity

The native (uncompressed) capacities of the DAT cartridges are dependent on the physical length of the tape. Refer to Table 30.

Table 30. DDS and DDS2 Tape Capacities

Length	Format	Native Capacity
60 m	DDS	1.3 GB typical
90 m	DDS	2.0 GB typical
120 m	DDS2	4.0 GB typical

Note: The 60 m and 90 m cartridges are automatically written in DDS format, and only 120 m cartridges are written in DDS2 format.

19.2.1.2 Operating Modes

The autoloader operates in two modes: sequential and random. The random mode is the mode of operation used by most backup software.

1. Sequential

   In sequential mode, the autoloader does not depend on host implementation of changer commands. You can select which cartridge you want by using the SELECT and LOAD buttons on the front panel. When a cartridge is ejected from the drive following a host UNLOAD command, the autoloader will automatically load the next available cartridge from the magazine into the drive.

2. Random (the usual mode of operation)

   In random mode, the host views the autoloader as two physical devices. It uses two physical unit numbers (PUNs), 0 and 1 of the SCSI ID:

   a. A changer mechanism that accepts SCSI medium changer commands.
   b. A tape drive that accepts SCSI sequential access commands.

   The controlling host computer therefore has full random access to any cartridge. Most PC backup software automatically loads, unloads, and cleans the autoloader using this mode.

19.2.1.3 Controls and Indicators

The autoloader has the following controls and indicators as shown in Figure 64.

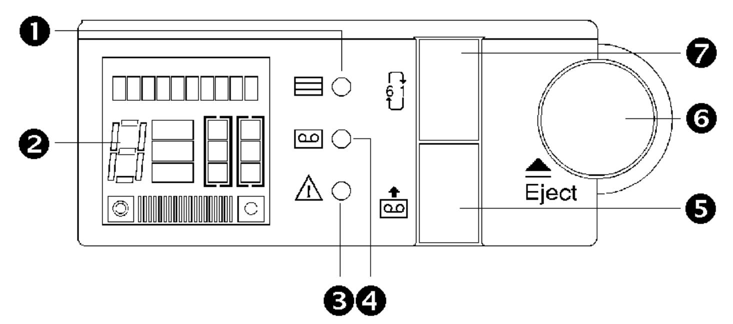


Figure 64. Control Panel for the 24/48 GB Autoloader

1. Magazine Present LED
2. LCD Display - shows the following information:
   - Message line showing status
   - Cartridge number currently selected
   - *DC* indicating that data compression is active
   - *WP* indicating the cartridge is write-protected
   - Six boxes showing which slots in the magazine have a cartridge
   - Bar showing how much tape has been used
3. Operator Attention Required LED
4. Tape Activity LED
5. Load Tape Button - loads the selected cartridge
6. Eject Button - unloads the cartridge and ejects the magazine from the unit
7. Select Button - selects a cartridge from the magazine

19.3 TR4 SCSI Tape Drive

The IBM TR4 SCSI Tape Drive provides a cost-effective tape backup solution for mid-range to high-end servers. This product is the third tier of the three-product Mini-QIC tape family. It is positioned above the DUALSTOR 800 and TR3 products as a high-performance, high-capacity tape backup product.

With a 4.0 GB native capacity, this tape drive can store up to 8.0 GB using software-controlled data compression offering a backup rate of up to 30 MB/min in native operation. This rate can be increased with software-controlled data compression to as much as 60 MB/min. The drive uses the new TR4 tape cartridges from 3M in the QIC-3095 tape format.

The TR4 drive is supplied as an internal and an external device. The corresponding part numbers are:

- TR4 Internal Tape Drive - 06H9716
- TR4 External Tape Drive - 3502900

Figure 65 shows the internal unit.

The external drives can be mounted either horizontally or vertically without affecting the drive operation.

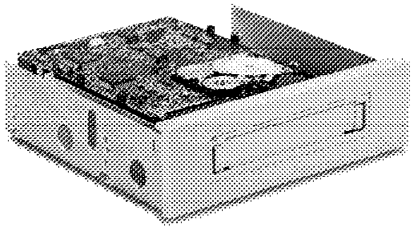


Figure 65. The IBM TR4 Internal Tape Drive

Physical Specifications

- Approximate Height: 76.2 mm (3.0 in.)
- Approximate Width: 238.8 mm (9.4 in.)
- Approximate Depth: 279.4 mm (11.0 in.)

Operating Environment

- Temperature: 5 to 45 °C (41 to 113 °F)
- Relative Humidity: 20 to 80 percent RH, non-condensing

The TR4 is supported in all IBM PC Servers equipped with an SCSI controller. It is not supported for use with RAID SCSI control devices. It is intended for use with non-RAID SCSI controllers only.

One of the following cables is used to connect the TR4 External SCSI-2 Tape Drive to an SCSI-2 host adapter:

- SCSI-2 Card to SCSI-2 Option (part #32G4089)
- SCSI Option to SCSI Option (part #6451042)
- SCSI Option to Single Port Option (part #31F4186)

Note: Use of this device with a wide SCSI adapter will require the installation of an SCSI wide-to-narrow cable converter.

The following application programs have been tested and found to be compatible with the 4.0/8.0 GB TR4 Internal SCSI Tape Drive:

- Novastor Novaback v2.1 for OS/2
- Back Again/2 v3.1 for OS/2
- Colorado Backup EXEC v1.1 for WIN '95
- Colorado Backup EXEC v3.01 for DOS v6.22
- Windows NT v3.51 native tape support

- Cheyenne ArcServe v2.0 for Windows NT
- Novell NetWare v3.12 native Sbackup
- Novell NetWare v4.1 native Sbackup
- Cheyenne ArcServe v6.0 for Novell NetWare v3.12
- Cheyenne ArcServe v6.0 for Novell NetWare v4.1
- Arcada Backup EXEC for NT, Single Server Edition, v6.0
- Arcada Backup EXEC for NetWare, Single Server Edition, v7.0
- Arcada Sytos Premium for OS/2, v2.x

The latter three programs are shipped on the CD-ROM included in the option kit.

Note: Please note that hardware device support is continuously updated by OEM software suppliers. Check with IBM if your software is not listed.

Cleaning cartridges can be obtained from your local Lexmark dealer (part #GEN100-1).

19.4 PC Server Advanced Systems Management Adapter

The IBM PC Server Advanced Systems Management Adapter (part #94G5570) is an full-length Industry Standard Architecture (ISA) card designed to provide comprehensive systems management capability to many IBM PC Server products.

The PC Server Advanced Systems Management Adapter integrates fully with TME 10 NetFinity to provide both local and remote management of the PC Server. The features of this adapter include:

- Exploitation of the I2C bus and other features of PC Server 325 system for additional systems management functions such as power on and off without a separate power unit option, system reset, and monitoring of additional temperature sensors
- Dial in and dial out capability through a modem even when the server is down using the optional PC Server Advanced Systems Management Power Unit
- Remote detection of certain hardware problems
- Auto restart and notification if the system does not respond
- Monitoring of internal server voltage and temperature

These features are described more fully in the sections below.

19.4.1 Remote Management

The PC Server Advanced Systems Management Adapter and related options, in conjunction with TME 10 NetFinity, lets you manage your LAN remotely so that you can:

- Dial in to the systems management card even when the system is down to:
  - Reset system
  - Browse a log of events detected by the service processor
  - Reconfigure the adapter to alert another source for problem resolution
  - Check voltages and temperature
  - Control system power
- Dial out to a pager or NetFinity Manager to alert the system administrator if an error is detected

- Detect and determine the cause of certain hardware problems remotely

19.4.2 Local Management

The PC Server Advanced Systems Management Adapter works with TME 10 NetFinity to help you manage your server locally. With this combination, you can:

- Monitor internal server voltages and temperature
- Control system power
- Restart the system
- Generate a notification to the system administrator if the server does not respond

You can use the adapter to power on and power off the machine. Obviously, you will need to continue to supply power to the PC Server Advanced Systems Management Adapter if you want to be able to bring the system back on line. The PC Server Advanced Systems Management Power Unit (part # 94G5571) is available for systems that do not have an internal continuous power feature. This option provides continuous power even if the system is powered off or is down due to a mechanical malfunction. PC Server 325 models do not require this option as it is provided within these specific servers.

The PC Server Advanced Systems Management Adapter card has an I2C bus with master I2C controller and can be connected to slaves through a planar connector or cables. The adapter has two 9-pin, D-shell, serial ports one of which supports attachment of a modem for dial-in/dial-out functions. Temperature and voltage sensors on the card provide a means to determine if critical thresholds set by using TME 10 NetFinity are within system tolerance.

The PC Server Advanced Systems Management Adapter is supported in all IBM PC Servers with an ISA or EISA bus architecture. You will also need a modem to support the dial in and dial out capabilities of the adapter.

In addition, you will also need TME 10 NetFinity installed and a supported network operating system such as OS/2 (R), Windows NT or NetWare.

The PC Server Systems Management Cable (part #94G6970) provides support for PC Server Advanced Systems Management Adapter and the PC Server 704. This cable is attached from the port on the PC Server 704 planar to the PC Server Advanced Systems Management Adapter connector. This option is required on PC Server 704 models. Cables to support PC Server 325 are provided with the PC Server Advanced Systems Management Adapter option.

Note

The PC Server 704 requires both the PC Server Advanced Systems Management Power Unit and PC Server Systems Management Cable to be installed for proper operation. Installation of the PC Server Systems Management Cable without the use of the PC Serve Advanced Systems Management Power Unit will cause the server to not power on properly.

19.5 PC Server 780 W Redundant Power Supply

The IBM PC Server 780 W Redundant Power Supply is a self-contained, redundant power supply for use in PC Server 500, 520, and 720 systems as well as the PC Server Enterprise Expansion Enclosure (IBM 3518). It allows these systems to continue functioning even if a power supply component fails. It is targeted for high-availability applications associated with mid-range to high-end server environments.

When installing the unit, the 470 watt primary power supply is removed and the redundant unit then installs and occupies the space normally used by the primary supply as well as the 220 watt auxiliary supply. Because it has a 780 watt capacity, the auxiliary supply is not needed even if the additional hot-swap drive banks are populated. The redundant power supply is comprised of two independent power supplies housed in one self-contained unit.

Depending upon the system in which it is installed, there are different ways in which the system notifications can occur. When installed in a PC Server 720, NetFinity is used to alert the system administrator if a temperature threshold is reached or the redundancy feature has been activated.

PC Server 500s, 520s, and the 3518 use the power good LED to indicate that a power supply module has failed. The power good LED blinks to notify the system administrator that redundancy has been activated. Neither the setting of temperature thresholds and alerts, nor receiving redundancy alerts through NetFinity are supported with these products.

The redundant power supply can be used in conjunction with an American Power Conversion Smart-UPS and NetFinity to provide a complete, power subsystem and systems management solution geared for high-availability PC Server applications.

The list below provides a summary of the product features:

- 780 watt, worldwide, voltage-sensing power supply
- Self-contained single unit, eliminates need for auxiliary power supply
- NetFinity alerts in PC Server 720
- Temperature-sensing fan

Chapter 20. SCSI Disk Drives

IBM offers a variety of SCSI hard disk drives either as standard drives in servers or as options. The following categories are tabulated here:

- SCSI-2 Fast
- SCSI-2 Fast/Wide
- SCSI-2 Hot Pluggable
- UltraSCSI

IBM also offers SCSI-2 Fast/Wide Differential drives, which allow SCSI cables of lengths over 6 meters and up to 25 meters. Differential drives are available in capacities of 2.25 GB, 4.51 GB and 5.31 GB and have the same characteristics as the SCSI-2 Fast/Wide equivalent drives.

A drive should be selected on its specifications: average access, latency, disk rotation as well as burst and sustainable data transfer rates.

It is also important to match the performance of the drive with that of the disk controller such that the relative slowness of the media is less of an impact on the overall system performance. SCSI-2 Fast/Wide drives should always be matched with Fast/Wide controllers.

20.1 SCSI-2 Fast Drives

Table 31 shows the currently offered SCSI-2 fast drives.

Table 31. SCSI-2 Fast Drives

Capacity	**1.27 GB**	**2.16 GB**	**2.25 GB**	**4.51 GB**
Part Number	75H8974	07H1124	94G3054	94G3196
Rotation Speed	5400 RPM	5400 RPM	7200 RPM	7200 RPM
Sustained Transfer		3.8-5.7 MBps	5.5-7.4 MBps	5.5-7.4 MBps
Burst Transfer (MBps)	10.0	10.0	10.0-12.6	10.0-12.6
Average Seek	11 ms	8.5 ms	7.5 ms	8.0 ms
Platters	2	3	4	8
Average Latency	5.6 ms	5.6 ms	4.2 ms	4.2 ms
Buffer (Adaptive)	128 KB	512 KB	512 KB	512 KB
Form Factor	3½″ slimline	3½″ slimline	3½″ slimline	3½″ half-high

Notes:

1. A slimline drive is 1″ high.
2. A half-high drive is 1.63″ high.
3. Sustained transfer rate is a measure of the typical time used to transfer data between the media and the drive's data buffer.

These internal drives can be used with an SCSI-2 Fast or Fast/Wide controller with the appropriate cabling. As these drives have a 50-pin connector, one of the following is required:

- 50-pin to 50-pin internal cable (for a Fast adapter)
- 68-pin to 68-pin internal cable (for a Fast/Wide adapter using a 68-to-50-pin converter)

The transfer rates of SCSI-2 Fast drives are limited to the 10 MBps by the SCSI interface.

20.2 SCSI-2 Fast/Wide Drives

IBM currently offers the following SCSI-2 Fast/Wide drives.

Table 32. SCSI-2 Fast/Wide Drives

Capacity	**1.12 GB**	**1.12 GB**	**2.25 GB**	**4.51 GB**
Part Number	94G3052	07H0386	94G3055	94G3057
Rotation Speed	7200 RPM	7200 RPM	7200 RPM	7200 RPM
Sustained Transfer	5.5-7.4 MBps	5.5-7.4 MBps	5.5-7.4 MBps	5.5-7.4 MBps
Burst Transfer (MBps)	20 MBps	10.0-12.6	10.0-12.6	20 MBps
Average Seek	6.9 ms	6.9 ms	7.5 ms	8.0 ms
Platters	2	2	4	8
Average Latency	4.2 ms	4.2 ms	6.7 ms	4.2 ms
Buffer (Adaptive)	512 KB	512 KB	512 KB	512 KB
Form Factor	3½″ slimline	3½″ slimline	3½″ slimline	3½″ half-high

Notes:

1. A slimline drive is 1″ high.
2. A half-high drive is 1.63″ high.
3. Sustained transfer rate is a measure of the typical time used to transfer data between the media and the drive′s data buffer.

20.3 SCSI-2 Fast/Wide Hot-swap Drives

IBM currently offers the following SCSI-2 Fast/Wide hot-swap drives.

Table 33. SCSI-2 Fast/Wide Hot-Swap Drives

Capacity	**1.12 GB HSIII**	**2.14 GB (704)**	**2.25 GB HSIII**	**4.41 GB HSIII**	**4.51 GB (704)**
Part Number	70G9861	94G5882	70G9862	70G9863	94G6679
Rotation Speed	7200	7200	7200	7200	7200
Sustained Transfer	4.9-8.2 MBps		4.9-8.2 MBps		
Burst Transfer (MBps)	20 MBps	20 MBps	20 MBps	20 MBps	20 MBps
Average Seek	8.0 ms	7.0 ms	8.0 ms	8.5 ms	8.0 ms
Average Latency	4.2 ms	4.2 ms	4.2 ms	4.2 ms	4.2 ms
Buffer (Adaptive)	512 KB	512 KB	512 KB	512 KB	
Form Factor	3½″ slimline	3½″ half-high	3½″ half-high	3½″ half-high	3½″ half-high

Notes:

1. A slimline drive is 1″ high.
2. A half-high drive is 1.63″ high.
3. HSIII - Hard disk is in a Hot-swap III drive tray.
4. 704 - Hard disk is for the PC Server 704 only.

20.4 UltraSCSI Hard Disks

IBM also offers a 2.16 GB UltraSCSI hard disk. The table below shows the specifications of this device.

Table 34. IBM UltraSCSI Disk Drive

Part number	07H1126
Rotation speed	5400 RPM
Average seek time	8.5 ms
Average latency	5.6 ms
Adaptive buffer	512 KB

Part 4. Appendixes

Appendix A. Sources of Drivers and Information

One of the challenges that LAN administrators face is ensuring that the adapters they want to use are supported and then finding the latest level of device drivers for these adapters. Another problem they face is trying to find information about known software bugs and their corresponding fixes. This appendix is included to help readers solve these problems.

For a general source of information and support on IBM PC Servers, go to the World Wide Web (WWW) at Universal Resource Locators (URLs):

```
http://www.pc.ibm.com/servers
http://www.pc.ibm.com/techlink/tcsrvsp1.html
```

A.1 Compatibility Information

IBM does extensive testing on each new system before it is introduced to ensure that it is compatible with the wide variety of IBM and non-IBM hardware and software. It is available from the following Web site:

```
http://www.pc.ibm.com/compat/compat.html
```

This site has a wealth of valuable information. It contains the following:

- A comprehensive listing of products that have been tested for compatibility with these servers and a technical description of each server.
- Compatibility reports and information about network and workstation operating systems, showing detailed information on hardware configurations, adapters, device drivers, and code levels used in the testing.
- Direct access to device drivers, flash BIOS updates, and other server code.
- An overview of how IBM conducts compatibility and certification tests.

This is an excellent source of information to ensure your hardware and software are supported by IBM servers. It is updated regularly with the latest NOS certifications and driver updates.

A.2 Device Drivers

The IBM PC Company Home Page has a very good file repository containing drivers and BIOS updates, as well as other documentation and drivers:

```
http://www.pc.ibm.com/files.html
```

In addition to this Web site, IBM has another site that contains a file that has LAN adapter device driver information. This file is a matrix of adapters and drivers and version levels. It is updated frequently as new drivers become available. The file is named DRVRLIST.EXE and is a self-extracting ZIP file containing a Word 6.0 document. It can be obtained from the PC Company BBS at either:

```
ftp://ftp.pcco.ibm.com/pub/network
```

or via modem: (919) 517-0001 in Raleigh, North Carolina in the US. There are also IBM bulletin boards in other countries where IBM operates.

If it is an IBM adapter, the driver itself can be obtained from the same site. In addition, the IBM Networking Hardware Division maintains an FTP site at URL:

ftp://lansupport.raleigh.ibm.com/pub/products/lanprods

For OEM adapters, many manufactures have web and FTP sites which work very similar to the process outlined above. Examples include:

http://www.adaptec.com
http://www.mylex.com
http://www.madge.com
http://www.3com.com

Similarly, operating system suppliers may also have useful information for your installation:

http://www.austin.ibm.com/pspinfo
http://www.lotus.com
http://www.microsoft.com/NTServer
http://www.sco.com/products
http://www.novell.com
http://www.banyan.com
http://www.sun.com/cgi-bin/show?sunsoft/solaris/index.body

For specifically OS/2 drivers and information, there is an OS/2 device driver home page:

http://www.europe.ibm.com/getdoc/psmemea/progserv/device/

and a database of known problems (and fixes) with OS/2:

http://ps.boulder.ibm.com/pbin-usa-ps/pub_search.pl

There are also useful Web sites ran by people not necessarily related to hardware or software vendors. One such site is the LAN Drivers Page at:

http://sunsite.unc.edu/~ towfiq/lan-drivers.html

Appendix B. IBM TechConnect Program

The IBM TechConnect program is a server training, certification and information resource program. It is designed to help network specialists increase their technical skills and proficiency on IBM servers as well as networking and open computing solutions.

IBM TechConnect will offer server technical specialists value-added capabilities by offering state of the industry hands-on PC Server training and certification as well as:

- Industry-leading PC Server products and technical field support
- Technical specialist peer networking opportunities
- Unique incentives and rewards in recognition of re-seller and individual′s technical achievements

In today′s competitive environment, business customers of every size are relying on increasingly sophisticated PC Server networks to support, run and profitably manage their businesses. In addition to requiring advanced, fault-tolerant PC Servers, customers require highly skilled, qualified technical professionals to design, install and support their growing networks. The IBM TechConnect program is designed to meet these customer challenges. TechConnect focuses on informing, training, educating, and developing highly skilled people to fully support IBM PC Server customers from a total networking solutions perspective.

The program works at two levels:

1. The *Associate path* is an information connection, a quarterly mailing of our TechLink pack that contains technical information designed specifically for networking specialists.

2. The *Certification path* includes hands-on lab training with real-life problem solving opportunities on IBM PC Server products. Course content will focus on underlying technology, in addition to specific products, to maximize value and provide transferable skills.

B.1 Certification

By attending the technical training being run through the IBM PC institute, individuals are working towards certification as an IBM Professional Server Specialist (PSS) or an IBM Professional Server Expert (PSE), which provides unique incentives and rewards in recognition of re-seller and individual′s technical achievements.

The courses descriptions and codes are:

- IBM Server Technical Training (V5051): PSS and PSE designation
- IBM Server/Novell NetWare Installation and Performance (V5052): PSE designation
- IBM Server/IBM LAN Server Installation and Performance (V5053): PSE designation
- IBM Server/Microsoft Windows NT (V5055): PSE designation

The professional roles being certified by IBM are:

- IBM Professional Server Specialist (PSS)

- IBM Professional Server Expert (PSE) - Novell NetWare
- IBM Professional Server Expert (PSE) - IBM LAN Server
- IBM Professional Server Expert (PSE) - Microsoft Windows NT
- IBM Professional Server Enterprise Expert (PSEE)

The Professional Server Enterprise Expert (PSEE) curriculum, open to those holding PSE designations, will include client/server-mainframe connectivity specialization paths with course work covering LAN, WAN, network management, backup/recovery and security.

To certify as a Professional Server Specialist (PSS) individuals must complete the IBM Server Technical Training course and gain a pass mark in the certification test. The PSS will receive a quarterly TechBox that includes the TechLink plus audio and video tapes and CDs.

To certify as a Professional Server Expert (PSE) individuals must complete the IBM server technical training course and one of the three IBM server/network operating system installation and performance courses. Individuals may choose more than one IBM server/network operating system installation and performance course.

In addition, two certification tests must be passed covering topics from the IBM server technical training course and one of the IBM server/network operating system installation and performance courses.

IBM will provide support to Professional Server Experts (PSEs) by offering the following benefits:

- Priority access to the IBM network and server support center via Internet to enable support of networks and effective problem determination and resolution of server/LAN issues.
- Quarterly technical mailing - TechBox.
- PSEs working for IBM business partners will enable their employer to take advantage of additional incentives on the purchase of IBM server product.
- Recognition programs will be run to recognize technical excellence in our IBM PSEs.

As a prerequisite to certifying as a Professional Server Expert, an individual must show proof of one of the following network operating systems certifications:

- Novell Certified NetWare Engineer
- Certified OS/2 LAN Server Engineer
- Windows NT Windows

For more information on TechConnect on the Web, go to the TechConnect home page:

`http://www.pc.ibm.com/techlink`

For information about OS/2 LAN Server certification, go to URL:

`http://www.austin.ibm.com/pspinfo/proroad.htm`

Appendix C. Special Notices

This publication is intended to help IBM customers, dealers, and other technical professionals select the most appropriate PC server for their requirements. The information in this publication is not intended as the specification of any programming interfaces that are provided by IBM. See the PUBLICATIONS section of the IBM Programming Announcement for the various products covered for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM′s product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM′s intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, 500 Columbus Avenue, Thornwood, NY 10594 USA.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer′s ability to evaluate and integrate them into the customer′s operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

DB2/2	EtherStreamer
IBM	LANStreamer
Micro Channel	NetFinity
NetView	OS/2
PowerPC	Predictive Failure Analysis
PS/2	SAA
Streamer	SystemView
WebExplorer	400

The following terms are trademarks of other companies:

C-bus is a trademark of Corollary, Inc.

PC Direct is a trademark of Ziff Communications Company and is used by IBM Corporation under license.

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

Windows is a trademark of Microsoft Corporation.

Java and HotJava are trademarks of Sun Microsystems, Inc.

Adaptec	Adaptec Incorporated
APC	American Power Conversion
Apple	Apple Computer, Incorporated
Banyan	Banyan Systems, Incorporated
Boca Research	Boca Research, Incorporated
Centronics	Centronics Data Computer Corporation
Cirrus Logic	Cirrus Logic, Incorporated
Compaq	Compaq Computer Corporation
Digital	Digital Equipment Corporation
ESS	American Telephone and Telegraph Company
GEM	Digital Research, Incorporated
Intel	Intel Corporation
InterNotes	Lotus Development Corporation
InterNotes Web Publisher	Lotus Development Corporation
IPX	Novell, Incorporated
i386	Intel Corporation
i486	Intel Corporation
i960	Intel Corporation
Lexmark	Lexmark International, Incorporated
Lotus Notes	Lotus Development Corporation
Mylex	Mylex Corporation
National Semiconductor	National Semiconductor Corporation
NCR	NCR Corporation
Netscape	Netscape Communications Corporation
Netscape Navigator	Netscape Communications Corporation
NetWare	Novell, Incorporated
Notes	Lotus Development Corporation
Novell	Novell, Incorporated
Open Server	The Santa Cruz Operation, Incorporated
OpenView	Hewlett-Packard Company
Pentium	Intel Corporation
Pentium Pro	Intel Corporation
Phoenix	Phoenix Technologies, Limited
PostScript	Adobe Systems, Incorporated
PowerChute	American Power Conversion
SCO	The Santa Cruz Operation, Incorporated
SCSI	Security Control Systems, Incorporated
Seagate	Seagate Technology, Incorporated
Smart-UPS	American Power Conversion
Solaris	Sun Microsystems, Incorporated
Tivoli	Tivoli Systems Inc., an IBM Company
TME 10	Tivoli Systems Inc., an IBM Company
UNIX	Use wording from NONIBM TMREQ.
UnixWare	Novell, Incorporated
VESA	Video Electronics Standards Association

3M	Minnesota Mining & Manufacturing Company
486	Intel Corporation

Other trademarks are trademarks of their respective companies.

Appendix D. Related Publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

D.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see "How To Get ITSO Redbooks" on page 213.

- *Advanced PS/2 Servers Planning and Selection Guide*, GG24-3927
- *IBM PC Server Disk Subsystem Configuration and Sizing*, SG24-4525
- *IBM PC Server and NetWare Integration Guide*, SG24-4576
- *IBM PC Server and OS/2 LAN Server Integration Guide*, SG24-4577
- *IBM PC Server and Windows NT Integration Guide*, SG24-4578
- *LAN Management Processes (Alerts/Monitoring) Using NetFinity*, SG24-4517
- *Workgroup Management Using SystemView for OS/2*, SG24-2596
- *Systems Management from an NT Server Point of View*, SG24-4723

D.2 Redbooks on CD-ROMs

Redbooks are also available on CD-ROMs. **Order a subscription** and receive updates 2-4 times a year at significant savings.

CD-ROM Title	Subscription Number	Collection Kit Number
System/390 Redbooks Collection	SBOF-7201	SK2T-2177
Networking and Systems Management Redbooks Collection	SBOF-7370	SK2T-6022
Transaction Processing and Data Management Redbook	SBOF-7240	SK2T-8038
AS/400 Redbooks Collection	SBOF-7270	SK2T-2849
RS/6000 Redbooks Collection (HTML, BkMgr)	SBOF-7230	SK2T-8040
RS/6000 Redbooks Collection (PostScript)	SBOF-7205	SK2T-8041
Application Development Redbooks Collection	SBOF-7290	SK2T-8037
Personal Systems Redbooks Collection	SBOF-7250	SK2T-8042

How To Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, CD-ROMs, workshops, and residencies. A form for ordering books and CD-ROMs is also provided.

This information was current at the time of publication, but is continually subject to change. The latest information may be found at URL http://www.redbooks.ibm.com.

How IBM Employees Can Get ITSO Redbooks

Employees may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **PUBORDER** — to order hardcopies in United States
- **GOPHER link to the Internet** - type GOPHER.WTSCPOK.ITSO.IBM.COM
- **Tools disks**

  To get LIST3820s of redbooks, type one of the following commands:

  ```
  TOOLS SENDTO EHONE4 TOOLS2 REDPRINT GET SG24xxxx PACKAGE
  TOOLS SENDTO CANVM2 TOOLS REDPRINT GET SG24xxxx PACKAGE (Canadian users only)
  ```

  To get lists of redbooks:

  ```
  TOOLS SENDTO WTSCPOK TOOLS REDBOOKS GET REDBOOKS CATALOG
  TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET ITSOCAT TXT
  TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET LISTSERV PACKAGE
  ```

  To register for information on workshops, residencies, and redbooks:

  ```
  TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ITSOREGI 1996
  ```

  For a list of product area specialists in the ITSO:

  ```
  TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ORGCARD PACKAGE
  ```

- **Redbooks Home Page on the World Wide Web**

  http://w3.itso.ibm.com/redbooks

- **IBM Direct Publications Catalog on the World Wide Web**

  http://www.elink.ibmlink.ibm.com/pbl/pbl

  IBM employees may obtain LIST3820s of redbooks from this page.

- **REDBOOKS category on INEWS**
- **Online** — send orders to: USIB6FPL at IBMMAIL or DKIBMBSH at IBMMAIL
- **Internet Listserver**

  With an Internet E-mail address, anyone can subscribe to an IBM Announcement Listserver. To initiate the service, send an E-mail note to announce@webster.ibmlink.ibm.com with the keyword subscribe in the body of the note (leave the subject line blank). A category form and detailed instructions will be sent to you.

How Customers Can Get ITSO Redbooks

Customers may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Online Orders** (Do not send credit card information over the Internet) — send orders to:

	IBMMAIL	**Internet**
In United States:	usib6fpl at ibmmail	usib6fpl@ibmmail.com
In Canada:	caibmbkz at ibmmail	lmannix@vnet.ibm.com
Outside North America:	dkibmbsh at ibmmail	bookshop@dk.ibm.com

- **Telephone orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	(long distance charges apply)
(+45) 4810-1320 - Danish	(+45) 4810-1020 - German
(+45) 4810-1420 - Dutch	(+45) 4810-1620 - Italian
(+45) 4810-1540 - English	(+45) 4810-1270 - Norwegian
(+45) 4810-1670 - Finnish	(+45) 4810-1120 - Spanish
(+45) 4810-1220 - French	(+45) 4810-1170 - Swedish

- **Mail Orders** — send orders to:

IBM Publications	IBM Publications	IBM Direct Services
Publications Customer Support	144-4th Avenue, S.W.	Sortemosevej 21
P.O. Box 29570	Calgary, Alberta T2P 3N5	DK-3450 Allerød
Raleigh, NC 27626-0570	Canada	Denmark
USA		

- **Fax** — send orders to:

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	(+45) 48 14 2207 (long distance charge)

- **1-800-IBM-4FAX (United States)** or **(+1) 415 855 43 29 (Outside USA)** — ask for:

  Index # 4421 Abstracts of new redbooks
  Index # 4422 IBM redbooks
  Index # 4420 Redbooks for last six months

- **Direct Services** - send note to softwareshop@vnet.ibm.com

- **On the World Wide Web**

  Redbooks Home Page http://www.redbooks.ibm.com
  IBM Direct Publications Catalog http://www.elink.ibmlink.ibm.com/pbl/pbl

- **Internet Listserver**

  With an Internet E-mail address, anyone can subscribe to an IBM Announcement Listserver. To initiate the service, send an E-mail note to announce@webster.ibmlink.ibm.com with the keyword subscribe in the body of the note (leave the subject line blank).

IBM Redbook Order Form

Please send me the following:

Title	Order Number	Quantity

- **Please put me on the mailing list for updated versions of the IBM Redbook Catalog.**

First name Last name

Company

Address

City Postal code Country

Telephone number Telefax number VAT number

- Invoice to customer number
- Credit card number

Credit card expiration date Card issued to Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

DO NOT SEND CREDIT CARD INFORMATION OVER THE INTERNET.

Index

Numerics

A

B

C

S

T

U

V

W

Z

Printed in U.S.A.

SG24-4899-00