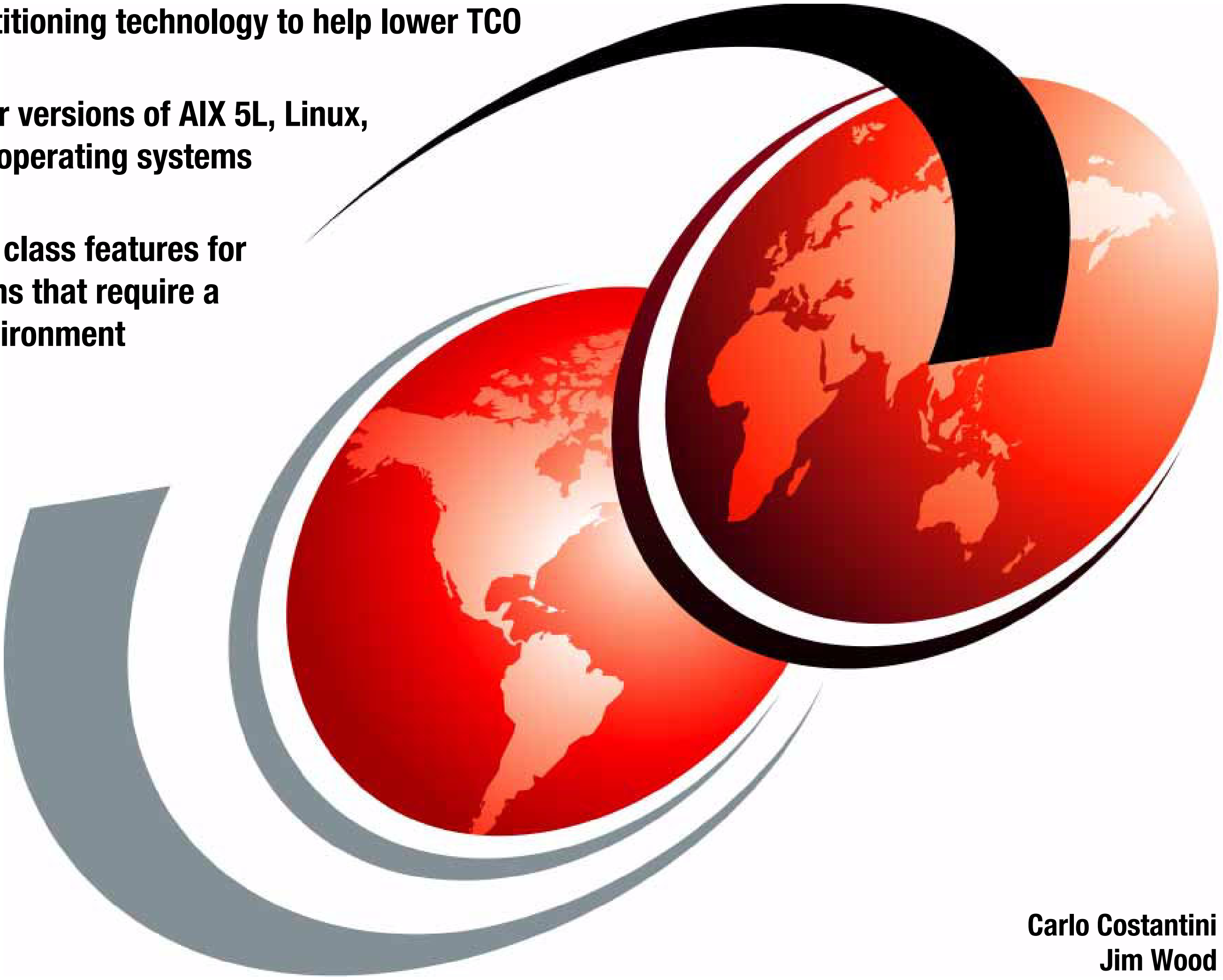
IBM eserver p5 590 and 595 Technical Overview and Introduction



- Finer system granularity using Micro-Partitioning technology to help lower TCO
- Support for versions of AIX 5L, Linux, and i5/OS operating systems
- Enterprise class features for applications that require a robust environment

Carlo Costantini
Jim Wood

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM @server p5 590 and 595 Technical Overview and Introduction

August 2005

Note: Before using this information and the product it supports, read the information in "Notices" on page vii.

First Edition (August 2005)

This edition applies to the IBM *@server* p5 590 and 595, and AIX 5L Version 5.3, product number 5765-G03.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described 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:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX 5L™
AIX®
Chipkill™
Electronic Service Agent™
Enterprise Storage Server®
@server®
@server®
Hypervisor™
HACMP™
i5/OS™
IBM®
ibm.com®
iSeries™
Micro-Partitioning™
Notes®
OpenPower™
Perform™
PowerPC®
POWER™
POWER4™
POWER4+™
POWER5™
pSeries®
Redbooks™
Redbooks (logo)™
Resource Link™
RS/6000®
TotalStorage®
Virtualization Engine™
WebSphere®
zSeries®

The following terms are trademarks of other companies:

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, and service names may be trademarks or service marks of others.

Preface

This document is a comprehensive guide covering the IBM® ⓔserver® p5 590 and 595 AIX 5L™ and Linux® operating system servers. We introduce major hardware offerings and discuss their prominent functions.

Professionals wishing to acquire a better understanding of IBM ⓔserver p5 products should consider reading this document. The intended audience includes:

- Clients
- Sales and marketing professionals
- Technical support professionals
- IBM Business Partners
- Independent software vendors

This document expands the current set of IBM ⓔserver documentation by providing a desktop reference that offers a detailed technical description of the p5-590 and p5-595 servers.

This publication does not replace the latest IBM ⓔserver marketing materials and tools. It is intended as an additional source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.

The team that wrote this Redpaper

This Redpaper was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center.

Carlo Costantini is a Certified IT Specialist for IBM and has over 27 years of experience with IBM and IBM Business Partners. He currently works in Italy Presales Field Technical Sales Support for IBM Sales Representatives and IBM Business Partners for all pSeries® and IBM ⓔserver p5 systems offerings. He has broad marketing experience. He is a certified specialist for pSeries and IBM ⓔserver p5 systems.

Jim Wood is a Technical Support Specialist for IBM and has 20 years of experience with IBM and IBM Business Partners. He currently works in the UK Hardware Front Office supporting customers and IBM Service Representatives for all pSeries and RS/6000® products. He holds a First Class Honours Degree in IT and Computing and is a Chartered Member of the British Computer Society. He is also an AIX 5L certified specialist.

The project that created this publication was managed by:
Scott Vetter

Thanks to the following people for their contributions to this project:

Arzu Gucer
International Technical Support Organization, Austin Center

Jim Mitchell, George H. Ahrens, Daniel Henderson, Todd Rosedahl, Pete Wendling, Gary Anderson, Salim A. Agha, Ajay K. Mahajan, Tenley Jackson, Ron Barker, Bill Mihaltse, Matt

Robbins, Robert Bluethman
IBM U.S.

Clive Benjamin, Derrick Daines, Dave Williams
IBM U.K.

Guliano Anselmi
IBM Italy

Timothy Gilson
Visa Europe Unix Platform Team

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and/or customers.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this Redpaper or other Redbooks™ in one of the following ways:

- Use the online **Contact us** review redbook form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbook@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. JN9B Building 905
  11501 Burnet Road
  Austin, Texas 78758-3493

1

General description

The IBM eserver p5 590 and IBM eserver p5 595 are the servers redefining the IT economics of enterprise UNIX® and Linux computing. The up to 64-way p5-595 server is the new flagship of the product line with nearly three times the commercial performance (based on rPerf estimates) and twice the capacity of its predecessor, the IBM eserver pSeries 690. Accompanying the p5-595 is the up to 32-way p5-590 that offers enterprise-class function and more performance than the pSeries 690 at a significantly lower price for comparable configurations.

As standard, these servers come with mainframe-inspired reliability, availability, serviceability (RAS) capabilities and IBM Virtualization Engine™ systems technology with breakthrough innovations such as Micro-Partitioning™ technology. Micro-Partitioning technology allows as many as ten logical partitions (LPARs) per processor to be defined. Both systems can be configured with up to 254 virtual servers with a choice of AIX 5L, Linux, and i5/OS™ operating systems in a single server, designed to enable cost-saving consolidation opportunities.

Note: Not all system features available under the AIX 5L operating system are available under the Linux operating system.

1.1 Model abstract for 9119-590 and 9119-595

The 9119 IBM @server p5 models 590 and 595 provide an expandable high-end enterprise solution for managing e-business computing requirements.

Table 1-1 represents the major product attributes of these models with the major differences highlighted by shading.

Table 1-1 9119-590 and 9119-595 attributes

Attribute	9119-590	9119-595
SMP processor configurations	8-way to 32-way	16-way, 32-way, 48-way, and 64-way
Maximum 16-way CPU books	2	4
Processor clock rate	1.65 GHz	1.65 GHz Standard or 1.9 GHz Turbo
Processor cache per processor pair	1.9 MB Level 2 36 MB Level 3	1.9 MB Level 2 36 MB Level 3
Processor packaging	MCM	MCM
64-bit copper technology POWER5™ processor	Y	Y
Maximum memory configuration	1 TB	2 TB
Rack space	42U 24-inch custom rack	42U 24-inch custom rack
Maximum number of I/O drawers	8	12
Maximum number of PCI-slots	160	240
Maximum number of 15 K rpm disks	128	192
Dual service processors	Y	Y
Integrated redundant power	Y	Y
Battery backup option	Y	Y
Powered expansion rack available	N	Y
Dynamic LPAR	Y	Y
Micro-Partitioning technology with up to 254 partitions	Y	Y
Acoustic rack doors available	Y	Y
Support for AIX 5L, Linux, and i5/OS	Y	Y

Each 16-way processor book also includes 16 slots for memory cards and six Remote I/O-2 attachment cards for connection of the system I/O drawers.

Each I/O drawer contains twenty 3.3-volt PCI-X adapter slots and up to sixteen disk bays.

The AIX 5L V5.2 and V5.3, Linux, and i5/OS V5R3 operating systems can run simultaneously in different partitions within the a server.

1.2 System frames

Both the p5-590 and p5-595 systems are based on the same 24-inch wide, 42 EIA height frame. Inside this frame all the server components are placed in predetermined positions. This design and mechanical organization offers advantages in optimization of floor space usage.

The p5-590 and p5-595 servers are designed with a basic server configuration that starts with a single *frame* (Figure 1-1) and is featured with optional and required components.

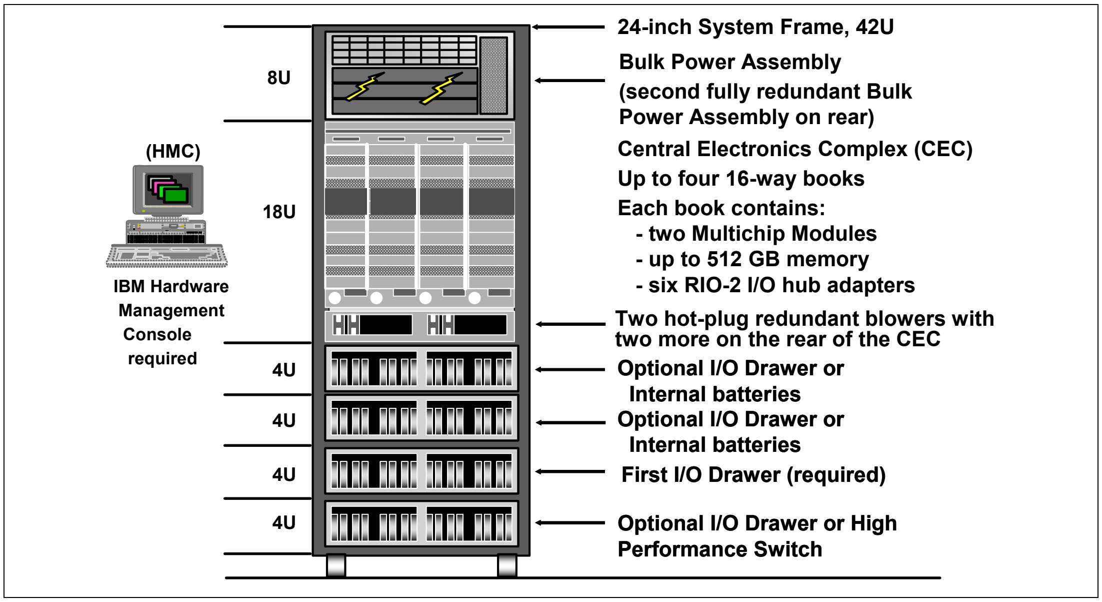

8U
18U
4U
4U
4U
4U
(HMC)
IBM Hardware Management Console required
24-inch System Frame, 42U
Bulk Power Assembly (second fully redundant Bulk Power Assembly on rear)
Central Electronics Complex (CEC)
Up to four 16-way books
Each book contains:
- two Multichip Modules
- up to 512 GB memory
- six RIO-2 I/O hub adapters
Two hot-plug redundant blowers with two more on the rear of the CEC
Optional I/O Drawer or Internal batteries
Optional I/O Drawer or Internal batteries
First I/O Drawer (required)
Optional I/O Drawer or High Performance Switch

Figure 1-1 Primary system frame organization

For additional capacity, either a powered or non-powered frame can be configured for a p5-595, or a non-powered frame for the p5-590, as shown in Figure 1-2.

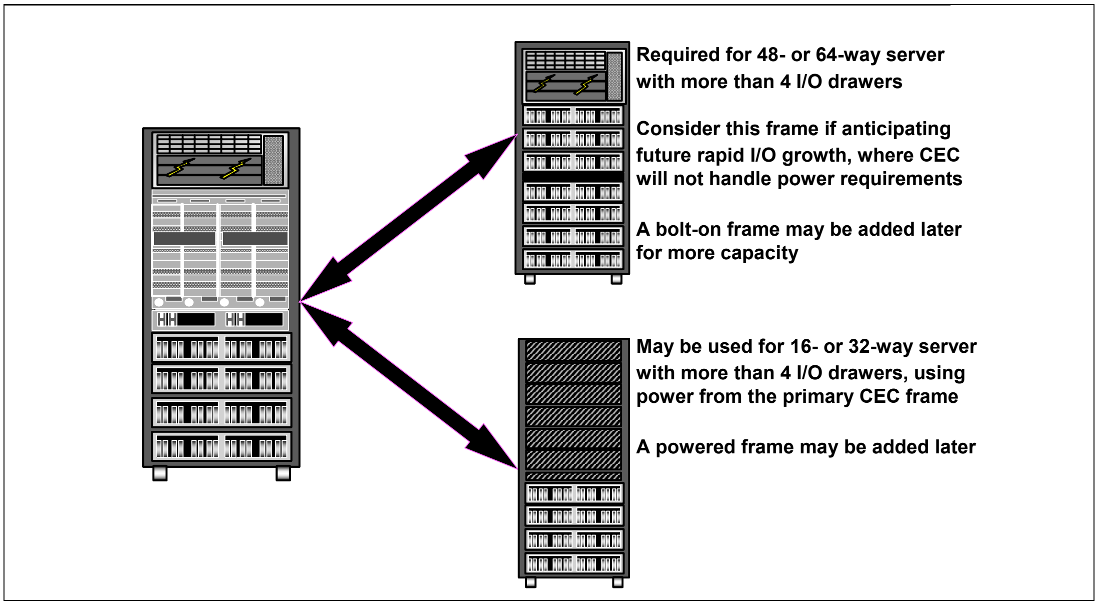

Required for 48- or 64-way server with more than 4 I/O drawers
Consider this frame if anticipating future rapid I/O growth, where CEC will not handle power requirements
A bolt-on frame may be added later for more capacity
May be used for 16- or 32-way server with more than 4 I/O drawers, using power from the primary CEC frame
A powered frame may be added later

Figure 1-2 Powered and non-powered bolt-on frames

1.3 Installation planning

Product installation and in-depth system cabling are beyond the scope of this paper. Complete installation instructions are shipped with each order. Comprehensive planning advice is available at this address:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

Key specifications are described in the following sections.

1.3.1 System specifications

Table 1-2 lists the general system specifications of the p5-590 and p5-595 servers.

Table 1-2 IBM eServer p5 590 and 595 server specifications

Description	Range
Recommended operating temperature (8-way, 16-way, 32-way)	10 degrees to 32 degrees C (50 degrees to 89.6 degrees F)
Recommended operating temperature (48-way and 64-way)	10 degrees to 28 degrees C (50 degrees to 82.4 degrees F)
Operating voltage	200 to 240, 380 to 415, or 480 volts AC
Operating frequency	50/60 plus or minus 0.5 Hz
Maximum power consumption (1.9 GHz processor)	22.7 kW
Maximum power consumption (1.65 GHz processor)	20.3 kW
Maximum thermal output (1.9 GHz processor)	77.5 kBtu/hr (British Thermal Unit)
Maximum thermal output (1.65 GHz processor)	69.3 kBtu/hr (British Thermal Unit)

1.3.2 Physical package

Table 1-3 lists the major physical attributes found on the p5-590 and p5-595 servers.

Table 1-3 IBM eServer p5 590 and p5 595 server physical packaging

Dimension	
Height	2025 mm (79.7 in.)
Width	785 mm (30.9 in.)
Depth	1326 mm (52.2 in.)[a] or 1681 mm (66.2 in.)[b]
Weight	
Minimum configuration	1241 kg (2735 lb)
Maximum configuration	2458 kg (5420 lb)

[a] With slimline doors installed
[b] With acoustical doors installed

1.3.3 Service clearances

There are several possible frame configurations of the p5-590 and p5-595 servers. FC 7960 is designed to provide improved access through doorways during shipment.

Figure 1-3 shows service clearances for double-frame systems with acoustical doors.

Note: The p5-595 server must be installed in a raised floor environment.

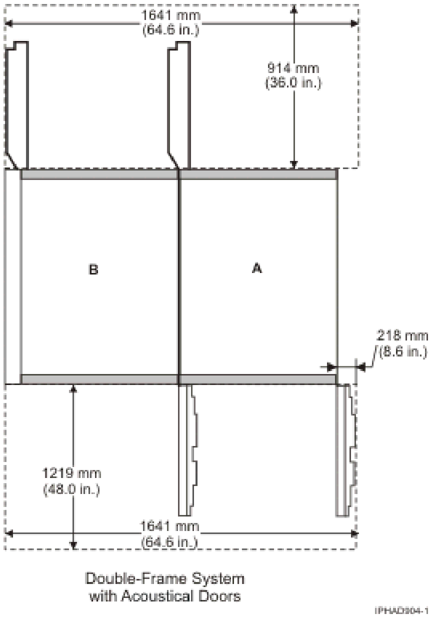


Figure 1-3 Service clearances

Service clearances for other configurations can be found at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/info/iphad/serviceclearance.htm#serviceclearance

1.4 Power and cooling

The p5-590 and p5-595 provide full power and cooling redundancy, with dual power cords and variable-speed fans and blowers, for both the Central Electronics Complex (CEC) and the I/O drawers. Redundant hot-plug power and cooling subsystems provide power and cooling backup in case units fail and allow for concurrent replacement. In the event of a complete power failure, Early Power Off Warning capabilities are designed to perform an orderly shutdown.

The primary system rack and powered Expansion Rack always incorporate two bulk power assemblies for redundancy. These provide 350 V dc power for devices located in those racks and associated nonpowered Expansion Racks. These bulk power assemblies are mounted in front and rear positions and occupy the top 8U of the rack. To help provide optimum system availability, these bulk power assemblies should be powered from separate power sources with separate line cords.

An optional Integrated Battery Backup (IBF) is available, if desired. The battery backup features are designed to protect against power line disturbances such as short-term power loss or brown-out conditions. The battery backup features each require 2U of space in the primary system rack or in the powered Expansion Rack. The battery backup features attach to the system bulk power regulators. The IBF is *not* an Uninterruptable Power Source (UPS) meant to keep the system powered on indefinitely in case of AC line outage.

In case of a fan or blower failure, the remaining fans automatically increase speed to compensate for the lost air flow from the failed component.

1.5 Minimum and optional features

The purpose of this section is to establish the minimum configuration for a p5-590 and p5-595. Appropriate feature codes for each system component are also provided. The IBM Configurator tool will also identify the feature code for each component used to build your system configuration.

> **Note:** Throughout this chapter all feature codes are referenced as FC *xxxx*, where *xxxx* is the appropriate feature code number of the particular item.

Table 1-4 identifies the components required to construct a minimum configuration for a p5-590.

Table 1-4 p5-590 minimum system configuration

Quantity	Component description	Feature code (FC)
One	IBM @server p5 590	9119-590
One	Media drawer for installation and service actions (additional media features may be required) without NIM	19-inch 7212-102 or FC 5795
One	16-way, POWER5 processor book, 0-way Active	FC 7981
Eight	1-way, processor activations	FC 7925
Two	Memory cards with a minimum of 8 GB of activated memory	Refer to the Sales Manual for valid memory configuration feature codes
Two	Processor clock cards, programmable	FC 7810
One	Power cable group, bulk power to CEC and fans	FC 7821
Three	Power converter assemblies, Central Electronics Complex	FC 7809
One	Power cable group, first processor book	FC 7822
Two	System service processors	FC 7811
One	Multiplexer card	FC 7812
Two	RIO-2 loop adapters, single loop	FC 7818
One	I/O drawer Note: requires 4U frame space	FC 5791 or FC 5794
One	Remote I/O (RIO) cable, 0.6 M Note: used to connect drawer halves	FC 7924
Two	Remote I/O (RIO) cables, 2.5 M	FC 3168
Two	15,000 rpm Ultra3 SCSI disk drive assemblies	FC 3277 or FC 3278

Quantity	Component description	Feature code (FC)
One	I/O drawer attachment cable group	FC 6122
One	Slim line or acoustic door kit	FC 6251 or FC 6252
Two	Bulk power regulators	FC 6186
Two	Bulk power controller assemblies	FC 7803
Two	Bulk power distribution assemblies	FC 7837
Two	Line cords	FC 86xx Refer to the Sales Manual for specific line cord feature code options.
One	Language specify	FC 9xxx Refer to the Sales Manual for specific language feature code options.
One	Hardware management console	7310-C04, 7310- CR3 (*)

Table 1-5 identifies the components required to construct a minimum configuration for a p5-595.

Table 1-5 p5-595 minimum system configuration

Quantity	Component description	Feature code (FC)
One	IBM eserver p5 595	9119-595
One	16-way, POWER5 Processor Book, 0-way Active	FC 7813 or FC 7988
Note: The following two components must be added to p5-595 servers with one processor book (FC 7813) One - Cooling Group (FC 7807) One - Power Cable Group (FC 7826)		
Sixteen	1-way, processor activations	FC 7815 or FC 7990
Two	Memory cards with a minimum of 8 GB of activated memory	Refer to the Sales Manual for valid memory configuration feature codes.
Two	Processor clock cards, programmable	FC 7810
One	Power cable group, bulk power to cec and fans	FC 7821
Three	Power converter assemblies, Central Electronics Complex	FC 7809
One	Power cable group, first processor book	FC 7822
One	Multiplexer card	FC 7812
Two	Service processors	FC 7811
Two	RIO-2 loop adapter, single loop	FC 7818
One	I/O drawer Note: 4U frame space required	FC 5791 or FC 5794
One	Remote I/O (RIO) cable, 0.6 M Note: Used to connect drawer halves	FC 7924

Quantity	Component description	Feature code (FC)
Two	Remote I/O (RIO) cables, 3.5 M	FC 3147
Two	15,000 rpm Ultra3 SCSI disk drive assembly	FC 3277 or FC 3278
One	PCI SCSI Adapter or PCI LAN Adapter for attachment of a device to read CD media or attachment to a NIM server	Refer to the Sales Manual for valid adapter feature code
One	I/O drawer attachment cable group	FC 6122
One	Slim line or acoustic door kit	FC 6251 or FC 6252
Two	Bulk power regulators	FC 6186
Two	Power controller assemblies	FC 7803
Two	Power distribution assemblies	FC 7837
Two	Line cords	FC 86xx Refer to your Sales Manual for specific line cord feature code options.
One	Language specify	FC 9xxx Refer to your Sales Manual for specific language feature code options.
One	Hardware management console	7310-C04, 7310- CR3 (*)

(*) An HMC is required, and two HMCs are recommended. A private network with the HMC providing DHCP services is mandatory on these systems; see “System management” on page 50.

The p5-590 and p5-595 servers support AIX 5L and Linux operating systems (OS) and require the following specific levels:

- AIX 5L Version 5.2 or Version 5.3, or later
- SUSE LINUX Enterprise Server 9 (SLES 9) for POWER™ or later
- Red Hat Enterprise Linux (RHEL AS 3) for POWER Version 3 or later
- i5/OS V5R3 or later

1.5.1 Processor features

The p5-590 system features base 8-way Capacity On Demand (CoD), 16-way, and 32-way configurations with the POWER5 processor running at 1.65 GHz. The p5-595 system features base 16-way, 32-way, 48-way, and 64-way configurations with the POWER5 processor running at 1.65 GHz or 1.9 GHz. Processors can be activated in increments of 1 (refer to 3.1.2, “Capacity Upgrade on Demand for memory” on page 70).

The p5-590 and p5-595 system configuration is based on the processor book. To configure it, it is necessary to order one or more of the following components:

- One or more 16-way processor book, 0-way active
- Activation codes to reach the expected configuration

Note: Any p5-595 or p5-590 system made of more than one processor book must have all processor cards running at the same speed.

For a list of available processor features, refer to Table 1-6.

Table 1-6 Available processor options

Feature code	Description
7813	16-way POWER5 Turbo CUoD Processor Book, 0-way active
7988	16-way POWER5 Standard CUoD Processor Book, 0-way active
7815	Activation, FC 7813 CUoD Processor Book, One Processor
7990	Activation, FC 7988 CUoD Processor Book, One Processor

Note: The POWER5 turbo processor uses 1.9 GHz clocking. The standard POWER5 processor uses 1.65 GHz clocking.

1.5.2 Memory features

The p5-590 and p5-595 have the following minimum and maximum configurable memory resource allocation requirements:

- Both p5-590 and p5-595 require a minimum of 8 GB of configurable system memory.
- Each processor book provides 16 memory card slots for a maximum of 32 memory cards (p5-590) or 64 memory cards (p5-595) per server.
- The p5-590 supports a maximum of 1024 GB of DDR1 configurable memory or 128 GB of DDR2 memory.
- The p5-595 supports a maximum of 2048 GB of DDR1 configurable memory or 256 GB of DDR2 memory.

Memory can be activated in increments of 1Gb (refer to 3.1.2, “Capacity Upgrade on Demand for memory” on page 70).

Table 1-7 lists the available memory features.

Table 1-7 Memory feature codes

Feature code	Description
7816	4 GB 266 MHz CUoD card with 2 GB active, DDR1
7835	8 GB 266 MHz CUoD card with 4 GB active, DDR1
7828	16 GB fully activated 266 MHz card, DDR1
7829	32 GB fully activated 200 MHz card, DDR1
8195	256 GB package of 32 fully activated 8 GB 266 MHz cards, DDR1
8197	512 GB package of 32 fully activated 16 GB 266 MHz cards, DDR1
8198	512 GB package of 16 fully activated 32 GB 200 MHz cards
7814	4 GB 533 MHz memory card, DDR2

1.5.3 USB diskette drive

In some situations, an external USB 1.44 MB diskette drive for p5-590 and p5-595 servers (FC 2591) is helpful. This lightweight USB 2.0 attached diskette drive takes its power requirements from the USB port. A USB cable is provided. The drive can be attached to the

USB adapter (FC 2738). A maximum of one USB diskette drive is supported per controller. The same controller can share a USB mouse and keyboard. Only features available through IBM are supported on the USB ports.

1.5.4 Hardware Management Console models

The Hardware Management Console (HMC) is a dedicated workstation that allows you to configure and manage partitions. The hardware management application helps you configure and partition the server through a graphical user interface. An HMC is mandatory for a p5-590 or p5-595 server; however, IBM highly recommends redundant HMCs.

Functions performed by the HMC include:

- Creating and maintaining a multiple partition environment
- Displaying a virtual operating system session terminal for each partition
- Displaying a virtual operator panel of contents for each partition
- Detecting, reporting, and storing changes in hardware conditions
- Powering managed systems on and off
- Acting as a service local point for service representatives to determine an appropriate service strategy
- Controling CoD resources

See “System management” on page 50 for detailed information on the HMC.

Table 1-8 lists the HMC options for POWER5 processor-based systems available at the time of writing.

Table 1-8 Available HMCs

Type-model	Description
7310-C04	IBM 7310 Model C04 Desktop Hardware Management Console
7310-CR3	IBM 7310 Model CR3 Rack-Mount Hardware Management Console

1.6 External disk subsystem

The p5-590 and p5-595 servers have internal hot-swappable drives supported in I/O drawers. The I/O drawers may be FC 5791 or FC 5794 or existing 7040-61D drawers migrated from a 7040 server. Internal disks are usually used for the base OS and paging space. Specific client requirements can be satisfied with several external disk possibilities that the p5-590 and p5-595 servers support.

The following section covers storage subsystems available at the time of writing. For further information about IBM disk storage subsystems, including withdrawn products such as 7133 SSA subsystems, and earlier models of those mentioned below, visit:

http://www-1.ibm.com/servers/storage/disk/

Note: External I/O Drawers 7311-D11 and 7311-D20 are not supported on the p5-590 and p5-595 servers.

1.6.1 IBM 2104 Expandable Storage Plus

The IBM 2104 Expandable Storage Plus Model DS4 is a low-cost 3U disk subsystem that supports up to 14 Ultra320 SCSI disks from 36.4 GB up to 146.8 GB. This subsystem can be used in splitbus mode, meaning the bus with 14 disks could be split into two buses with seven disks each. In this configuration, one additional LPAR (using a dedicated adapter) could be provided with up to seven disks by using one Ultra3 SCSI adapter (FC 5712) or one Ultra3 SCSI RAID adapter (FC 5703).

1.6.2 IBM TotalStorage® Storage servers

The IBM TotalStorage DS4000 Storage server family consists of several models: Models DS4100, DS4300, DS4400, DS4500, and DS4800. The Model DS4100 is the smallest model that scales up to 14 TB, and Model DS4500 is the largest, which scales up to 32 TB of disk storage at the time this publication was written. The IBM TotalStorage DS4800 is the most powerful in the highly successful IBM TotalStorage DS4000 Series. The p5-590 or p5-595 server is connected to the TotalStorage Storage Servers using Fibre Channel, either directly, or over a storage area network (SAN).

1.6.3 IBM TotalStorage DS6000 and DS8000 series

The IBM TotalStorage DS6000 series (recently announced) is designed to deliver the resiliency, performance, and many of the key features of the IBM TotalStorage Enterprise Storage Server® (ESS) in an amazingly small, modular package. The DS6000 series can be scaled from 292 GB to 67.2 TB. The IBM TotalStorage DS8000 series are the high-end premier storage solutions for use in storage area networks. Created specifically for medium and large enterprises, the IBM TotalStorage DS8000 series offers high-capacity storage systems that are designed to deliver performance, scalability, resiliency, and value. The DS8000 series uses 64-bit IBM POWER5 microprocessors in dual 2-way (for the DS8100) or dual 4-way (for the DS8300) processor complexes to help reduce cycle times and accelerate response times, giving users fast access to vital information. The DS8000 series is designed to offer outstanding performance scalability—scaling up nearly linearly in disk, cache, and fabric infrastructure, with processor (2-way, 4-way, etc.). The physical storage capacity of the DS8000 series systems can range from 1.1 TB to 192 TB of physical capacity, and it has an architecture designed to scale up to a petabyte (one thousand terabytes). Both models would normally connect to the p5-590 and p5-595 using Fibre Channel—either directly, or over a storage area network (SAN).

1.7 Operating system support

The p5-590 and p5-595 servers are capable of running IBM AIX 5L for POWER, i5/OS and support appropriate versions of Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM eserver systems.

1.7.1 AIX 5L

If installing AIX 5L on the p5-590 and p5-595 servers, the following minimum requirements are needed:

- AIX 5L for POWER V5.2 with the 5200-04 Recommended Maintenance Package (APAR IY56722), or later, plus APAR IY60347
- AIX 5L for POWER V5.3 with APAR IY60349, or later

Note: The Advanced POWER Virtualization is not supported on AIX 5L for POWER Version 5.2; it requires AIX 5L Version 5.3.

IBM periodically releases maintenance packages for the AIX 5L operating system. These packages are available on CD-ROM or they can be downloaded from the Internet at:

http://www.ibm.com/servers/eserver/support/pseries/aixfixes.html

Information on how to obtain the CD-ROM can be found on the Web page mentioned above.

You can also get individual operating system fixes and information about obtaining AIX 5L service at this site. In AIX 5L Version 5.3, there is also the **suma** command available, which helps the administrator to automate the task of checking and downloading operating system downloads. For more information about the **suma** command functionality refer to:

http://techsupport.services.ibm.com/server/fixget

If you have problems downloading the latest maintenance level, ask your IBM Business Partner or IBM representative for assistance.

1.7.2 Linux

For the p5-590 and p5-595 servers, Linux distributions are available through Novell SUSE LINUX and Red Hat at the time this publication was written. The p5-590 and p5-595 servers require the following versions of Linux distributions:

- SUSE LINUX Enterprise Server 9 for POWER, or later
- Red Hat Enterprise Linux AS for POWER Version 3, or later

Note: Not all p5-590 and p5-595 server features available on the AIX 5L operating system are available on the Linux operating systems.

Note: Dynamic LPAR is not supported by Red Hat Enterprise Linux AS for POWER Version 3.

Information on features and external devices supported by Linux on the p5-590 or p5-595 can be found at:

http://www.ibm.com/servers/eserver/pseries/linux/

Information about SUSE LINUX Enterprise Server 9 can be found at:

http://www.novell.com/products/linuxenterpriseserver/

For information about Red Hat Enterprise Linux AS for pSeries from Red Hat, see:

http://www.redhat.com/software/rhel/details/

For the latest in IBM Linux news, subscribe to the Linux Line at:

https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=linuxline

Many of the features described in this document are operating system dependant and may not be available on Linux. For more information see:

http://www.ibm.com/servers/eserver/linux/power/whitepapers/linux_overview.html

Note: IBM only supports the Linux systems of clients with a SupportLine contract covering Linux. Otherwise, the Linux distributor should be contacted for support.

1.7.3 i5/OS V5R3

IBM i5/OS™ on *e*server p5 servers is intended for clients with a relatively small amount of i5/OS applications, whose focus and IT strategy are centered on UNIX. It is available as a solution for server consolidation for a partitioned system.

You cannot transfer an i5/OS license from an existing iSeries server to an *e*server p5 server running i5/OS partitions. The i5/OS licenses apply to the IBM *e*server p5 server itself and not to the underlying I/O subsystem. Also, you cannot order the i5/OS as the only operating system on *e*server p5 server.

i5/OS V5R3 or later has the following dependencies:

- Supported on 1.65 GHz POWER5 models only.
- Only one or two processors on the p5-590 and p5-595 systems can be dedicated to i5/OS.
- Only an AIX or Linux logical partition can be designated as the service partition. An i5/OS partition cannot be designated as the service partition on an *e*server p5 server.

You can create the following i5/OS partitions:

- One partition that uses one or two dedicated processors
- Two partitions that use one dedicated processor each
- One partition that uses uncapped shared processing units, with a maximum of two virtual processors for the partition
- Two partitions that use uncapped shared processing units, with a maximum of one virtual processor for each partition
- One partition that uses one dedicated processor and one partition that uses uncapped shared processing units, with a maximum of one virtual processor for the partition that uses uncapped shared processing units
- One partition that uses one dedicated processor and from one to ten partitions that use capped shared processing units, with a minimum of 0.10 processing units for each partition that uses capped shared processing units
- One partition that uses uncapped shared processors, with a maximum of one virtual processor for each partition that uses uncapped shared processing units, and from one to ten partitions that use capped shared processing units, with a minimum of 0.10 processing units for each partition that uses capped shared processing units

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1. The major components of this diagram are described in the following sections. The bandwidths provided throughout this section are theoretical maximums provided for reference. We always recommend that you obtain real-world performance measurements using production workloads.

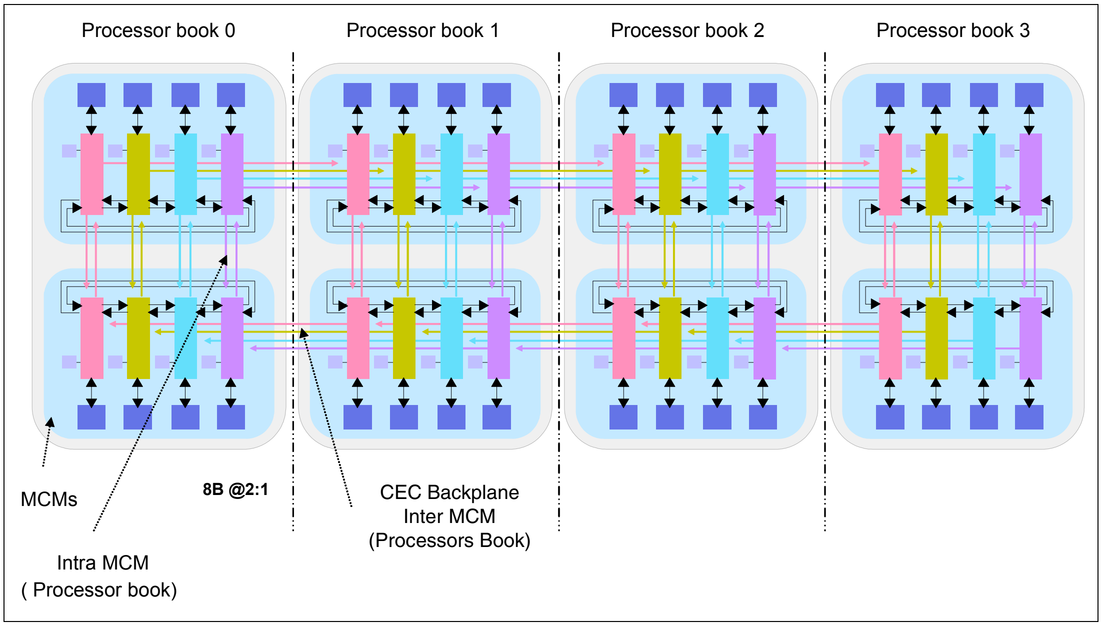


Figure 2-1 p5-590 and p5-595 64-way processor book

2.1 System design

Both the p5-590 and p5-595 servers are based on a modular design, where all components are mounted in 24-inch racks. Inside this rack all the server components are placed in specific positions. This design and mechanical organization offers advantages in optimization of floor space usage.

There are three major subsystems:

- The Central Electronics Complex (CEC)
- The power subsystem
- The I/O subsystem

In addition, all these components are managed by an external management workstation, called the Hardware Management Console (HMC).

2.1.1 Central Electronics Complex

The Central Electronics Complex is an 18 EIA unit drawer that houses:

- 1 to 4 processors books (nodes)

  The processor book contains the POWER5 processors, the L3 cache modules located in Multichip modules, and memory and RIO-2 attachment cards.

- CEC backplane (double-sided passive backplane) that serves as the system component mounting unit

  Processor books plug into the front side of the backplane. The node distributed converter assemblies (DCA) plug into the back side of the backplane. The DCAs are the power supplies for the individual processor books.

  A Fabric bus structure on the backplane board provides communication between books.

- Service processor unit

  Located in the panel above the distributed converter assemblies (DCA). It contains redundant service processors and Oscillator cards.

- Remote I/O (RIO) adapters to support attached I/O drawers
- Fans and blowers for CEC cooling
- Light strip (front, rear)

Figure 2-2 on page 17 provides a logical view of the CEC components.

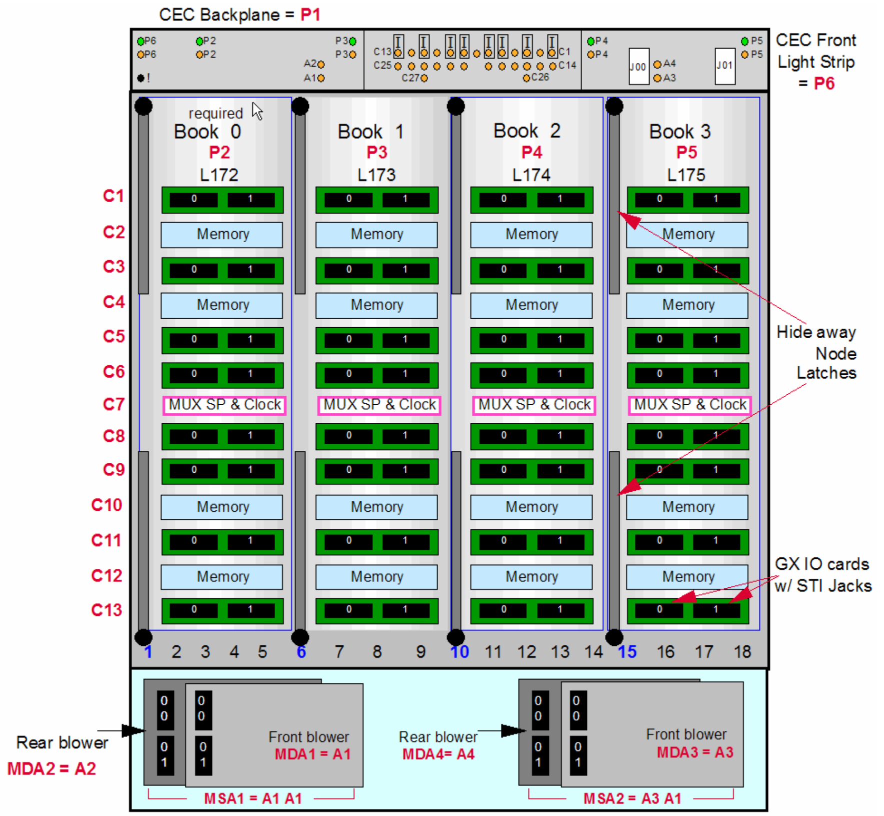


Figure 2-2 CEC components

2.1.2 CEC backplane

The top view of p5-595 CEC is shown in Figure 2-3. There are no physical differences between the p5-590 backplane and the p5-595 backplane.

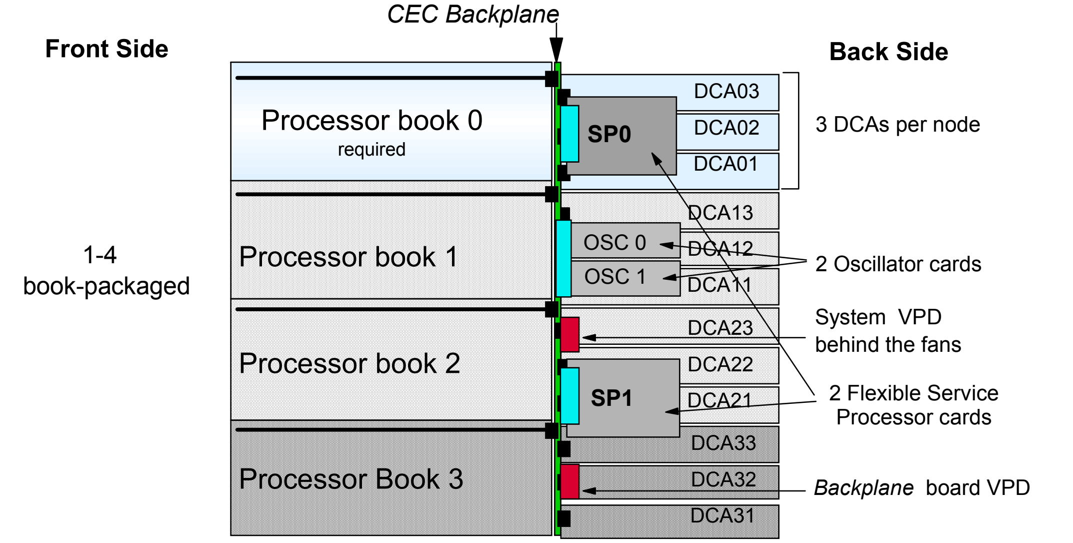


Figure 2-3 p5-595 backplane

Note: In the p5-590 configuration, book 2 and book 3 are not available.

The backplane is positioned vertically in the center of the CEC, and provides mount spaces for processor books. This is a double-sided passive backplane. Figure 2-3 on page 17 depicts component population on both the front side and back side of the backplane.

The CEC backplane provides the following types of slots:

- Slots for up to four processor books

  Processor books plug into the front side of the backplane and are isolated into their own power planes, which allows the system to power on/off individual nodes within the CEC.

- Slots for up to 12 distributed converter assemblies DCA

  Three DCAs per processor book provide N+1 logic redundancy. The DCA trio is located on the rear CEC, behind the processor book they support.

- Fabric bus for communications between processor books

Located in the panel above the CEC DCAs are:

- Service processor and OSC unit assembly
- VPD card

2.1.3 Processor books

In the p5-590 and p5-595 systems, the POWER5 chip has been packaged with the L3 cache chip into a cost-effective multi-chip module package. The storage structure for the POWER5 processor chip is a distributed memory architecture that provides high-memory bandwidth. Each processor can address all memory and sees a single shared memory resource. As such, two MCMs with their associated L3 cache and memory are packaged on a single processor book. Access to memory behind another processor is accomplished through the fabric buses. The p5-590 supports up to two processor books (each book is a 16-way), and the p5-595 supports up to four processor books. Each processor book has dual MCMs containing POWER5 processor chips and 36 MB L3 modules. Each 16-way processor book also includes 16 slots for memory cards and six remote I/O attachment cards (RIO-2) for connection of the system I/O drawers, as shown in Figure 2-13 on page 31.

2.1.4 The POWER5 chip

The POWER5 chip features single and simultaneous multithreading execution. The POWER5 processor maintains both binary and architectural compatibility with existing POWER4™ processor-based systems and is designed to allow binaries to continue executing properly and application optimizations to carry forward to newer systems.

The POWER5 microprocessor provides additional enhancements such as virtualization; simultaneous multithreading support; improved reliability, availability, and serviceability at both chip and system levels; and it has been designed to support interconnection of 64 processors along with higher clock speeds.

Figure 2-4 on page 19 shows the high-level structures of POWER5 processor-based systems. The POWER5 processor supports a 1.9 MB on-chip L2 cache, implemented as three identical slices with separate controllers for each. Either processor core can independently access each L2 controller. The available L3 cache with a capacity of 36 MB operates as a backdoor with separate buses for reads and writes that operate at half processor speed.

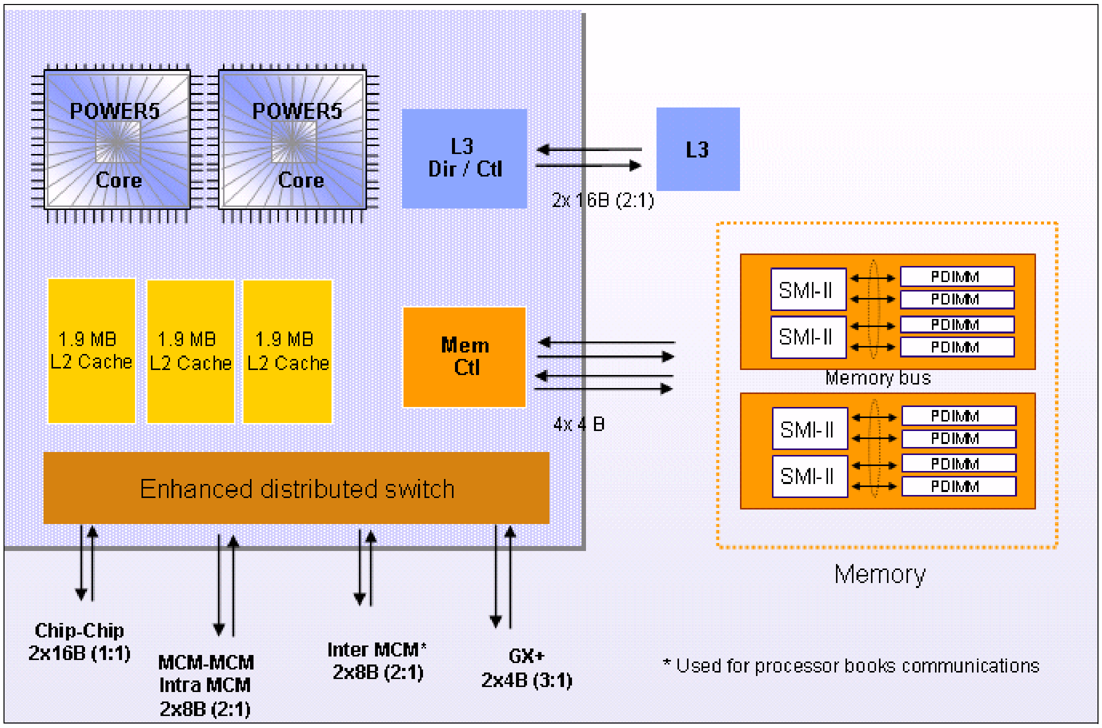


Figure 2-4 POWER5 processor

The storage structure for the POWER5 chip is a distributed memory architecture that provides high memory bandwidth, although each processor can address all memory and sees a single shared memory resource. The processors are interfaced to eight memory slots, controlled by two SMI-2 controllers, which are located in close physical proximity to the processor book modules. I/O connects to the p5-590 and p5-595 processor module using the GX+ bus. The processor module provides a single GX+ bus. The GX+ bus provides an interface to I/O devices through the RIO-2 connections.

Table 2-1 highlights the differences between the POWER4 and POWER5 processors.

Table 2-1 POWER4 to POWER5 comparison

	POWER4 design	**POWER5 design**
L1 data cache	2-way set-associative FIFO[a]	4-way set-associative LRU[b]
L2 cache	8-way set-associative 1.44 MB	10-way set-associative 1.9 MB
L3 cache	32 MB 118 clock cycles	36 MB ~80 clock cycles
Chip interconnect type Intra MCM data bus Inter MCM data bus	Distributed switch 1/2 processor speed 1/2 processor speed	Enhanced distributed switch processor speed 1/2 processor speed
Memory bandwidth	4 GB/s per chip	~16 GB/s per chip
Simultaneous multithreading	No	Yes
Processor addressing	1 processor	1/10th of processor
Dynamic power management	No	Yes

	POWER4 design	POWER5 design
Size	412 mm	389 mm

a. FIFO stands for First In First Out.
b. LRU stands for Least Recently Used.

Figure 2-5 shows the high-level structures of POWER5 processor-based systems. The POWER4 processors scale up to a 32-way symmetric multiprocessor. Going beyond 32 processors with POWER4 architecture could increase interprocessor communication, resulting in higher traffic on the interconnection fabric bus. This can cause greater contention and negatively affect system scalability.

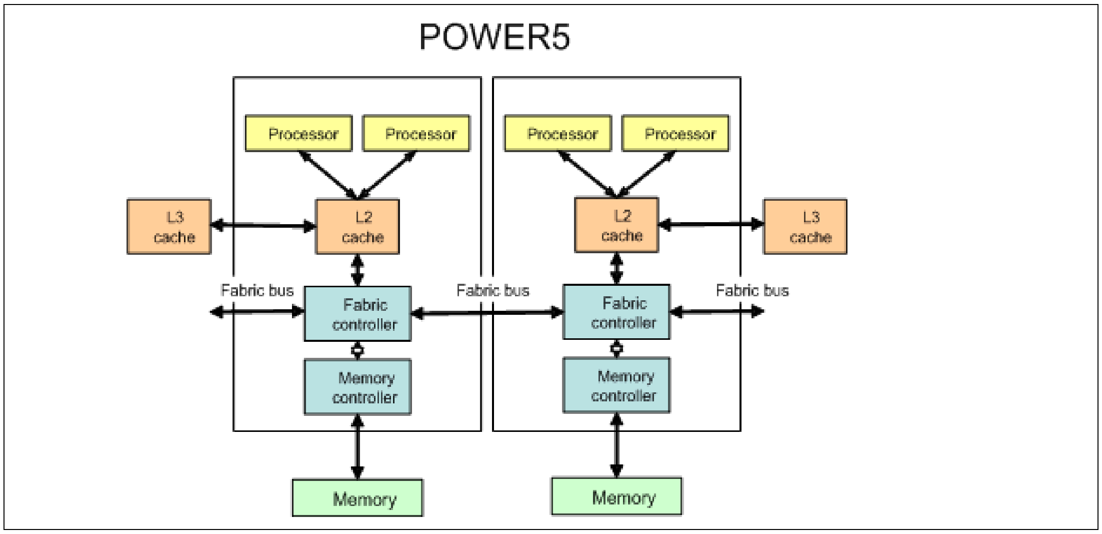


Figure 2-5 POWER5 system structures

Moving the L3 cache reduces traffic on the fabric bus and enables POWER5 processor-based systems to scale to higher levels of symmetric multi-processing. The POWER5 processor supports a 1.9 MB on-chip L2 cache, implemented as three identical slices with separate controllers for each. Either processor core can independently access each L2 controller. The L3 cache, with a capacity of 36 MB, operates as a backdoor with separate buses for reads and writes that operate at half the processor speed.

Because of the higher transistor density of the POWER5 0.13-µm technology (over the original POWER4), it was possible to move the memory controller on-chip and eliminate a chip that was previously needed for the memory controller function. These changes in the POWER5 processor also have the significant side benefits of reducing latency to the L3 cache and main memory, as well as reducing the number of chips that are necessary to build a system.

The POWER5 processor supports the 64-bit PowerPC® architecture. A single die contains two identical processor cores, each supporting two logical threads. This architecture makes the chip appear as a four-way symmetric multiprocessor to the operating system. The POWER5 processor core has been designed to support both enhanced simultaneous multithreading and single-threaded (ST) operation modes.

POWER5 design provides additional enhancements such as virtualization, reliability, availability, and serviceability (RAS) features at both chip and system levels.

Key enhancements introduced into the POWER5 processor and system design include:

- Simultaneous multithreading
- Dynamic resource balancing to efficiently allocate system resources to each thread
- Software-controlled thread prioritization
- Dynamic power management to reduce power consumption without affecting performance
- Micro-Partitioning technology
- Virtual storage, virtual Ethernet
- Enhanced scalability, parallelism
- Enhanced memory subsystem

2.1.5 Multichip module and system interconnect

POWER5 chips can be packaged in several ways such as multichip module (MCM), dual chip module (DCM), or mounted on a system planar.

MCMs are used as basic building blocks on high-end SMPs. Each MCM is an eight-way building block. The p5-590 and p5-595 MCMs have four POWER5 chips and four L3 cache chips each.

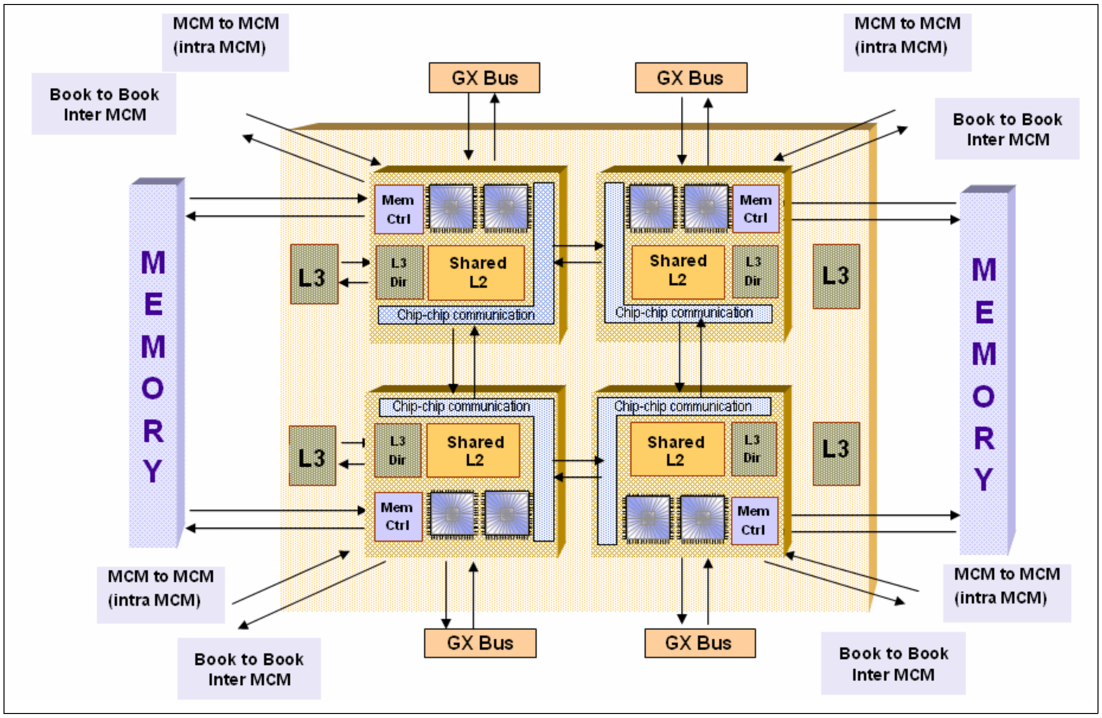


Figure 2-6 MCM logical view

Figure 2-7 on page 22 shows an actual picture of a POWER5 MCM.

Each MCM houses four POWER5 chips (eight processor cores) that are connected through chip-to-chip ports, with their associated L3 chips. The POWER5 chips are mounted on the MCM such that they are all rotated 90 degrees from one another. This arrangement minimizes the interconnect distances, which improves the speed of the inter-chip communication. There are separate communication buses between processors in the same MCM, and processors in different MCMs, as shown in Figure 2-6.

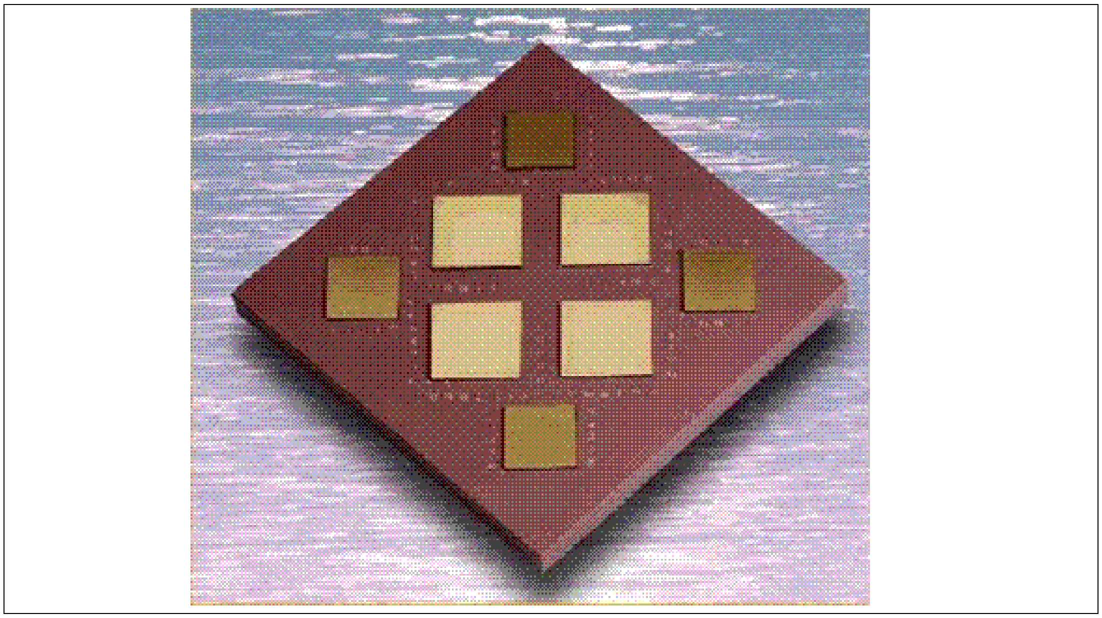


Figure 2-7 Multichip module

Two POWER5 MCMs can be tightly coupled to form a book, as shown in Figure 2-8. These books are interconnected again to form larger SMPs, up to 64-way. The MCMs and books can be interconnected to form 8-way, 16-way, 32-way, 48-way, and 64-way SMPs with one, two, four, six, and eight MCMs, respectively.

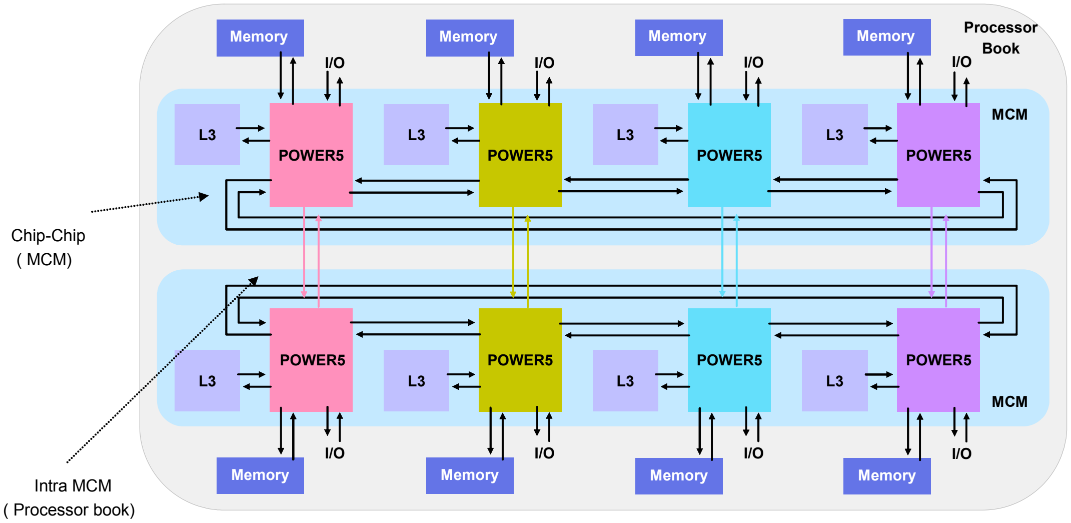


Figure 2-8 16-way processor book interconnect layout

POWER5 processor is not just a chip, but rather an architecture of how a set of chips is designed together to build a system. As such, POWER5 can be considered a technology in its own right. POWER5 exploits the enhanced distributed switch for interconnects. In that light, systems are built by interconnecting POWER5 chips to form up to 64-way symmetric multiprocessors.

The connecting buses between the MCMs exploit an enhanced version of the distributed switch from the POWER4 processor. All chip interconnections operate at half-processor frequency and scale with processor frequency. Intra-MCM buses have been enhanced from

Power4 to allow operation at full processor speeds. The inter-MCM buses continue to operate at half-processor speeds. Figure 2-8 on page 22 shows the processor book interconnect layout.

Figure 2-1 on page 15 provides an interconnect layout for a 64-way p5-595 system.

2.1.6 Simultaneous multithreading

To provide improved performance at the application level, simultaneous multithreading functionality is embedded in the POWER5 chip technology. Applications developed to use process level parallelism (multi-tasking) and thread-level parallelism (multi-threads) can shorten their overall execution time. Simultaneous multithreading is the next stage of processor saturation, for throughput-oriented applications, to introduce the method of instruction-level parallelism to support multiple pipelines to the processor.

The simultaneous multithreading mode maximizes the usage of the execution units. In the POWER5 chip, more rename registers have been introduced (for floating-point operation, rename registers increased to 120), which are essential for out-of-order execution, and then vital for simultaneous multithreading.

If simultaneous multithreading is activated:

- More instructions can be executed at the same time.
- The operating system views twice the number of physical processors installed in the system.
- Support is provided in mixed environments:
  - Capped and uncapped partitions
  - Virtual partitions
  - Dedicated partitions
  - Single partition systems

> **Note:** Simultaneous multithreading is supported on POWER5 processor-based systems running AIX 5L Version 5.3 or Linux operating system-based systems at an appropriate level. AIX 5L Version 5.2 does not support this function.

IBM has documented simultaneous multithreading performance benefit at 30 percent.

For more information, see the following URL:

`http://www.ibm.com/servers/eserver/pseries/hardware/system_perf.html`

The simultaneous multithreading policy is controlled by the operating system and is thus partition specific. AIX 5L provides the **`smtctl`** command that turns simultaneous multithreading on and off either immediately or on next reboot. For a complete listing of flags, see:

`http://publib.boulder.ibm.com/infocenter/pseries/index.jsp`

For Linux, an additional boot option must be set to activate simultaneous multithreading after a reboot.

Enhanced simultaneous multithreading features

To improve simultaneous multithreading performance for various workloads and provide robust quality of service, the POWER5 processor provides two features:

- Dynamic resource balancing

  Dynamic resource balancing is designed to ensure that the two threads executing on the same processor flow smoothly through the system. Depending on the situation, the POWER5 processor resource balancing logic has different thread throttling mechanisms (a thread reaching a threshold of L2 cache misses will be throttled to allow other threads to pass the stalled thread).

- Adjustable thread priority

  Adjustable thread priority allows software to determine when one thread should have a greater (or lesser) share of execution resources. The POWER5 processor supports eight software-controlled priority levels for each thread.

Single threading operation

Having threads executing on the same processor will not increase the performance of applications with execution unit limited performance, or applications that consume all the chip's memory bandwidth. For this reason, the POWER5 processor supports the single threading execution mode. In this mode, the POWER5 processor gives all the physical resources to the active thread, allowing it to achieve higher performance than a POWER4 processor based-system at equivalent frequencies. Highly optimized scientific codes are one example where single threading operation may provide more throughput.

2.1.7 Dynamic power management

In current Complementary Metal Oxide Semiconductor (CMOS) technologies, chip power is one of the most important design parameters. With the introduction of simultaneous multithreading, more instructions execute per cycle per processor core, thus increasing the core's and the chip's total switching power. To reduce switching power, POWER5 chips use a fine-grained, dynamic clock gating mechanism extensively. This mechanism gates off clocks to a local clock buffer if dynamic power management logic knows the set of latches driven by the buffer will not be used in the next cycle. This allows substantial power saving with no performance impact. In every cycle, the dynamic power management logic determines whether a local clock buffer that drives a set of latches can be clock gated in the next cycle.

2.1.8 Available processor speeds

The p5-590 system features base 8-way (CoD), 16-way, and 32-way configurations with the POWER5 processor running at 1.65 GHz. The p5-595 system features base 16-way, 32-way, 48-way, and 64-way configurations with the POWER5 processor running at 1.65 GHz and 1.9 GHz.

> **Note:** Any p5-595 or p5-590 system made of more than one processor book must have all processors running at the same speed.

To determine the processor characteristics on a running system, use one of the following commands:

lsattr -El proc*X* Where *X* is the number of the processor; for example, proc0 is the first processor in the system. The output from the command[1] would be similar to this:

[1] The output of the **lsattr** command has been expanded with AIX 5L to include the processor clock rate.

```
frequency 165640000     Processor Speed       False
smt_enabled true        Processor SMT enabled False
smt_threads 2           Processor SMT threads False
state enable            Processor state       False
type powerPC_POWER5     Processor type        False
```

(False, as used in this output, signifies that the value cannot be changed through an AIX 5L command interface.)

pmcycles -m This command (AIX 5L Version 5.3 and later) uses the performance monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. This is the sample output of a 16-way p5-590 running at 1.65 GHz system:

```
Cpu 0 runs at 1656 MHz
Cpu 1 runs at 1656 MHz
Cpu 2 runs at 1656 MHz
Cpu 3 runs at 1656 MHz
...
Cpu 13 runs at 1656 MHz
Cpu 14 runs at 1656 MHz
Cpu 15 runs at 1656 MHz
```

2.2 System flash memory configuration

In the p5-590 and p5-595, a serial electronically erasable programmable read-only memory (sEEPROM) adapter plugs into the back of the Central Electronics Complex backplane. The platform firmware binary image is programmed into the system sEEPROM, also known as *system FLASH memory*. FLASH memory is initially programmed during manufacturing of the p5-590 and p5-595 systems. However, this single binary image can be reprogrammed to accommodate firmware fixes provided to the client using the hardware management console.

The firmware binary image contains boot code for the p5-590 and p5-595. This boot code includes, but is not limited to, system service processor code; code to initialize the POWER5 processors, memory, and I/O subsystem components; partition management code; and code to support the virtualization features. The firmware binary image also contains hardware monitoring code used during partition runtime.

During boot time, the system service processor dynamically allocates the firmware image from flash memory into main system memory. The firmware code is also responsible for loading the operating system image into main memory.

2.2.1 Vital product data and system smart chips

Vital product data (VPD) carries all of the necessary information for the service processor to determine if the hardware is compatible and how to configure the hardware and chips on the card. The VPD also contains the part number and serial number of the card used for servicing the machine, as well as the location information of each device for failure analysis. Since the VPD in the card carries all information necessary to configure the card, no card device drivers or special code has to be sent with each card for installation.

Smart chips are micro-controllers used to store vital product data (VPD). The smart chip provides a means for securely storing data that cannot be read, altered, or written other than by IBM privileged code. The smart chip provides a means of verifying IBM eServer On/Off Capacity on Demand and IBM eServer Capacity Upgrade on Demand activation codes that only the smart chip on the intended system can verify. This allows clients to purchase additional spare capacity and pay for use only when needed. The smart chip is the basis for the CoD function and verifying the data integrity of the data stored in the card.

2.3 Light strip

There is no operator panel on p5-590 and p5-595 servers. The p5-590 and p5-595 have a light strip on both the front and the rear of the system unit. The light strips contain several LEDs, each representing the status of a particular field replaceable unit (FRU) or component. Under normal system operating conditions:

- No amber LEDs are lit. An amber LED that is lit indicates a problem with the component associated with that LED.
- If an active component has a green LED associated with it, that LED is lit.
- Processor Books, oscillator cards, SP cards, and the light strips themselves are FRUs for which both green and amber LEDS are assigned. If, for example, PU Book 3 is not active (as would be the case in a 48-way system), both the green and amber LEDs would be unlit.
- If the System Attention LED is lit, a serviceable event has been detected and recorded by the system.

A light strip is composed of a printed circuit card mounted on a plastic bezel.

Note: The following abbreviations are used:

- Proc. book - Processor book
- AMD - Air moving device (blower)
- MCM - Multichip module
- OSC - Oscillator
- SP - Service processor
- DCA - Distributed converter assembly
- CEC - Central electronic complex

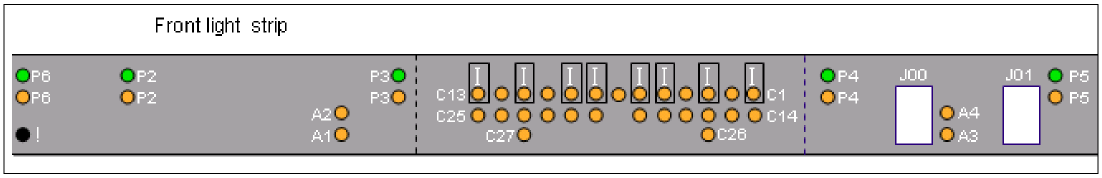


Figure 2-9 p5-590 and p5-595 front light strip

Figure 2-9 shows the front light strip. The following three tables (left, middle, and right sections) help identify LED meanings. Table 2-2, Table 2-3 on page 27, and Table 2-4 on page 27 are grouped by light strip section.

Table 2-2 Front light strip left section

Front light strip left section		
P6 (upper) Front light strip (green)	P2 (upper) Proc. book 0 (green)	A2- AMD2 (amber)

Front light strip left section		
P6 (lower) Front light strip (amber)	P2 (lower) Proc. book 0 (amber)	A1- AMD1 (amber)
!- system System attention	P3 (upper) Proc. book 1 (green)	P3 (lower) Proc. book 1 (amber)

Table 2-3 Front light strip middle section

Front light strip middle section (all amber)		
C27- M1 MCM	C13 - D8 RIO adapter	C12 - MC04 memory card
C26 - M0 MCM	C11 - D7 RIO adapter	C10 - MC03 memory card
C9 - D6 RIO adapter	C8 - D5 RIO adapter	C7 mux card
C6 - D4 RIO adapter	C5 - D3 RIO adapter	C4 - MC02 memory card
C3 - D2 RIO adapter	C2 - MC04 memory card	C1 - D1 RIO adapter
C25 - MC16 memory card	C24 - MC15 memory card	C23 - MC14memory card
C22 - MC13 memory card	C21 - MC12 memory card	C20 - MC11 memory card
C19 - MC10 memory card	C18 - MC09 memory card	C17 - MC08 memory card
C16 - MC07 memory card	C15 - MC06 memory card	C14 - MC05 memory card

Table 2-4 Front light strip right section

Front light strip right section		
P4 (upper) Proc. book 2 (green)	A4- AMD 4 (amber)	P5 (upper) Proc. book 3 (green)
P4 (lower) Proc. book 2 (amber)	A3- AMD3 (amber)	P5(lower) Proc. book 3 (amber)

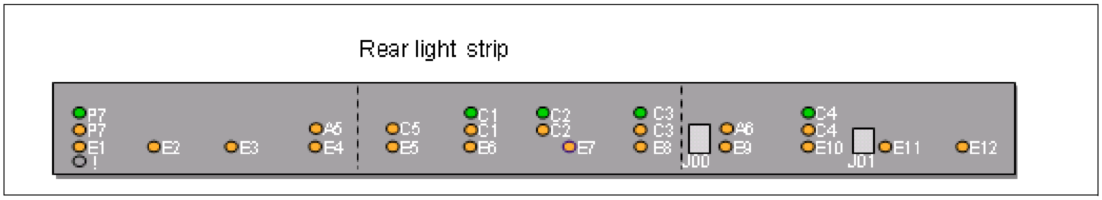


Figure 2-10 p5-590 and p5-595 rear light strip

Figure 2-10 shows the rear light strip. The following three tables (left, middle, and right sections) help identify LED meanings. Table 2-5, Table 2-6 on page 28, and Table 2-7 on page 28 are grouped by light strip section.

Table 2-5 Rear light strip left section

Rear light strip left section		
P7 (upper) Rear light strip (green)	A5- AMD5 (amber)	E1 - DCA 30 (amber)

Rear light strip left section		
P7 (lower) Rear light strip (amber)	E2 - DCA 31 (amber)	E3 - DCA 32 (amber)
!- system System attention	E4 - DCA 20 (amber)	

Table 2-6 Rear light strip middle section

Rear light strip middle section		
C1 (upper) SP 1 (green)	C2 (upper) OSC 1 (green)	C3 (upper) OSC 2 (green)
C1 (lower) SP 1 (amber)	C2 (lower) OSC 1 (amber)	C3 (lower) OSC 2 (amber)
C5 - Anchor 1 (amber)	E5 - DCA 21 (amber)	E6- DCA 22 (amber
P1 - CEC backplane	E7 - DCA 10 (amber)	E8 - DCA 11 (amber)

Table 2-7 Rear light strip right section

Rear light strip right section		
A6- AMD6 (amber)	C4 (upper) SP 0 (green)	E9 - DCA 12 (amber)
E10 - DCA 00 (amber)	C4 (lower) SP 0 (amber)	E11 - DCA 01 (amber)
E12 - DCA 02(amber)		

2.4 Memory subsystem

The p5-590 and p5-595 memory controllers are internal to the POWER5 chip. The memory controller interfaces to four Synchronous Memory Interface II (SMI-II) buffer chips and eight DIMM cards per processor chips, as shown in Figure 2-11 on page 29. There are 16 memory card slots per processor book and each processor chip on an MCM owns a pair of memory cards.

The p5-590 and p5-595 use Double Data Rate (DDR) DRAM memory cards. The two types of DDR memory used are DDR1 and the higher-speed DDR2. Memory migration from previous systems is not supported.

Note: Because the DDR1 and DDR2 modules use different voltages, mixing of the memory technologies is not allowed within a server.

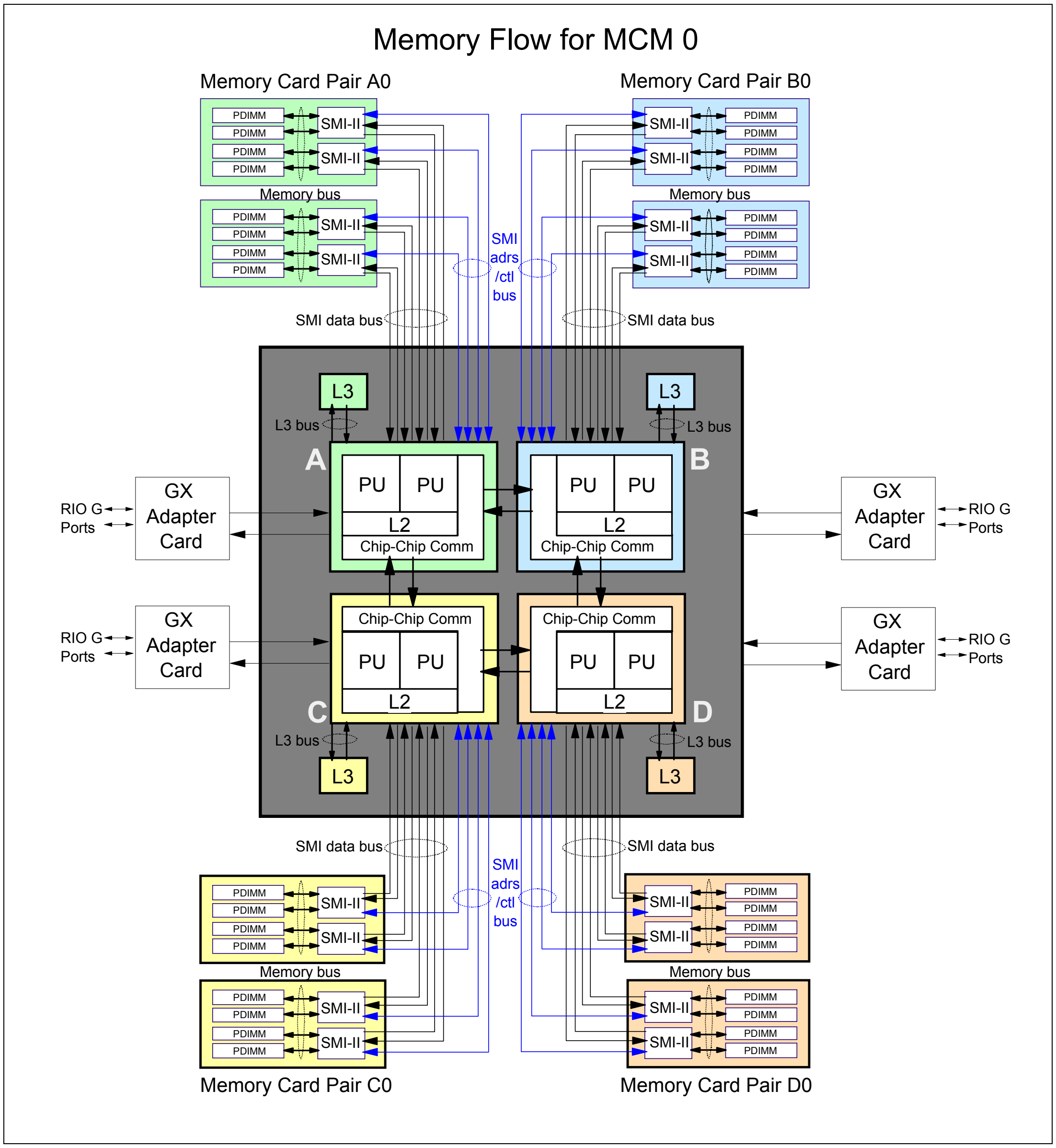


Figure 2-11 Memory flow diagram for MCM0

2.4.1 Memory cards

On the p5-590 and the p5-595 systems the memory is seated on a memory card, shown in Figure 2-12 on page 30. Each memory card has four soldered DIMM cards and two SMI-II chips for address/controls and data buffers. Individual DIMM cards cannot be removed or added, and memory cards have a fixed amount of memory.

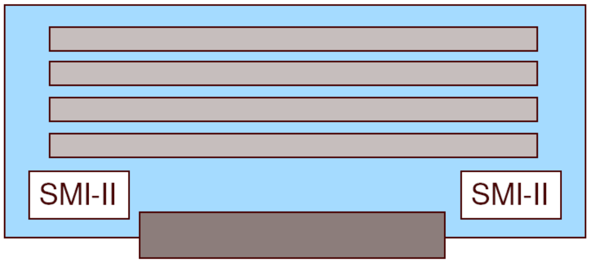


Figure 2-12 Memory card with four DIMM slots

The memory features that are available for the p5-590 and the p5-595 at the time of writing are listed in Table 2-8.

Table 2-8 Types of available memory cards for p5-590 and p5-595

Memory type	Size	Speed	Number of memory cards	Feature code
DDR1 COD	4 GB (2 GB active)	266 MHz	1	7816
	8 GB (4 GB active)	266 MHz	1	7835
DDR1	16 GB	266 MHz	1	7828
	32 GB	200 MHz	1	7829
	256 GB package	266 MHz	32 * 8 GB	8195
	512 GB package	266 MHz	32 * 16 GB	8197
	512 GB package	200 MHz	16 * 32 GB	8198
DDR2	4 GB	533 MHz	1	7814

2.4.2 Memory configuration and placement

The minimum memory for a p5-590 processor-based system is 2 GB and the maximum installable memory is 1024 GB using DDR1 memory DIMM technology (128 GB using DDR2 memory DIMM). The total memory depends on the number of available processors (16 per processor book).

The minimum memory for a p5-595 processor-based system is 8 GB and the maximum installable memory is 2,048 GB using DDR1 memory DIMM technology (256 GB using DDR2 memory DIMM). The total memory depends on the number of available processors.

Table 2-9 lists the possible memory configurations.

Table 2-9 Memory configuration table

System	p5-590	p5-595
Min. configurable memory	8 GB	8 GB
Max. configurable memory using DDR1 memory	1,024 GB	2,048 GB
Max. configurable memory using DDR2 memory	128 GB	256 GB
Max. number of memory cards	32	64

The memory locations for each processor chip in the MCMs are illustrated in Figure 2-13.

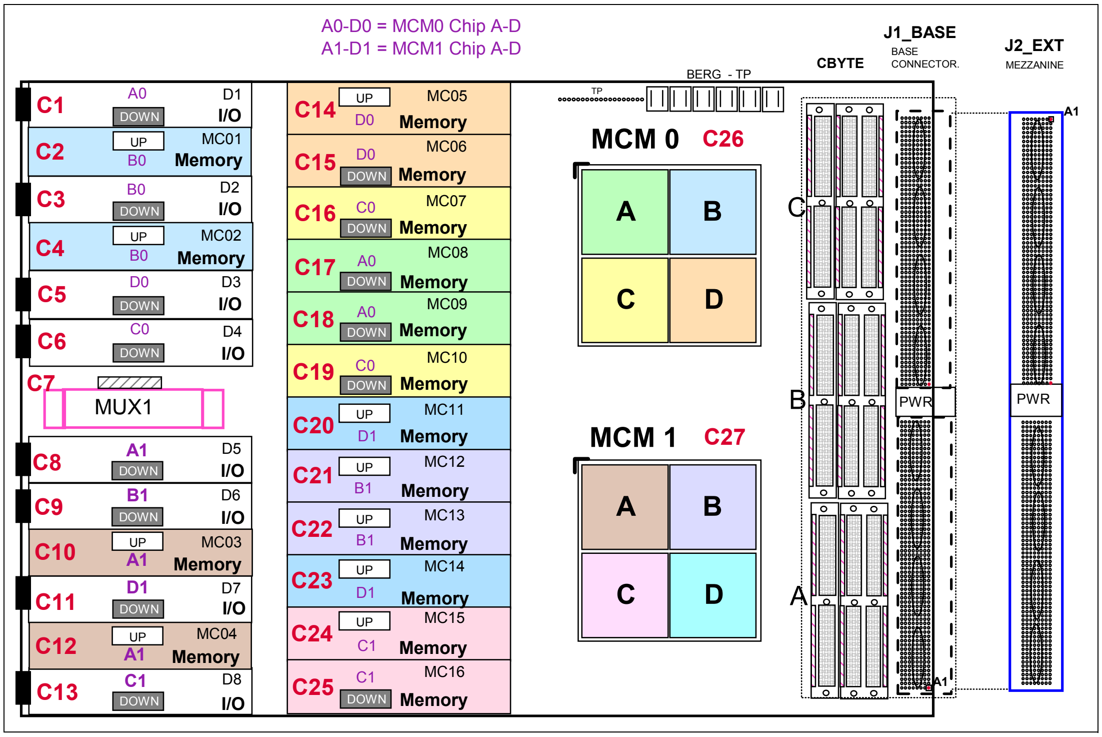


Figure 2-13 Memory placement for the p5-590 and p5-595

The following rules *must* be observed:

- Memory must be installed in identical pairs.
- Servers with one processor book must have a minimum of two memory cards installed.
- Servers with two processor books must have a minimum of four memory cards installed per processor book (two per MCM).

The following memory configuration guidelines are *recommended*:

- The same amount of memory should be used for each MCM (two per processor book) in the system.
- Each 8-way MCM (two per processor book) should have some memory.
- No more than two different sizes of memory cards should be used in each processor book.
- All MCMs (two per processor book) in the system should have the same aggregate memory size.
- A minimum of half of the available memory slots in the system should contain memory.
- It is better to install more cards of smaller capacity than fewer cards of larger capacity. Cards with larger capacity would occupy fewer slots, providing room for future expansion and the re-use of the currently installed memory.

For p5-590 and p5-595 servers being used for high-performance computing, the following are *strongly recommended*:

- Use DDR2 memory.
- Install some memory in support of each 8-way MCM (two MCMs per processor book).
- Use the same sized memory cards across all MCMs and processor books in the system.

2.4.3 Memory throughput

Figure 2-14 illustrates the peak bandwidths per processor chip.

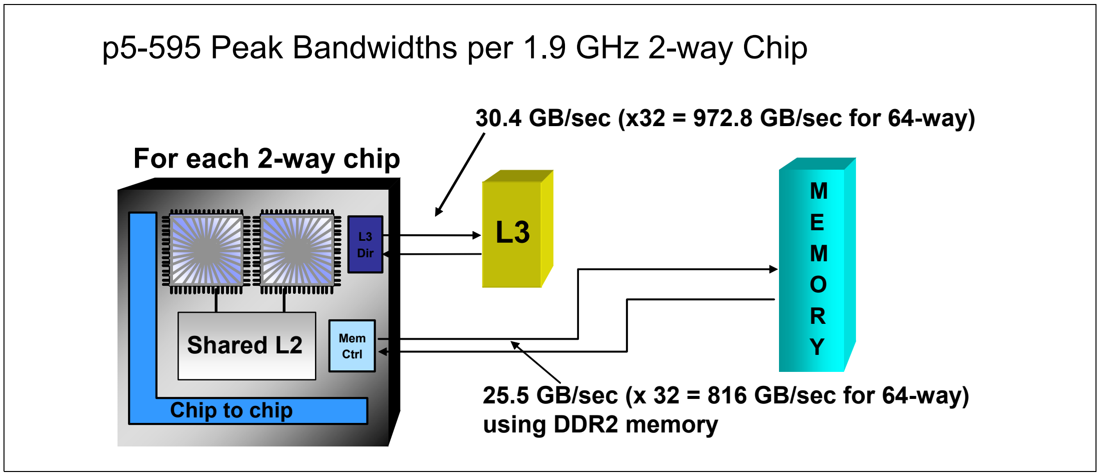

p5-595 Peak Bandwidths per 1.9 GHz 2-way Chip
30.4 GB/sec (x32 = 972.8 GB/sec for 64-way)
For each 2-way chip
L3
Dir
L3
Shared L2
Mem
Ctrl
Chip to chip
M
E
M
O
R
Y
25.5 GB/sec (x 32 = 816 GB/sec for 64-way)
using DDR2 memory

Figure 2-14 Peak bandwidths per processor chip

The L3 cache, with a capacity of 36 MB, operates as a backdoor with separate buses for reads and writes that operate at half the processor speed. There is a 16 byte read bus and a 16 byte write bus. Therefore the total bandwidth of the L3 bus on a 1.9 GHz processor will be (32 x 0.95 x 10^9) = 30.4 GB/s.

A DDR bus allows double reads or writes per clock cycle. There are four 4 byte read buses and four 2 byte write buses between each chip, and an associated pair of memory giving a total of 24 bytes available for simultaneous read and write operations. Since DDR2 runs at 533 MHz, this provides a total bandwidth of (2 x 24 x 533 x 10^6) = 25.5 GB/s.

Note: The throughput calculations use the definition of 1 GB = 10^9 bytes.

2.5 System buses

The following sections provide additional information related to the internal GX buses. For information regarding other buses refer to 2.1.4, “The POWER5 chip” on page 18.

2.5.1 GX+ and RIO-2 buses

The processor module provides a GX+ bus that is used to connect to the I/O subsystem.

- Each processor book has eight GX+ slots, which support communication with the I/O drawer.
- Each processor on an MCM may own 1 GX+ I/O Card.
- Each GX+ I/O Card has two RIO-2 ports.

  Remote I/O (RIO-2) links allow for connectivity to external I/O drawers and PCI-X technology.
- Figure 2-10 on page 33 provides a GX+ card layout in relation to the MCM core.

Table 2-10 GX to MCM relation

RIO card	MCM	Chip
C1	0	A
C3	0	B
C5	0	D
C6	0	C
C8	1	A
C9	1	B
C11	1	D
C13	1	C

Note: GX+ bus clock frequency shows a CPU to GX+ ratio of 3:1.

2.6 Internal I/O subsystem

The p5-590 and p5-595 use remote I/O drawers (that are 4U) for directly attached PCI or PCI-X adapters and SCSI disk capabilities. A minimum of one I/O drawer (FC 5791 or FC 5794) is required per system.

Note: The p5-590 supports up to eight I/O drawers, while the p5-595 supports up to twelve I/O drawers.

2.6.1 I/O drawer

The I/O drawers provide internal storage and I/O connectivity to the system. Figure 2-15 on page 34 shows a view of an I/O drawer, with the PCI slots and riser cards that connect to the RIO ports in the I/O books.

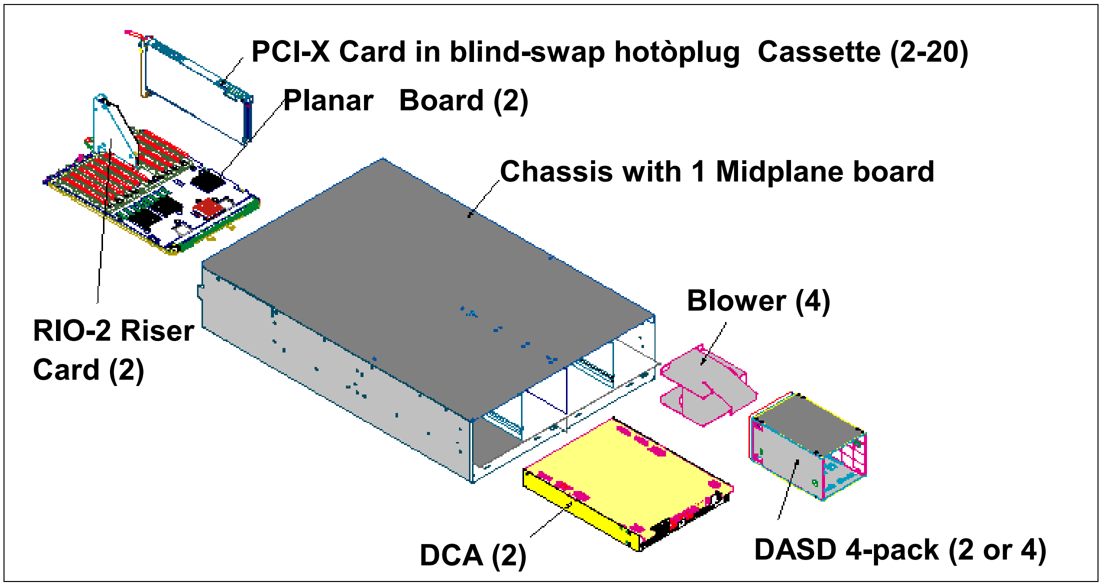

PCI-X Card in blind-swap hotòplug Cassette (2-20)
Planar   Board (2)
Chassis with 1 Midplane board
Blower (4)
RIO-2 Riser Card (2)
DCA (2)
DASD 4-pack (2 or 4)

Figure 2-15   I/O drawer details

Each I/O drawer is divided into two separate halves. Each half contains 10 blind-swap PCI-X slots (3.3 volt) and one or two Ultra3 SCSI 4-pack backplanes for a total of 20 PCI slots and up to 16 hot-swap disk bays per drawer (these internal SCSI backplanes do not support external SCSI device attachments). Each half of the I/O drawer is powered separately.

Existing 7040-61D I/O drawers may be attached to a p5-590 and p5-595 server as additional I/O drawers, if available.

- Only 7040-61D I/O drawers containing FC 6571 PCI-X planars are supported. Any FC 6563 PCI planars must be replaced with FC 6571 PCI-X planars before the drawer can be attached.
- Only adapters supported on the p5-590 I/O drawers are supported in 7040-61D I/O drawers, if attached. Unsupported adapters must be removed before attaching the drawer to the p5-590 server.

A minimum of one I/O drawer (FC 5791 or FC 5794) is required per system. I/O drawer FC 5791 contains 20 PCI-X slots and 16 disk bays, and FC 5794 contains 20 PCI-X slots and eight disk bays.

A maximum of eight I/O drawers can be connected to a p5-590. Fully configured, the p5-590 can support 160 PCI adapters and 128 disks at 15,000 rpm.

A maximum of 12 I/O drawers can be connected to a p5-595. Fully configured, the p5-595 can support 240 PCI adapters and 192 disks at 15,000 rpm.

A blind-swap hot-plug cassette (equivalent to those in FC 4599) is provided in each PCI-X slot of the I/O drawer. Cassettes not containing an adapter will be shipped with a plastic filler card installed to help ensure proper environmental characteristics for the drawer. If additional blind-swap hot-plug cassettes are needed, FC 4599 should be ordered.

All 10 PCI-X slots on each I/O drawer planar are capable of supporting either 64-bit or 32-bit PCI or PCI-X adapters. Each I/O drawer planar provides 10 PCI-X slots capable of supporting 3.3 V signaling PCI or PCI-X adapters operating at speeds up to 133 MHz.

2.6.2 I/O drawer attachment

System I/O drawers are connected to the p5-590 and p5-595 CEC using RIO-2 loops. Drawer connections are made in loops to help protect against a single point-of-failure resulting from an open, missing, or disconnected cable. If a fault is detected the system can reduce the speed on a cable, or disable part of the loop. Systems with non-looped configurations could experience degraded performance and serviceability. The system has a non-looped configuration if only one RIO-2 path is running.

RIO-2 loop connections operate at 1 GHz. RIO-2 loops connect to the system CEC using RIO-2 loop attachment adapters (FC 7818). Each of these adapters has two ports and can support one RIO-2 loop. Up to six of the adapters can be installed in each 16-way processor book. Up to 8 or 12 I/O drawers can be attached to the p5-590 or p5-595, depending on the model and attachment configuration.

I/O drawers may be connected to the CEC in either single-loop or dual-loop mode.

- Single-loop (Figure 2-16 on page 36) mode connects an entire I/O drawer to the CEC using one RIO-2 loop (2 ports). The two I/O planars in the I/O drawer are connected together using a short RIO-2 cable. Single-loop connection requires one RIO-2 Loop Attachment Adapter (FC 7818) per I/O drawer.
- Dual-loop (Figure 2-17 on page 36) mode connects each I/O planar in the drawer to the CEC separately. Each I/O planar is connected to the CEC using a separate RIO-2 loop. Dual-loop connection requires two RIO-2 Loop Attachment Adapters (FC 7818) per I/O drawer. With dual-loop configuration, the RIO-2 bandwidth for the I/O drawer is higher.

> **Note:** Dual-loop mode is recommended whenever possible, as it provides the maximum bandwidth between the I/O drawer and the CEC.

Table 2-11 lists the number of single-looped and double-looped I/O drawers that can be connected to a p5-590 or p5-595 server based on the number of processor books installed.

Table 2-11 Number of RIO drawers that can be connected

Number of processor books	Single-looped	Dual-looped
1	6	3
2	8 (590) 12 (595)	6
3	12 (p5-595)	9 (p5-595)
4	12 (p5-595)	12 (p5-595)

On initial orders of p5-590 or p5-595 servers, IBM manufacturing will place dual-loop-connected I/O drawers as the lowest numerically designated drawers followed by any single-looped I/O drawers.

2.6.3 Single loop (full-drawer cabling)

For an I/O drawer, the following connections are required for a single loop cabling.

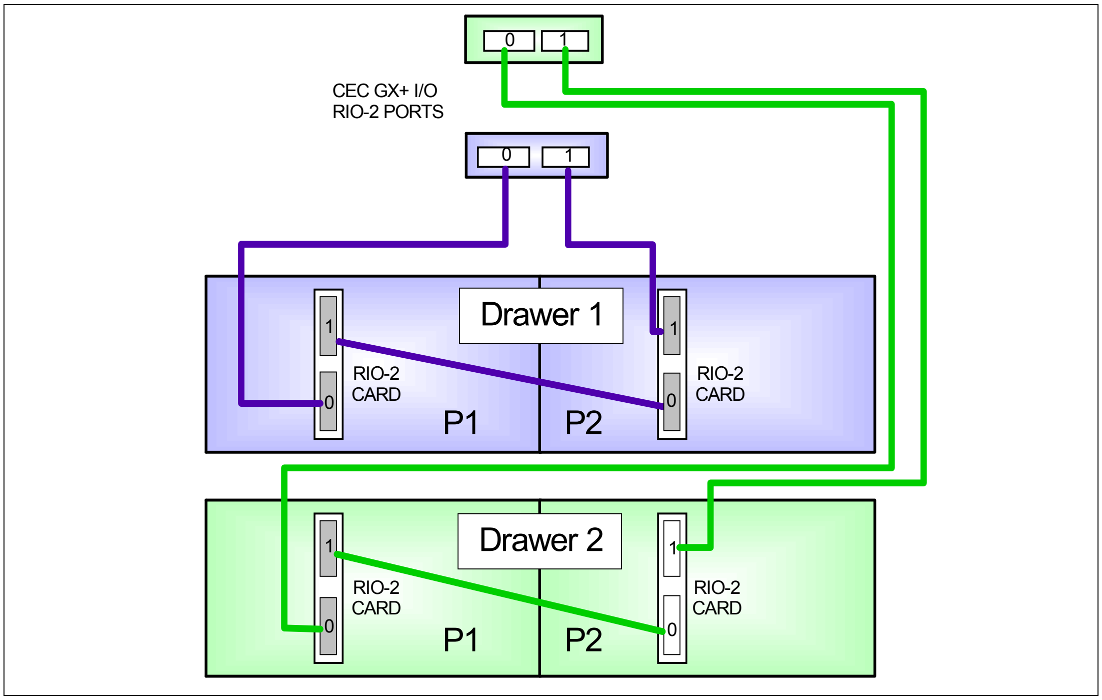


Figure 2-16 Single loop I/O drawer (FC 5791)

The short RIO-2 cable connecting the two halves of the drawer ensures that each side of the drawer (P1 and P2) can be accessed by the CEC I/O (RIO-2 adapter) card, even if one of the cables is damaged. Each half of the I/O drawer can communicate with the CEC I/O card for its own uses or on behalf of the other side of the drawer.

2.6.4 Dual looped (Half-drawer cabling)

Although I/O drawers will not be built in half-drawer configurations, they can be cabled to, and addressed by the CEC, in half drawer increments (Figure 2-17).

For an I/O drawer, the following connections are required for dual loop cabling.

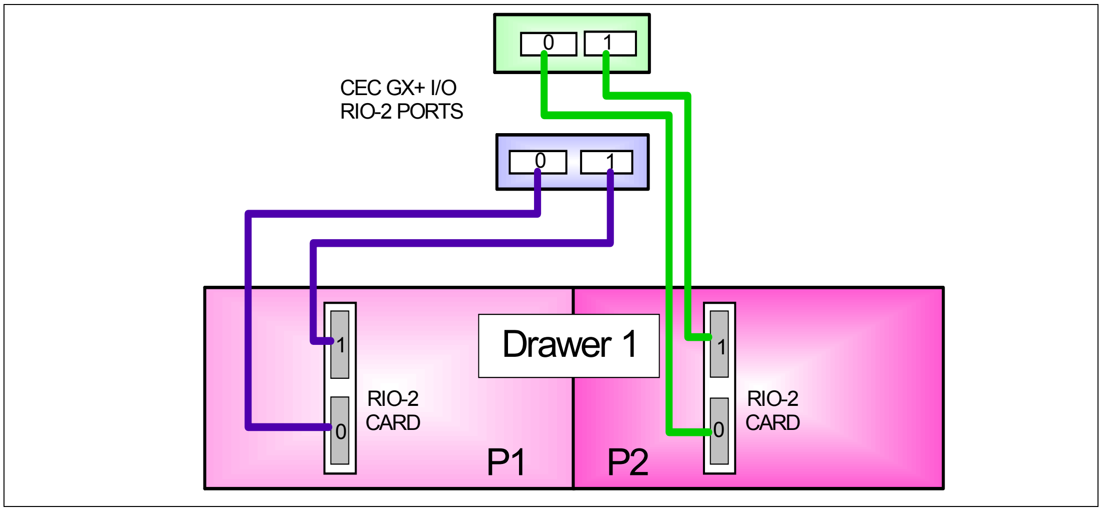


Figure 2-17 Dual loop I/O drawer (FC 5791)

However, to simplify the management of the server we strongly recommend that I/O loops be configured as described in the IBM @server Information Center, and to only follow a different order when absolutely necessary.

In any case, it becomes extremely important for the management of the system to keep up-to-date cabling documentation of your systems, because it may be different from the cabling diagrams of the installation guides.

2.6.5 Disks and boot devices

A minimum of two internal SCSI hard disks are required per server. It is recommended that these disks be used as mirrored boot devices. These disks should be mounted in the first I/O drawer whenever possible. This configuration provides service personnel with the maximum amount of diagnostic information if the system encounters errors in the boot sequence.

Boot support is also available from local SCSI, SSA, and Fibre Channel adapters, or from networks using Ethernet or token-ring adapters.

If the boot source other than internal disk is configured, the supporting adapter should also be in the first I/O drawer.

2.6.6 Media options

The p5-590 and p5-595 servers must have access either to a device capable of reading CD media or to a NIM server.

- The recommended devices for reading CD media are the rack-mounted media drawer (FC 5795) or an IBM Storage Device Enclosure. The Media Drawer (FC 5795) is mounted in the 13U location of the CEC Rack; the 7212-102, 7210-025, and 7210-030 enclosures attach using a PCI SCSI adapter in one of the system I/O drawers.
- If a NIM server is used, it must attach through a PCI LAN adapter in one of the system I/O drawers. An Ethernet connection is recommended.

Rack-mounted Media Drawer (FC 5795) provides a 1U high internal media drawer for use with the p5-590 and p5-595 servers. The media drawer displaces any I/O drawer or battery backup feature components that would be located in the same location of the primary system rack. The Media Drawer provides a fixed configuration that must be ordered with three media devices and all required SCSI and power attachment cabling.

This media drawer can be mounted in the CEC rack with three available media bays, two in the front and one in the rear. The device in the rear is only accessible from the rear of the system. New storage devices for the media bays include:

- 16X/48X IDE DVD-ROM drive (FC 2634)
- 4.7 GB, SCSI DVD-RAM drive (FC 5752)
- 36/72 GB, 4 mm internal tape drive (FC 6258)

This offering is preferred to the 7212-102 Storage Device Enclosure, which cannot be mounted in the p5-590 and p5-595 CEC rack.

2.6.7 PCI-X slots and adapters

PCI-X, where the X stands for extended, is an enhanced PCI bus, delivering a theoretical peak bandwidth of up to 1 GB/s, running a 64-bit bus at 133 MHz. PCI-X is backward compatible, so the p5-590 and p5-595 I/O drawers can support existing 3.3 volt PCI adapters.

Most PCI and PCI-X adapters for the p5-590 and p5-595 servers are capable of being hot-plugged. Any PCI adapter supporting a boot device or system console should not be hot-plugged. The following adapters are not hot-plug-capable:

- POWER GXT135P Graphics Accelerator with Digital Support (FC 2849)
- 2-Port Multiprotocol PCI Adapter (FC 2962)

System maximum limits for adapters and devices may not provide optimal system performance. These limits are given to help assure connectivity and function.

Configuration limitations have been established to help ensure appropriate PCI or PCI-X bus loading, adapter addressing, and system and adapter functional characteristics when ordering I/O drawers. These I/O drawer limitations are in addition to individual adapter limitations shown in the feature descriptions section of the Sales Manual.

The maximum number of a specific PCI or PCI-X adapters allowed per p5-590 and p5-595 server may be less than the number allowed per I/O drawer multiplied by the maximum number of I/O drawers.

The PCI-X slots in the I/O drawers of p5-590 and p5-595 servers support Extended Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet generated from the affected PCI-X slot hardware by calling system firmware, which is designed to examine the affected bus, allow the device driver to reset it, and continue without a system reboot.

> **Note:** As soon as a p5-590 or p5-595 server is connected to a Hardware Management Console, the POWER Hypervisor™ software will prevent the system from using non-EEH OEM adapters.

To find more information about PCI adapter placement look at the IBM @server Hardware Information Center by placing a search for *PCI placement*.

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

2.6.8 LAN adapters

Table 2-12 lists additional LAN adapters available at the time of writing. IBM supports an installation with Network Installation Manager (NIM) using Ethernet adapters (CHRP[2] is the platform type).

Table 2-12 Available LAN adapter

Feature code	Adapter description	Size	Max 595/590
4962	IBM 10/100 Mbps Ethernet PCI Adapter II	Short	192/160
5700	IBM Gigabit Ethernet-SX PCI-X Adapter	Short	192/160
5701	IBM 10/100/1000 Base-TX Ethernet PCI-X Adapter	Short	192/160
5706	IBM 2-Port 10/100/1000 Base-TX Ethernet PCI-X Adapter	Short	192/160
5707	IBM 2-Port Gigabit Ethernet-SX PCI-X Adapter	Short	192/160
5718	IBM 10 Gigabit Ethernet-SR PCI-X Adapter	Short	24/16

[2] CHRP stands for Common Hardware Reference Platform, a specification for PowerPC processor-based systems that can run multiple operating systems.

Feature code	Adapter description	Size	Max 595/590
5719	IBM 10 Gigabit Ethernet-LR PCI-X Adapter	Short	24/16

2.6.9 SCSI adapters

The p5-590 and p5/595 I/O drawers have four integrated Ultra3 SCSI adapters. Integrated adapters support the SCSI Enclosure Services (SES hot-swappable control functions). The integrated Ultra3 SCSI adapters are for internal storage only and cannot be used for external disks.

Table 2-13 lists additional SCSI adapters available at the time of writing. All listed adapters can be used as boot adapters.

Table 2-13 Available SCSI adapters

Feature code	Adapter description	Size	Max 595 and 590
6204	PCI Universal Differential Ultra SCSI Adapter	Short	32
5710	PCI-X Dual Channel Ultra320 SCSI Blind Swap Adapter	Long	62
5711	PCI-X Dual Channel Ultra320 SCSI RAID Blind Swap Adapter	Long	62

2.6.10 Internal storage

Each I/O drawer contains four integrated Ultra3 SCSI adapters and SCSI Enclosure Services (SES hot-swappable control functions). In drawer FC 5791 all four are used, each being connected to a SCSI 4-pack backplane. In drawer FC 5794 only two 4-packs are installed.

Each of the 4-packs supports up to four hot-swappable Ultra3 SCSI disk drives, which can be used for installation of the operating system or storing data.

Table 2-14 lists hot-swappable disk drives.

Table 2-14 Hot-swappable disk drive options

Feature code	Description
3277	36.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive
3278	73.4 GB 15,000 rpm Ultra320 SCSI hot-swappable disk drive

Note: Disks with 10 K rpm rotational speeds from earlier systems are not supported.

Prior to the hot-swap of a disk in the hot-swappable capable bay, all necessary operating system actions must be undertaken to ensure that the disk is capable of being deconfigured. After the disk drive has been deconfigured, the SCSI enclosure device will power-off the bay, enabling safe removal of the disk. You should ensure that the appropriate planning has been given to any operating-system-related disk layout, such as the AIX 5L Logical Volume Manager, when using disk hot-swap capabilities. For more information, see *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496.

2.7 Logical partitioning and virtualization

Dynamic logical partition (DLPAR) and virtualization increase utilization of system resources. The following section provides details and configuration specifications on this topic. The virtualization discussion includes virtualization enabling technologies that are standard on the system, such as the POWER Hypervisor component and the Advanced POWER Virtualization feature.

2.7.1 Dynamic logical partitioning

Introduced with the POWER4 processor product line and the AIX 5L Version 5.1 operating system, logical partitioning (LPAR) became available. This technology offered the capability to divide a pSeries system into separate logical systems (partitions), allowing each LPAR to run an operating environment on dedicated attached devices, such as processors, memory, and I/O components.

Later, dynamic LPAR increased the flexibility, allowing selected system resources, such as processors, memory, and I/O components, to be added and deleted from dedicated partitions while they are running. AIX 5L Version 5.2, with all the necessary enhancements to enable dynamic LPAR, was introduced in 2002. The ability to reconfigure dynamic LPARs enables system administrators to dynamically redefine all available system resources to reach the optimum capacity for each defined dynamic LPAR.

For information about minimum required hardware and software, refer to “Hardware and software guidelines for dynamic LPAR and virtualization” on page 45.

2.7.2 Virtualization

With the introduction of the POWER5 processor, partitioning technology moved from a dedicated resource allocation model to a virtualized shared resource model. This section briefly discusses the key components of virtualization on IBM @server p5 servers.

More information on virtualization can be found at the following URL:

http://www.ibm.com/servers/eserver/about/virtualization/

POWER Hypervisor

Combined with features designed into the POWER5 processor, the POWER Hypervisor component delivers functions that enable other system technologies including Micro-Partitioning technology, virtualized processors, IEEE VLAN compatible virtual switch, virtual SCSI adapters, and virtual consoles. The POWER Hypervisor component is part of system firmware that is always active, regardless of system configuration.

The POWER Hypervisor component performs the following tasks:

- Provides an abstraction layer between the physical hardware resources and the logical partitions using them
- Enforces partition integrity by providing a security layer between logical partitions
- Controls the dispatch of virtual processors to physical processors
- Saves and restores all processor state information during logical processor context switch
- Controls hardware I/O interrupt management facilities for logical partitions
- Provides virtual LAN channels between physical partitions that help to reduce the need for physical Ethernet adapters for inter-partition communication

Three types of virtual I/O adapters are supported by the POWER Hypervisor firmware.

Virtual SCSI

The POWER5 server uses SCSI as the mechanism for virtual storage devices. This is accomplished using two paired adapters: A virtual SCSI server adapter and a virtual SCSI client adapter.

Virtual Ethernet

The POWER Hypervisor component provides a virtual Ethernet switch function that allows partitions on the *same server* a means for fast and secure communication. Virtual Ethernet working on LAN technology allows a transmission speed in the range of 1 to 3 GB/s depending on the MTU[3] size. Virtual Ethernet requires a POWER5 system with either AIX 5L Version 5.3 or the appropriate level of Linux and an HMC to define the virtual Ethernet devices. Virtual Ethernet does not require the purchase of any additional features or software such as the Advanced POWER Virtualization feature.

Virtual Ethernet features include:

- A partition supports 256 virtual Ethernet connections, where a single virtual Ethernet resource can be connected to another virtual Ethernet, a real network adapter, or both in a partition. Each virtual Ethernet adapter can also be configured as a trunk adapter.
- Each partition operating system detects the virtual local area network (VLAN) switch as an Ethernet adapter, without the physical link properties and asynchronous data transmit operations. Layer-2 bridging to a physical Ethernet adapter is also included in the virtual Ethernet features. The virtual Ethernet network is extendable outside the server to a physical Ethernet network.

Note: Virtual Ethernet is based on the IEEE 802.1Q VLAN standard. No physical I/O adapter is required when creating a VLAN connection between partitions, and no access to an outside network is required.

Virtual (TTY) console

Each partition needs to have access to a system console. Tasks such as operating system install, network setup, and some problem analysis activities require a dedicated system console. The POWER Hypervisor component provides virtual console using a virtual TTY or serial adapter and a set of Hypervisor calls to operate on them. Virtual TTY, or **`vterm`**, does not require the purchase of any additional features or software such as the Advanced POWER Virtualization feature.

Depending on the system configuration, the operating system console can be provided by the Hardware Management Console virtual TTY, or from a terminal emulator connected to an synchronous adapter in an I/O drawer.

[3] Maximum transmission unit

Note: The POWER5 Hypervisor component is active when the server is running in partition and non-partition mode, and also when not connected to the HMC. The Hypervisor memory requirements should be considered when planning the amount of system memory required. Use the LPAR Validation Tool for calculating the Hypervisor memory requirements.

http://www-1.ibm.com/servers/eserver/pseries/lpar/planning.html

In AIX 5L Version 5.3, the **lparstat** command using the -h and -H flags will display Hypervisor statistical data. Using the -h flag adds summary Hypervisor statistics to the default **lparstat** output.

2.7.3 Advanced POWER Virtualization feature

The Advanced POWER Virtualization is a standard feature (FC 7992) for p5-590 and p5-595 servers. This feature enables the implementation of virtual partitions on IBM @server p5 servers. For p5-590 and p5-595 servers the configurator selects FC 7992 to order the Advanced POWER Virtualization feature. See Figure 2-16 on page 45 for operating system support for this feature.

The Advanced POWER Virtualization includes:

- Firmware enablement for Micro-Partitioning technology
  - Supports up to 254 partitions, 1/10th of processor granularity
- Installation image for the Virtual I/O Server (VIOS) software that supports:
  - Ethernet adapter sharing
  - Virtual SCSI Server
  - VIOS ships on a CD
  - Software supports AIX 5L and Linux
- Partition Load Manager (AIX 5L Version 5.3 only)
  - Automated CPU and memory reconfiguration
  - Real-time partition configuration and load statistics
  - Graphical user interface
  - Software ships on a CD

For more details on Advanced POWER Virtualization and virtualization see the following URL:

http://www.ibm.com/servers/eserver/pseries/ondemand/ve/resources.html

Note: The Advanced POWER Virtualization feature (FC 7992) is not supported on AIX 5L for POWER Version 5.2, or previous versions.

Micro-Partitioning technology

Micro-Partitioning technology is designed to allow the resource definition of a partition to allocate fractions of processors to the partition. Micro-Partitioning technology is only available with POWER5 systems. From an operating system perspective, a virtual processor is indistinguishable from a physical processor, unless the operating system had been enhanced to be made aware of the difference. Physical processors are abstracted into virtual processors that are available to partitions.

A partition may be defined with a processor capacity as small as 10 processor units. This represents 1/10 of a physical processor. Each processor can be shared by up to 10 shared processor partitions and each partition can then can be incremented fractionally by as little as 1/100th of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under control of the POWER Hypervisor component. The shared processor partitions are created and managed by the HMC. Dedicated and micro-partitioned processors can co-exist on the same POWER5 server as long as they are available.
Table 2-15 lists processor partitioning information related to the p5-590 and p5-590 servers.

Table 2-15 Processor partitioning overview of the p5-590 and p5-595 servers

Partitioning implementation	Model 590	Model 595
Processors (maximum configuration)	32	64
Dedicated processor partitions (maximum configuration)	32	64
Shared processor partitions (maximum configuration)	254	254

It is important to point out that the maximums stated are supported by the hardware, but the practical limits based on production workload demands may be significantly lower. Also see Table 2-16 on page 45 for operating systems supported using Micro-Partitioning technology.

Virtual I/O Server

The Virtual I/O Server is a special purpose partition that provides virtual I/O resources to client partitions. The Virtual I/O Server owns the real resources that are shared with the other LPARs. The Virtual I/O technology allows a physical adapter assigned to a VIO Server partition to be shared by one or more partitions, enabling clients to minimize the number of physical adapters. The Virtual I/O Server eliminates the requirement that every partition own a dedicated network adapter, disk adapter, and disk drive.

Figure 2-18 shows an organization view of Micro-Partitioning technology including the Virtual I/O Server. The figure also includes virtual SCSI and Ethernet connections and mixed operating system partitions.

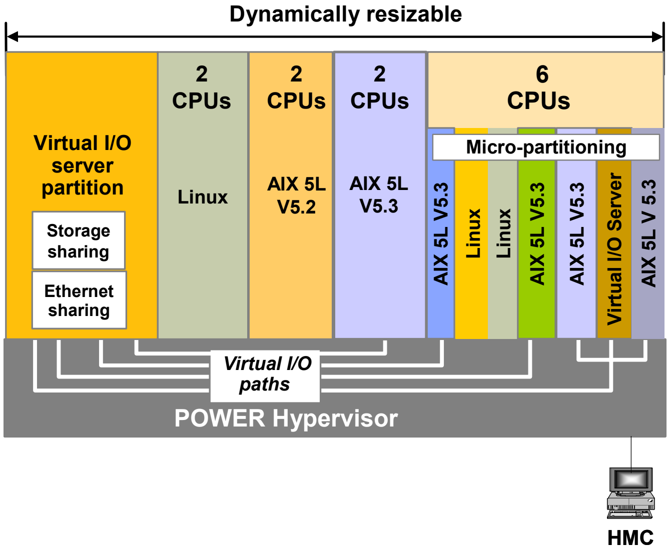


Figure 2-18 Virtual partition organization view

Since the Virtual I/O Server is an AIX 5L Version 5.3 operating system-based appliance, redundancy for physical devices attached to the Virtual I/O Server can be provided by using capabilities such as Multipath I/O and IEEE 802.3ad Link Aggregation.

Installation of the Virtual I/O Server partition is performed from a special mksysb CD that is provided to clients who order the Advanced POWER Virtualization feature. This is dedicated software only for the Virtual I/O Server operations, so the Virtual I/O Server software is only supported in Virtual I/O Server partitions.

The Virtual I/O Server can be installed by:

- Media (assigning the DVD-ROM drive to the partition and booting from the media)
- The HMC (inserting the media in the DVD-ROM drive on the HMC and using the `installios` command)
- Using Network Install Manager (NIM)

> **Note:** To increase the performance of I/O intensive applications, dedicated physical adapters are preferred using dedicated partitions.
>
> It is recommended that you install the Virtual I/O Server in a partition with dedicated resources to help ensure consistent performance.
>
> The Virtual I/O Server supports logical mirroring and RAID configurations. Logical volumes created on RAID or JBOD configurations are bootable, and the number of logical volumes is limited to the amount of storage available and architectural limits of the LVM.

Two major functions are provided with the Virtual I/O Server: A shared Ethernet adapter and Virtual SCSI.

Shared Ethernet adapter

A shared Ethernet adapter is a new service that acts as a layer 2 network switch to route network traffic from a virtual Ethernet to a real network adapter. The shared Ethernet adapter must be assigned to the Virtual I/O Server partition.

Virtual SCSI

Access to real storage devices is implemented through the Virtual SCSI services, a part of the Virtual I/O Server partition. This is accomplished using a pair of virtual adapters: A virtual SCSI server adapter and a virtual SCSI client adapter. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands it receives. It is owned by the Virtual I/O Server partition. The virtual SCSI client adapter allows the client partition to access standard SCSI devices and LUNs that are assigned to the client partition.

All current storage device types, such as SAN, SCSI, and RAID, are supported. iSCSI and SSA are not supported.

More information on specific storage devices supported can be found at:

http://techsupport.services.ibm.com/server/virtualization/vios/documentation/datasheet.html

> **Note:** The shared Ethernet adapter and virtual SCSI function is provided in the Virtual I/O Server that is included with the Advanced POWER Virtualization feature services.

Table 2-16 on page 45 lists details on requirements for configuring virtual partitions.

Partition Load Manager

Partition Load Manager (PLM) provides automated processor and memory distribution between a dynamic LPAR and Micro-Partitioning capable logical partition running AIX 5L. The PLM application is based on a client/server model to share system information, such as processor or memory events, across the concurrent present logical partitions.

The following events are registered on all managed partition nodes:

- Memory-pages-steal high thresholds and low thresholds
- Memory-usage-high thresholds and low thresholds
- Processor-load-average high threshold and low threshold

Note: PLM is supported on AIX 5L Version 5.2 and AIX 5L Version 5.3.

2.7.4 Hardware and software guidelines for dynamic LPAR and virtualization

This section covers the main considerations regarding dynamic LPAR and virtualization.

Operating system support for dynamic LPAR and virtualization

Table 2-16 lists AIX 5L and Linux support for dynamic LPAR and virtualization.

Table 2-16 Operating system supported function

Function	AIX 5L Version 5.2	AIX 5L Version 5.3	Linux SLES 9	Linux RHEL AS 3	Linux RHEL AS 4
Dynamic LPAR					
Processor	Y	Y	Y	N	Y
Memory	Y	Y	N	N	N
I/O	Y	Y	Y	N	Y
Virtualization					
Micro-partitions (1/10th of processor)	N	Y	Y	Y	Y
Virtual Storage	N	Y	Y	Y	Y
Virtual Ethernet	N	Y	Y	Y	Y
Partition Load Manager	Y	Y	N	N	N

Dynamic LPAR minimum requirements

The minimum resources that are needed per LPAR (not per system) are the following:

- At least one processor per partition for a dedicated processor partition or at least 1/10th of a processor when using Micro-Partitioning technology.
- At least 128 MB of physical memory per additional partition.
- At least one disk (either physical or virtual) to store the operating system.
- At least one disk adapter (either physical or virtual) or integrated adapter to access the disk.
- At least one Ethernet adapter (either physical or virtual) per partition to provide a network connection to the HMC, as well as general network access.

Note: It is recommended to use separate adapters for the management and the public LAN to protect the access of your system's management functions.

- A partition must have an installation method, such as NIM or CD/DVD, and a means of running diagnostics, such as network diagnostics.

Processor

Each LPAR requires at least one physical processor if virtualization is not used. Based on this, the maximum number of dynamic LPARs without virtualization is 32 for the p5-590 server and 64 for the p5-595 server. With the use of the Advanced POWER Virtualization feature, the number of partitions per processor is 10.

Memory

It is important to highlight that the IBM @server p5 and OpenPower™ servers and their associated virtualization features have adopted an even more dynamic memory allocation policy than the previous partition-capable pSeries servers.

In a partitioned environment, some of the physical memory areas are reserved by several system functions to enable partitioning in the partitioning-capable server. You can assign unused physical memory to a partition. You do not have to specify the precise address of the assigned physical memory in the partition profile, because the system selects the resources automatically.

The Hypervisor firmware requires memory to support the logical partitions on the server. The amount of memory required by the Hypervisor firmware varies according to several factors. Factors influencing the Hypervisor memory requirements include the following:

- Number of logical partitions
- Partition environments of the logical partitions
- Number of physical and virtual I/O devices used by the logical partitions
- Maximum memory values given to the logical partitions

Generally, you can estimate the amount of memory required by server firmware to be approximately eight percent of the system installed memory. The actual amount required will generally be less than eight percent. However, there are some server models that require an absolute minimum amount of memory for server firmware, regardless of the previously mentioned considerations.

The minimum amount of physical memory for each partition is 128 MB, but in most cases the actual requirements and recommendations are between 256 MB and 512 MB for AIX 5L, Red Hat, and Novell SUSE LINUX. After that, you can assign further physical memory to partitions in increments of 16 MB. This is supported for partitions running AIX 5L Version 5.2 with the 5200-04 Recommended Maintenance package, AIX 5L Version 5.3, Red Hat Enterprise Linux AS 3 (no dynamic LPAR), Red Hat Enterprise Linux AS 4, and SUSE LINUX Enterprise Server 9. There are implications on how big a partition can grow based on the amount of memory allocated initially. For partitions that are initially sized less than 256 MB, the maximum size is 16 times the initial size. For partitions initially sized 256 MB or larger, the maximum size is 64 times the initial size.

Note: For a more detailed impression of the amount of memory required by the server firmware, use the LPAR Validation Tool (LVT). Refer to "LPAR validation tool" on page 48.

I/O

The I/O devices are assigned on a slot level to the LPARs, meaning an adapter (either physical or virtual) installed in a specific slot can only be assigned to one LPAR.

If an adapter has multiple devices, such as the 4-port Ethernet adapter or the Dual Ultra3 SCSI adapter, all devices are automatically assigned to one LPAR and cannot be shared.

Devices connected to an internal controller must be treated as a group. A group can only be assigned together to one LPAR and cannot be shared.

Therefore, the following integrated devices can be independently assigned to LPARs:

- I/O drawer (FC 5791and FC 5794) Integrated Ultra320 SCSI controller

  All SCSI resources in the disk bays must be assigned together to the same LPAR. There is no requirement to assign them to a particular LPAR; in fact, they can remain unassigned if the LPAR minimum requirements are obtained using devices attached to a SCSI adapter installed in the system.

- Media devices

  The p5-590 and p-595 servers can be configured with an optional rack-mounted media drawer (FC 5795) or storage device enclosure (IBM 7212-102).

  These devices must belong to only a single LPAR at a time; therefore, all devices in the media bays will be available to only one LPAR at a time.

Virtual I/O devices are also assigned to dynamic LPARs on a slot level. Each partition is capable of handling up to 256 virtual I/O slots. Therefore each partition can have up to:

- 256 virtual Ethernet adapters with each virtual Ethernet capable of being associated with up to 21 VLANs.
- 256 virtual SCSI adapters.

> **Note:** For more detailed planning of the virtual I/O slots and their requirements, use the LPAR validation tool.

Every LPAR requires disks (either physical or virtual) for the operating system.

Partitions must be assigned to the boot adapter and disk drive from the following options:

- An internal disk drive inserted in one of the 4-pack disk bays on I/O drawer (FC 5791 or FC 5794) and the SCSI controller on the drawer. Each of the disk bays is connected to a separate internal SCSI controller on the drawer.
- A boot adapter inserted in one of 20 PCI-X slots in a I/O drawer connected to the system. A bootable external disk subsystem is connected to this adapter.

Therefore, for additional LPARs without using virtualization, external disk space is necessary, which can be accomplished by using external disk subsystems. The external disk space must be attached with a separate adapter for each LPAR by using SCSI or Fibre Channel adapters, depending on the subsystem.

For additional LPARs using virtualization, the required disk drives for each partition are provided by the Virtual I/O Server partitions. Physical disks owned by the Virtual I/O Server partition can either be exported and assigned to a client partition whole, or can be partitioned into several logical volumes. The logical volumes can then be assigned to different partitions.

For the p5-590 and p5-595 servers, additional disk space can be provided by using an external storage subsystem such as the IBM TotalStorage DS4500, for LPARs using virtualization or direct attachment. For more detailed information about the available IBM disk subsystems, refer to “External disk subsystem” on page 10.

For additional LPARs without using virtualization, an additional Ethernet adapter is necessary. As stated previously, it is highly recommended to use separate Ethernet adapters for connection to the management LAN and public LAN.

Additional partitions using virtualization can implement the required Ethernet adapters as virtual Ethernet adapters. Virtual Ethernet adapters can be used for all kinds of inter-partition communication. To connect the virtual Ethernet LANs to an external network, one or more Shared Ethernet Adapters (SEA) can be used in the Virtual I/O Server partition.

2.7.5 LPAR validation tool

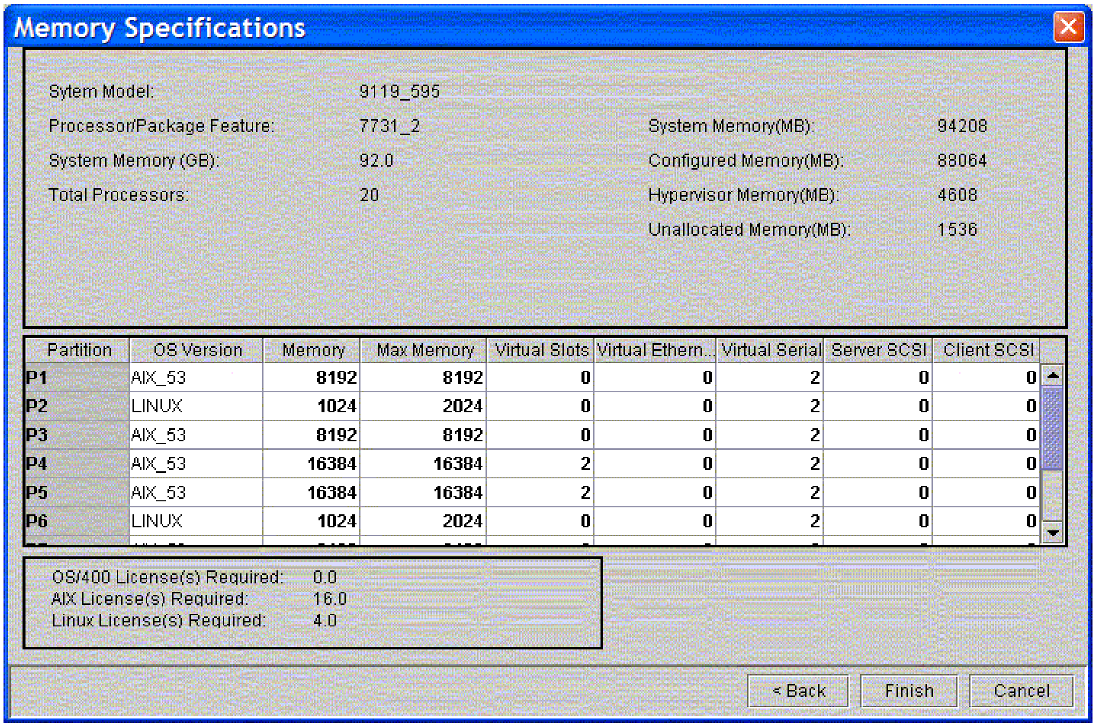
When configuring dynamic or virtual partitions on @server p5 systems, the LPAR Validation Tool (LVT) can be used to verify system resource requirements. With the LVT, you can customize the partition design by selecting PCI slots for given adapters, specific drives to selected bays, and much more. The LVT provides a useful report that can complement the organization and validation of features required for configuration of a complex partition solution. The LVT supports IBM @server p5 and @server i5 servers, iSeries™, and OpenPower systems. A proficient knowledge of LPAR design requirements, limitations, and best practices facilitates the use of this tool.

The LVT tool provides the following functions:

- Support for partitions running AIX 5L Version 5.2 and Version 5.3, and Linux
- Validation of dynamic LPAR design
- Validation of virtual partition design, including Virtual I/O Server and virtual clients
- Calculates unallocated memory and shared processor pool
- Calculates Hypervisor memory requirements
- Calculates number of operating system licenses needed to support partition design
- Validates number of virtual slots required for partitions

> **Important:** We recommend the use of the LVT to calculate Hypervisor requirements to determine memory resources required for all partitioned and non-partitioned servers.

Figure 2-19 on page 49 shows the calculated Hypervisor memory requirements based on sample partition requirements.



Sytem Model:	9119_595
Processor/Package Feature:	7731_2
System Memory (GB):	92.0
Total Processors:	20

System Memory(MB):	94208
Configured Memory(MB):	88064
Hypervisor Memory(MB):	4608
Unallocated Memory(MB):	1536

Partition	OS Version	Memory	Max Memory	Virtual Slots	Virtual Ethern...	Virtual Serial	Server SCSI	Client SCSI
P1	AIX_53	**8192**	**8192**	**0**	**0**	2	**0**	**0**
P2	LINUX	**1024**	**2024**	**0**	**0**	2	**0**	**0**
P3	AIX_53	**8192**	**8192**	**0**	**0**	2	**0**	**0**
P4	AIX_53	**16384**	**16384**	2	**0**	2	**0**	**0**
P5	AIX_53	**16384**	**16384**	2	**0**	2	**0**	**0**
P6	LINUX	**1024**	**2024**	**0**	**0**	2	**0**	**0**

OS/400 License(s) Required:	0.0
AIX License(s) Required:	16.0
Linux License(s) Required:	4.0

Figure 2-19 LVT screen showing Hypervisor requirements

The LVT is a standalone Java™ application that runs on a Microsoft® Windows® 95 or later workstation with 128 MB minimum of free memory.

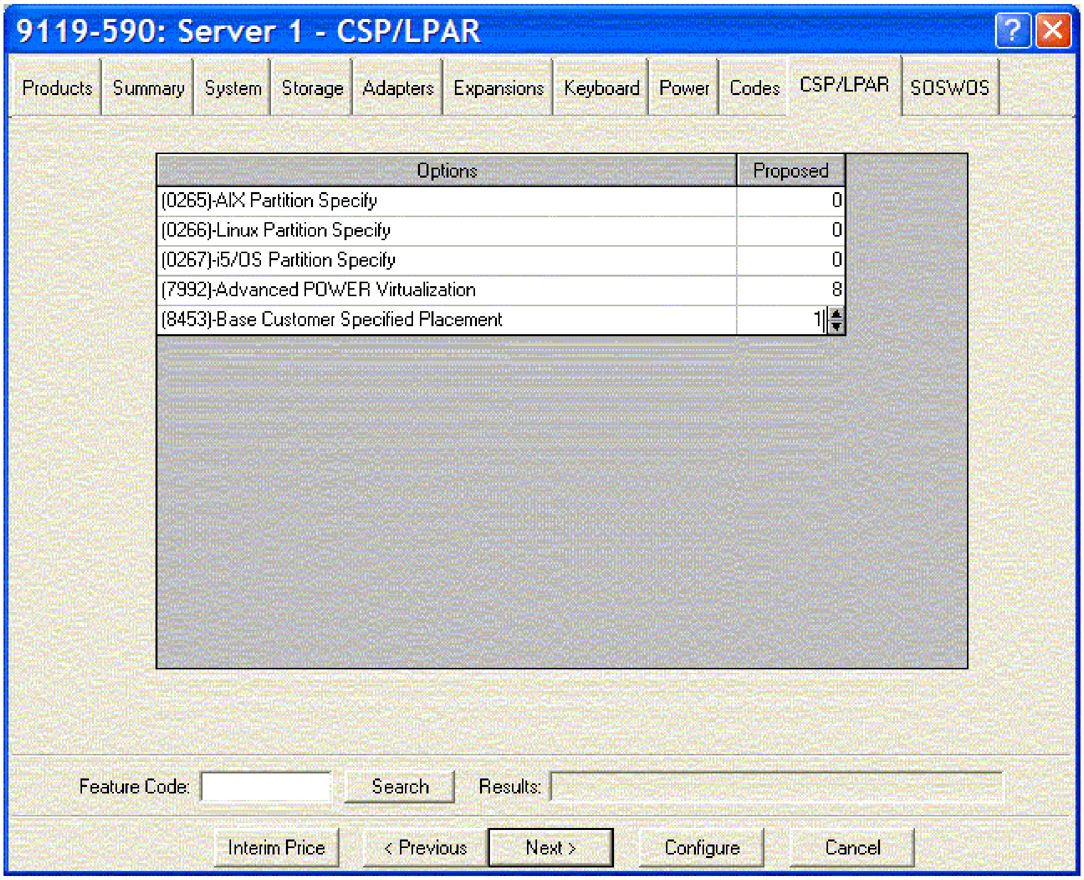
For download and installation information, including the user’s guide, visit:

http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm

2.7.6 Client-specific placement and eConfig

The LVT also provides the output report that is used for the Customer Specified Placement (CSP) offering. The LVT output is uploaded on the CSP site for submission to manufacturing. The CSP offering enables the placement of adapters and disks for an exact built-to-order system based on a client’s specifications. Manufacturing uses the LVT output to custom build the server. The server is then shipped configured with the features placed as indicated in the LVT.lvt output file.

The server configuration must include the CSP FC 8453. This CSP feature code is selected on the *Code tab* in the IBM Configurator for e-business (eConfig) wizard. Figure 2-20 on page 50 shows a screen shot of FC 8453 in the eConfig wizard.



Options	Proposed
(0265)-AIX Partition Specify	0
(0266)-Linux Partition Specify	0
(0267)-i5/OS Partition Specify	0
(7992)-Advanced POWER Virtualization	8
(8453)-Base Customer Specified Placement	1

Figure 2-20 eConfig tab for CSP selection

Note: An order submitted must include the CSP FC 8453 on the configuration, and the form submitted to the CSP site must include the .lvt output file and the order number of the system.

CSP is available on a limited number of POWER5 systems. See the following URL for details:

`http://www.ibm.com/servers/eserver/power/csp/index.html`

2.8 System management

The following section provides an overview of powering on the managed system, the service processor, HMC, and firmware.

2.8.1 Power on

There is no physical operator panel on the p5-590 and p5-595. System status on earlier models could be determined, to some extent, by the operator panel LCD display. For the p5-590 and p5-595 it is necessary to refer to the BPC front panel (Figure 2-21 on page 51) or the HMC (Figure 2-22 on page 51).

The power on sequence is described in Table 2-17 on page 51.

Table 2-17 Power on sequence

State	Indication[a]
Power cords connected - EPO switch Off	On BPCs, the UEPO Power LED is on. (Also, Power Good LED - on the far right of the panel will be on throughout this sequence.)
Power available - EPO switch ON	UEPO Power LED remains on, UEPO CMPLT turns on. The service processors become active. Power STBY LED is flashing. During this step, the service processors and BPCs will get their IP addresses from the HMC (DHCP server).
Standby Power Complete	Power STBY LED is solid. Fans are running in *flushing* mode.
System Power On	DCAs power on (via commands received from the Ethernet). Light Strips powered on. Fans running in *fast* mode. (Note: it is quite normal for the fans to continue to run in fast mode for about the first 20 minutes.)

[a] See Figure 2-21 on page 51 for LED locations.

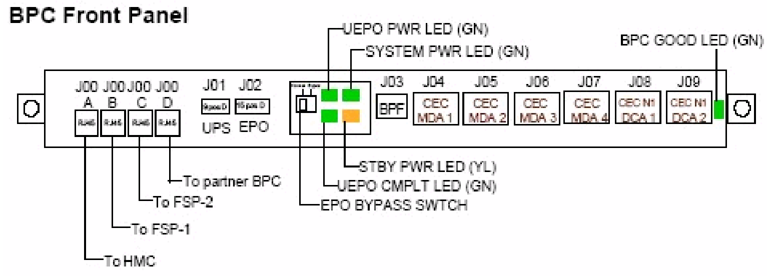
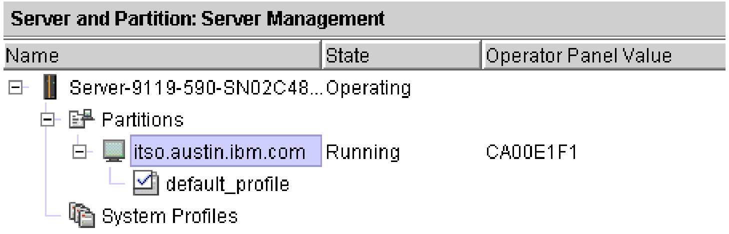


Figure 2-21 BPC front panel


Server and Partition: Server Management

Name	State	Operator Panel Value
Server-9119-590-SN02C48...	Operating	
Partitions		
itso.austin.ibm.com	Running	CA00E1F1
default_profile		
System Profiles		

Figure 2-22 Operating states as seen from the HMC

2.8.2 Service processor

Unlike other @server p5 servers, the p5-590 and p5-595 have two service processors.

The p5-590 and p5-595 service processor function is located on redundant service processor cards (same role of any service processor of @server p5 systems) in the CEC; one is considered the primary and the other secondary. The two service processor cards are in the same assembly with two redundant oscillator cards (OSC) shared by all processor books.

Figure 2-23 on page 52 details the oscillator and service processor assembly.

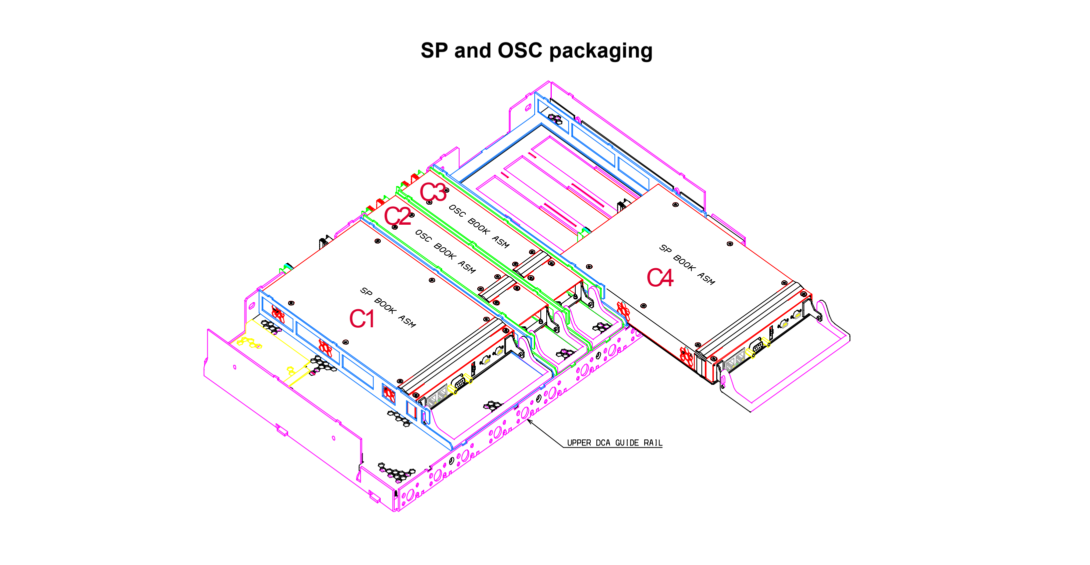


Figure 2-23   Oscillator and Service Processor Assembly

The service processor is an embedded controller based on a PowerPC 405GP processor (PPC405) implementation running the service processor internal operating system. The service processor operating system contains specific programs and device drivers for the service processor hardware.

The key components include a flexible service processor-base (FSP-B) and an extender chipset (FSP-E).

See Figure 2-21 on page 51 for more details.

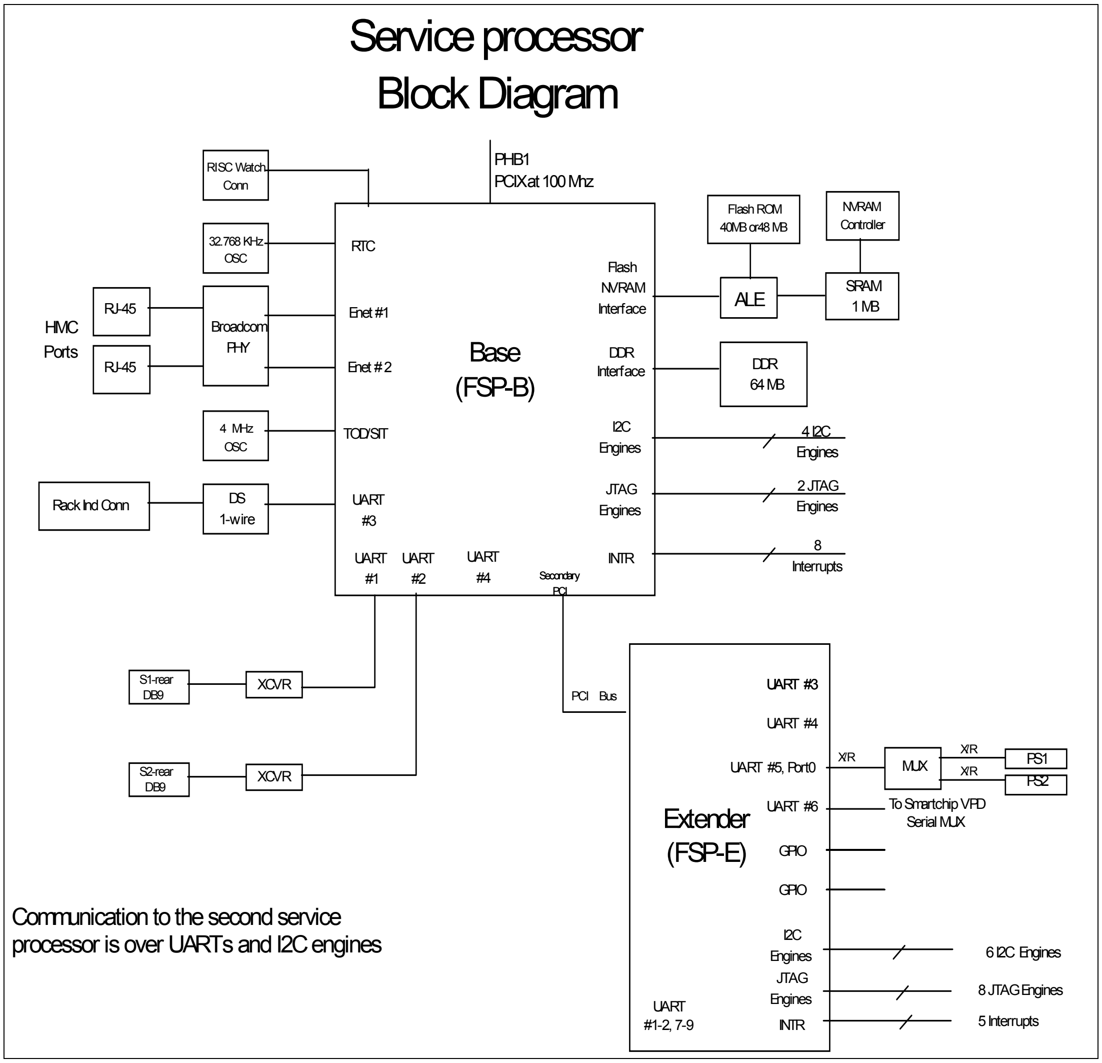


Figure 2-24 Service processor block diagram

The PPC405 core is five-stage pipeline instruction processor and contains 32-bit general purpose registers. The Flash ROM contains a compressed image of a software load. NVRAM contains configuration data and is backed up by battery in the event of power loss.

FSP-B has four UART cores, which provide a full duplex serial interface. As shown in Figure 2-24, UART #1 and UART #2 are used for RS232 Serial Port #1 and RS232 Serial Port #2, respectively. UART #3 is used for Rack VPD/Light interface. UART #4 is not used.

Server use communications ports

Each p5-590 or p5-595 server must be connected to a Hardware Management Console (HMC) for system control, LPAR, Capacity Upgrade on Demand, and service functions. The HMC is capable of supporting multiple p5 servers. It is strongly recommended to have two HMCs connected.

The p5-590 or p5-595 servers do not connect directly to HMC but through an Ethernet hub connection provided by the Bulk Power Controllers (FC 7803) part of the Bulk Power Assembly. The p5-590 and p5-595 are designed with dual bulk power controllers (BPC).

Bulk Power Controller (FC 7803) provides the base power distribution and control for the internal power assemblies and communications hub function for the HMC and the BPC. BPC is part of Bulk Power Assembly (BPA).

Each bulk power controller BPCs has two sides, commonly referred to as A and B sides. Each BPC hub has four 10/100 Ethernet ports to connect various system components. See Figure 2-26 on page 57 for details.

The BPC connectivity scheme is presented in Table 2-18.

Table 2-18 BPC connections

BPC Ethernet hub port	Connected component
BPC Port A	Connects to the Hardware Management Console (HMC)
BPC Port B	Connects to service processor 0
BPC Port C	Connects to service processor 1
BPC Port D	Connects to the partner BPC

Note: Two Bulk Power Controller Assemblies (FC 7803) are required for the 9119 system rack. Two additional FC 7803 BPCs are required when the optional Powered Expansion Rack (FC 5792) is ordered.

Figure 2-25 provides a full illustration of the service processor card.

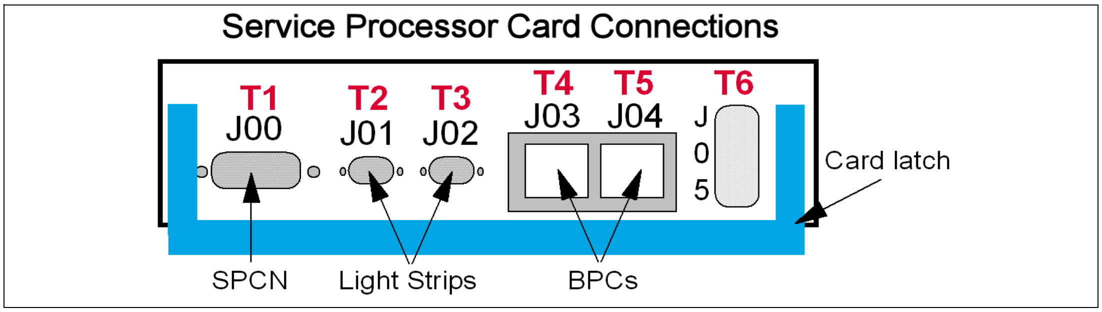


Figure 2-25 Service processor (front view)

Table 2-19 details the service processor cable connections.

Table 2-19 Table of service processor card location codes

Jack ID	Location code	Service processor 0	Service processor 1	Function
J00	T1			System Power Control Network (SPCN) connection
J01	T2	J00 CEC Front Light Strip	J01 CEC Front Light Strip	Light Strip connection
J02	T3	J01 CEC Back Light Strip	J00 CEC Back Light Strip	
J03	T4	J00C BPA-A side	J00B BPA-A side	Ethernet port 0 to Bulk Power Controller (BPC)

Jack ID	Location code	Service processor 0	Service processor 1	Function
J04	T5	J00C BPA-B side	J00B BPA-B side	Ethernet port 1 to Bulk Power Controller (BPC)
J05	T6	Unused		

2.8.3 HMC

The Hardware Management Console (HMC) is a dedicated workstation that controls managed systems, including IBM *e*server hardware, logical partitions, and Capacity on Demand. To provide flexibility and availability, there are different ways to implement HMCs, including the local HMC, remote HMC, redundant HMC, and the Web-based System Manager Remote Client. One HMC is capable of controlling multiple POWER5 processor-based systems.

Note: At the time of writing, one HMC supports up to 48 POWER5 processor-based systems and up to 256 LPARs using the HMC machine code Version 4.5.

Local HMC

A local HMC is any physical HMC that is directly connected to the system it manages through a private service network. An HMC in a private service network is a DHCP[4] server from which the managed system obtains the address for its service processor and bulk power controller.

Remote HMC

A remote HMC is a stand-alone HMC or an HMC installed in a 19-inch rack that is used to access another HMC. A remote HMC may be present in an open network.

Redundant HMC

A redundant HMC manages a system that is already managed by another HMC. When two HMCs manage one system, those HMCs are peers and can be used simultaneously to manage the system.

Web-based System Manager remote client

The Web-based System Manager Remote Client is an application that is usually installed on a PC. You can then use this PC to access HMCs remotely. Web-based System Manager Remote Clients can be present in private and open networks. You can perform most management tasks using the Web-based System Manager Remote Client.

The remote HMC and the Web-based System Manager Remote Client allows you the flexibility to access your managed systems (including HMCs) from multiple locations using multiple HMCs.

To install the remote client, complete the following address, and enter it into your machine’s Web browser, and download and run setup.exe.

hostname`/remote_client.html`

Note: To allow a Web-based System Manager connection, you must allow traffic on Port 9090 over the HMC’s public network interface. See **HMC Configuration** → **Customize Network Settings** → **Lan Adapter Details** → **Firewall**.

[4] DHCP stands for Dynamic Host Control Protocol.

For more detailed information about usage of the HMC, refer to the IBM @server Hardware Information Center:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm?info/iphby/asmihmc.htm

2.8.4 HMC connectivity

There are some significant differences regarding the HMC connection to the managed server with the p5-590 and p5-595 servers and other POWER5 processor-based systems.

- At least one HMC is mandatory, and two are recommended.
- The first (or only) HMC is connected using a private network to Bulk Power Controller (BPC-A). The HMC must be setup to provide DHCP addresses on that private (eth0) network.
- A secondary (redundant) HMC is connected using a separate private network to BPC-B. The second HMC must be set up as a DHCP server to use a different range of addresses for DHCP.
- An additional provision has to be made for a HMC connection to the BPC in a powered expansion frame.

Note: DHCP must be used, as the BPCs are dependent upon the HMC to provide them with addresses. There is no way to set a static address on a BPC.

If there is a single managed server (with powered expansion frame), then no additional LAN components are required (Ethernet Cables FC 7801 or FC 7802 or a client provided cable is required). However, if there are multiple managed servers, additional LAN switches and cables will be needed for the HMC private networks. These switches and cables must be planned for.

Figure 2-26 shows two p5-590s controlled with dual HMCs.

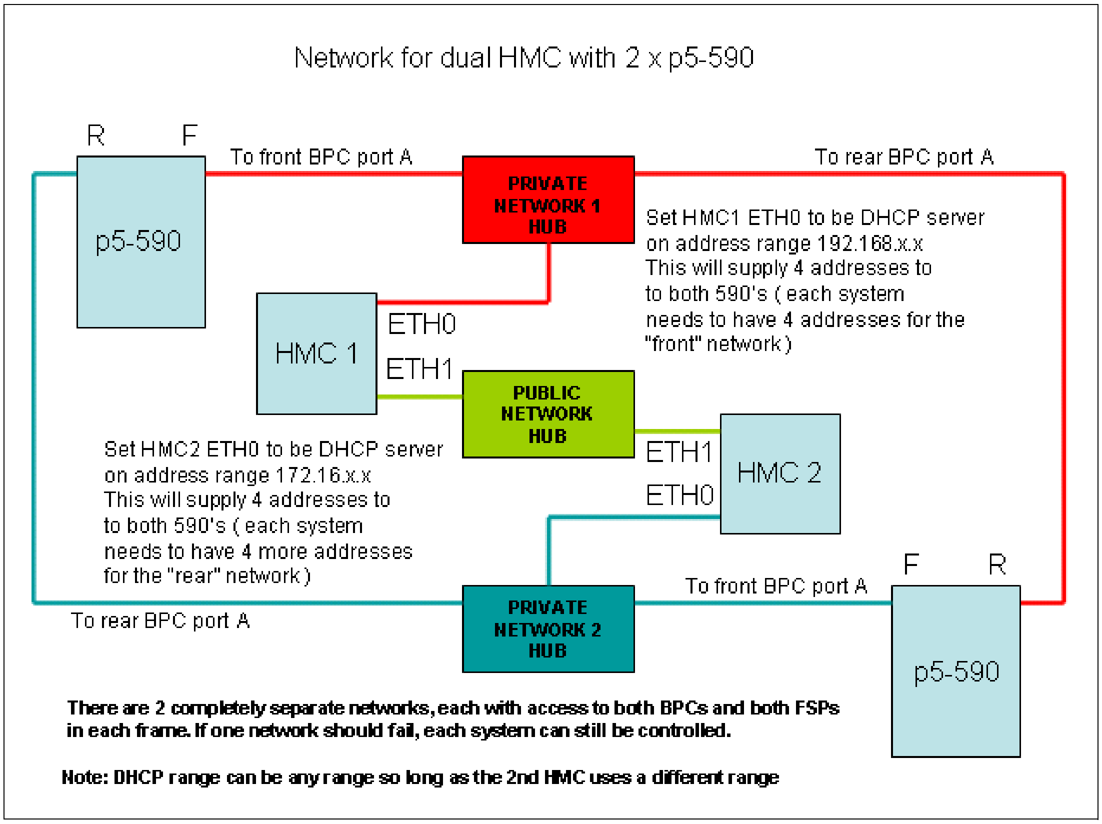

Network for dual HMC with 2 x p5-590
R
F
To front BPC port A
PRIVATE NETWORK 1 HUB
To rear BPC port A
p5-590
Set HMC1 ETH0 to be DHCP server on address range 192.168.x.x This will supply 4 addresses to to both 590's (each system needs to have 4 addresses for the "front" network)
HMC 1
ETH0
ETH1
PUBLIC NETWORK HUB
Set HMC2 ETH0 to be DHCP server on address range 172.16.x.x This will supply 4 addresses to to both 590's (each system needs to have 4 more addresses for the "rear" network)
ETH1
ETH0
HMC 2
F
R
To front BPC port A
To rear BPC port A
PRIVATE NETWORK 2 HUB
p5-590
There are 2 completely separate networks, each with access to both BPCs and both FSPs in each frame. If one network should fail, each system can still be controlled.
Note: DHCP range can be any range so long as the 2nd HMC uses a different range

Figure 2-26 Network for two p5-590s controlled with dual HMCs

HMC network interfaces

The HMC supports up to three separate physical Ethernet interfaces. In the desktop version of the HMC, this consists of one integrated Ethernet and up to two plug-in adapters. In the rack-mounted version, this consists of two integrated Ethernet adapters and up to one plug-in adapter. Use each of these interfaces in the following ways:

- One network interface can be used exclusively for HMC-to-managed system communications (and must be the eth0 connection on the HMC). This means that only the HMC, Bulk-Power Controllers (BPC), and service processors of the managed systems would be on that network. Even though the network interfaces into the service processors are SSL encrypted and password protected, having a separate dedicated network can provide a higher level of security for these interfaces.
- Another network interface would typically be used for the network connection between the HMC and the logical partitions on the managed systems, for the HMC-to-logical partition communications.
- The third interface is an optional additional Ethernet connection that can be used for remote management of the HMC. This third interface can also be used to provide a separate HMC connection to different groups of logical partitions. For example, you could do any of the following:
  - An administrative LAN that is separate from the LAN on which all the usual business transactions are running. Remote administrators could access HMCs and other managed units using this method.
  - Different network security domains for your partitions, perhaps behind a firewall with different HMC network connections into each of those two domains.

Note: With the rack-mounted HMC, if an additional (third) Ethernet port is installed in the HMC (by using a PCI Ethernet card), then that PCI-card becomes the eth0 port. Normally (without the additional card), eth0 is the first of the two integrated Ethernet ports.

2.8.5 HMC code

For updates of the machine code and HMC functions and hardware prerequisites refer to the following Web page:

https://techsupport.services.ibm.com/server/hmc/power5

POWER4 HMC models, such as the 7315-CR2 or 7315-C03 and others, can be upgraded to support POWER5 processor-based systems.

Note: It is not possible to connect POWER4 and POWER5 processor-based systems simultaneously to the same HMC.

To upgrade an existing POWER4 HMC:

- Order FC 0961 for your existing HMC. Contact your IBM Sales Representative for help.
- Call an IBM Service Center and order APAR MB00691.
- Order the CD online by selecting **Version 4.5 machine code updates** → **Order CD** → **Go** at the Hardware Management Console Support for IBM @server i5, @server p5, pSeries, and iSeries Web page at:

  https://techsupport.services.ibm.com/server/hmc/power5

Note: You must have an IBM ID to use this freely available service. Registration information and online registration form can be found at the above Web page.

2.8.6 Hardware management user interfaces

In the following sections we give you a brief overview of the different p5-590 and p5-595 server hardware management user interfaces available.

Advanced System Management Interface

The Advanced System Management Interface (ASMI) is the interface to the service processor that allows you to set flags that affect the operation of the server, such as auto power restart, and to view information about the server, such as the error log and vital product data.

This interface is accessible using a Web browser on a client system that is connected to the service processor on an Ethernet network. It can also be accessed using the HMC. The service processor and the ASMI are standard on all IBM @server i5, @server p5, and OpenPower servers.

Accessing the ASMI using a Web browser

The Web interface to the Advanced System Management Interface is accessible through Microsoft Internet Explorer 6.0, Netscape 7.1, or Opera 7.23 running on a PC or mobile computer connected to the service processor. The Web interface is available during all phases of system operation including the initial program load and runtime. However, some of the menu options in the Web interface are unavailable during IPL or runtime to prevent usage or ownership conflicts if the system resources are in use during that phase.

Accessing the ASMI using a HMC

To access the Advanced System Management Interface using the Hardware Management Console, complete the following steps:

1. Ensure that the HMC is set up and configured.
2. In the navigation area, expand the managed system you want to work with.
3. Expand Service Applications and click **Service Focal Point**.
4. In the content area, click **Service Utilities**.
5. From the Service Utilities window, select the managed system you want to work with.
6. From the Selected menu on the Service Utilities window, select **Launch ASM menu**.

For more detailed information about usage of ASMI please refer to the IBM ⓔserver Hardware Information Center.

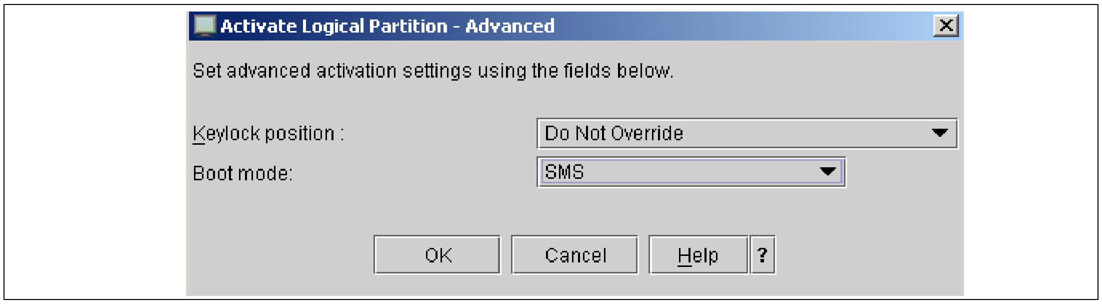
http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm?info/iphau/usingsms.htm

System Management Services

Use the system management services (SMS) menus to view information about your system or partition, and to perform tasks such as setting a password, changing the boot list, and setting the network parameters.

To start the system management services, do the following:

1. Use the HMC to activate the server or partition and choose **Advanced**.
2. From the Boot Mode pull-down menu, select **SMS**. See Figure 2-27.



Activate Logical Partition - Advanced

Set advanced activation settings using the fields below.

Keylock position : Do Not Override

Boot mode: SMS

Figure 2-27 Boot options

For more detailed information about usage of SMS please refer to the IBM ⓔserver Hardware Information Center:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm?info/iphau/usingsms.htm

2.8.7 Determining HMC serial number

For some HMC or service processor troubleshooting situations an IBM service representative will have to sign into the HMC. The service password changes daily and is not available for normal client use. If the PE determines a local service engineer can sign on to the HMC, the service representative may request the HMC serial number.

To find the HMC serial number, open a restricted shell window and run the following command:

```
#lshmc -v
```

2.8.8 Firmware

Depending on your service environment, you can download your server firmware fixes using different interfaces and methods. The p5-590 and p5-595 servers must use the HMC to install server firmware fixes. Firmware is loaded on to the server and to the bulk power controller over the HMC to the frame's Ethernet network.

Server firmware

Server firmware is the part of the Licensed Internal Code that enables hardware, such as the service processor. Check for available server firmware fixes regularly, and download and install the fixes if necessary. The server firmware binary image is a single image that includes code for the service processor, the POWER Hypervisor firmware, and platform partition firmware. This server firmware binary image is stored in the service processor's flash memory and executed in the service processor main memory.

Since there are dual service processors per CEC, both service processors are updated when firmware updates are applied and activated using the Licensed Internal Code Updates section of the HMC.

Firmware is available for download at:

```
http://techsupport.services.ibm.com/server/mdownload/systems.html
```

> **Note:** Firmware version 01SF230_120 and later provides support for concurrent firmware maintenance (CFM). Only updates within a release can be concurrent. However, some updates that are for critical problems may be designated disruptive even if they are *within* a release. See "Firmware level naming convention" on page 62 for information regarding *release* and *fixpack*.

Power subsystem firmware

Power subsystem firmware is the part of the Licensed Internal Code that enables the power subsystem hardware in the model p5-590 and p5-595 servers. You must use an HMC to update or upgrade power subsystem firmware fixes.

The bulk power controller (BPC) has its own service processor. The power firmware not only has the code load for the BPC service processor itself, but it also has the code for the distributed converter assemblies (DCAs), bulk power regulators (BPRs), fans, and other more granular field replaceable units that have firmware to help manage the frame and its power and cooling controls. The BPC service processor code load also has the firmware for the cluster switches that may be installed in the frame.

In the same way that the Central Electronics Complex (CEC) has dual service processors, the power subsystem has dual BPCs. Both are updated when firmware changes are made using the Licensed Internal Code Updates section of the HMC.

The BPC initialization sequence Central Electronics Complex after the reboot is unique. The BPC service processor must check the code levels of all the power components it manages, including DCAs, BPRs, fans, cluster switches, and it must load those if they are different than what is in the active flash side of the BPC. Code is cascaded to the downstream power components over universal power interface controller (UPIC) cables.

Platform initial program load

The main function of the p5-590 and p5-595 service processors is to initiate platform initial program load (IPL), also referred to as platform boot. The service processor has a self-initialization procedure, and then initiates a sequence of initializing and configuring many components on the CEC backplane.

The service processor has various functional states, which can be queried and reported to the POWER Hypervisor component. Service processor states include, but are not limited to, standby, reset, power up, power down, and runtime. As part of the IPL process, the primary service processor will check the state of the backup. The primary service processor is responsible for reporting the condition of the backup service processor to the POWER Hypervisor component. The primary service processor will wait for the backup service processor to indicate that it is ready to continue with the IPL (for a finite time duration). If the backup service processor fails to initialize in a timely fashion, the primary will report the backup service processor as a non-functional device to the POWER Hypervisor component and will mark it as a *garded* resource before continuing with the IPL. The backup service processor can later be integrated into the system.

Open Firmware

IBM @server p5 and OpenPower servers have one instance of Open Firmware, both when in the partitioned environment and when running as a full system partition. Open Firmware has access to all devices and data in the system. Open Firmware is started when the system goes through a power-on reset. Open Firmware, which runs in addition to the Hypervisor firmware in a partitioned environment, runs in two modes: Global and partition. Each mode of Open Firmware shares the same firmware binary that is stored in the flash memory.

In a partitioned environment, partition Open Firmware runs on top of the global Open Firmware instance. The partition Open Firmware is started when a partition is activated. Each partition has its own instance of Open Firmware and has access to all the devices assigned to that partition. However, each instance of partition Open Firmware has no access to devices outside of the partition in which it runs. Partition firmware resides within the partition memory and is replaced when AIX 5L takes control. Partition firmware is needed only for the time that is necessary to load AIX 5L into the partition system memory.

The global Open Firmware environment includes the partition manager component. That component is an application in the global Open Firmware that establishes partitions and their corresponding resources (such as CPU, memory, and I/O slots), which are defined in partition profiles. The partition manager manages the operational partitioning transactions. It responds to commands from the service processor external command interface that originate in the application that is running on the HMC.

For more information on Open Firmware refer to *Partitioning Implementations for IBM @server p5 Servers,* SG24-7039, at:

```
http://www.redbooks.ibm.com/redpieces/abstracts/SG247039.html?Open
```

Temporary and permanent side of the service processor

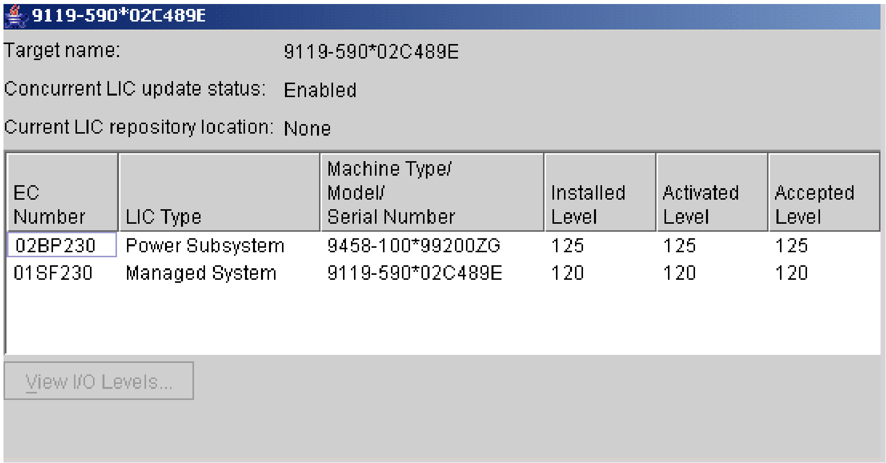
The service processor and the BPC maintain two copies of the firmware.

- One copy is considered the permanent or backup copy and is stored on the permanent side, sometimes referred to as the *p* side.
- The other copy is considered the installed or temporary copy and is stored on the temporary side, sometimes referred to as the *t* side. It is recommended that you start and run the server from the temporary side.
- The copy actually booted from is called the activated level, sometimes referred to as *b*.

The concept of *sides* is an abstraction. The firmware is located in flash memory and pointers in nvram determine which is *p* and *t*.

Note: The default value the system will boot is *temporary*.

To view the firmware levels on the HMC, select **Licensed Internal Code → Updates Change Internal Code → Select managed system → View System Information → None**, and the screen in Figure 2-28 on page 62 is displayed. (The power subsystem is always machine type 9458, whereas the server is machine type 9119.)



9119-590*02C489E

Target name: 9119-590*02C489E

Concurrent LIC update status: Enabled

Current LIC repository location: None

EC Number	LIC Type	Machine Type/ Model/ Serial Number	Installed Level	Activated Level	Accepted Level
02BP230	Power Subsystem	9458-100*99200ZG	125	125	125
01SF230	Managed System	9119-590*02C489E	120	120	120

Figure 2-28 p5-590 and p5-595 code levels

The levels are:

- The Installed Level indicates the level of firmware that has been installed and will be installed into memory after the managed system is powered off and powered on using the default temporary side.
- The Activated Level indicates the level of firmware that is active and running in memory.
- The Accepted Level indicates the backup level (or permanent side) of firmware. You can return to the backup level of firmware if you decide to remove the installed level.

The following example is the output of the `lsmcode` command for AIX 5L and Linux, showing the firmware levels as they are displayed in the outputs:

- AIX 5L:

  ```
  The current permanent system firmware image is SF230_120
  The current temporary system firmware image is SF230_120
  The system is currently booted from the temporary firmware image.
  ```

  The `lsmcode` command is part of bos.diag.util.
- Linux:

  ```
  system:SF230_120 (t) SF230_120 (p) SF230_120 (b)
  ```

Firmware level naming convention

The naming convention is:

- 01SF225_096 → PPNNSSS_FFF
  - PP - Package identifier: 01 = managed system, 02 = power code
  - NN - Machine type; SF = POWER5, system BP = bulk power code
  - SSS - Release level
  - FFF - Fixpack number

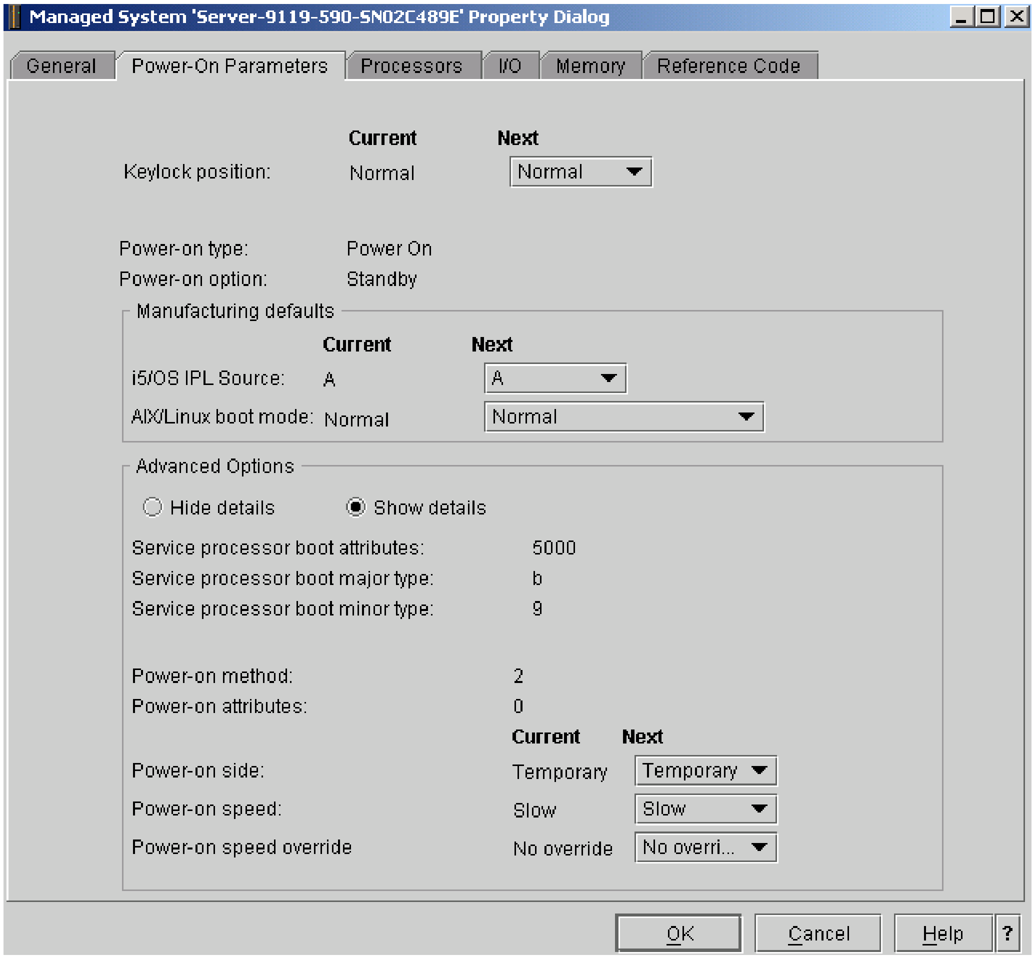
Note: The following points are of special interest:

- The server firmware fix is installed on the temporary side only after the existing contents of the temporary side are permanently installed on the permanent side (the service processor performs this process automatically when you install a server firmware fix).
- If you want to preserve the contents of the permanent side, you need to remove the current level of firmware (copy the contents of the permanent side to the temporary side) before you install the fix.
- However, if you get your fixes using Advanced features on the HMC interface and you indicate that you do not want the service processor to automatically accept the firmware level, the contents of the temporary side are not automatically installed on the permanent side. In this situation, you do not need to remove the current level of firmware to preserve the contents of the permanent side before you install the fix.

You might want to use the new level of firmware for a period of time to verify that it works correctly. When you are sure that the new level of firmware works correctly, you can permanently install the server firmware fix. When you permanently install a server firmware fix, you copy the temporary firmware level from the temporary side to the permanent side.

Conversely, if you decide that you do not want to keep the new level of server firmware, you can remove the current level of firmware. When you remove the current level of firmware, you copy the firmware level that is currently installed on the permanent side from the permanent side to the temporary side.

Choosing which firmware to use when powering on the system is done using the Power-On Parameters tab in the server properties box, as shown in Figure 2-29 on page 64.



	Current	Next
Keylock position:	Normal	Normal

Power-on type: Power On

Power-on option: Standby

Manufacturing defaults

	Current	Next
i5/OS IPL Source:	A	A
AIX/Linux boot mode:	Normal	Normal

Advanced Options

○ Hide details ● Show details

Service processor boot attributes: 5000

Service processor boot major type: b

Service processor boot minor type: 9

Power-on method: 2

Power-on attributes: 0

	Current	Next
Power-on side:	Temporary	Temporary
Power-on speed:	Slow	Slow
Power-on speed override	No override	No overri...

Figure 2-29 Power on parameters

For a detailed description of firmware levels refer to the IBM @server Hardware Information Center and select **Service and support** → **Customer service and support** → **Getting fixes** → **Firmware (Licensed Internal Code) fixes** → **Concepts** → **Temporary and permanent side of the service processor** at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

Get server firmware fixes using an HMC

You use an HMC to manage your server, and you have configured several partitions on the server. Periodically, you need to download and install fixes for your server and power subsystem firmware.

You want to use the HMC to perform this task. How you get the fix depends on whether the HMC or server is connected to the Internet.

- If the HMC or server is connected to the Internet:

  There are several repository locations from which you can download the fixes using the HMC. For example, you can download the fixes from your service provider's Web site or support system, from optical media that you order from your service provider, or from an FTP server on which you previously placed the fixes.

- If neither the HMC nor your server is connected to the Internet:

  You will need to download your new system firmware level to a CD-ROM media or FTP server.

For both of these options, you can use the interface on the HMC to install the firmware fix (from one of the repository locations or from the optical media). The Change Internal Code wizard on the HMC provides a step-by-step process for you to perform the required steps to install the fix.

1. Ensure that you have a connection to the service provider (if you have an Internet connection from the HMC or server).
2. Determine the available levels of server and power subsystem firmware.
3. Create optical media (if you do not have an Internet connection from the HMC or server).
4. Use the Change Internal Code wizard to update your server and power subsystem firmware.

   > **Note:** The tasks in the Licensed Internal Code Updates view on the HMC vary according to HMC code level:
   >
   > - Version 4.4.x and earlier:
   >   - Use Change Internal Code to update within a release.
   >   - Use Manufacturing Equipment Specification Upgrade to upgrade to a new release.
   > - Version 4.5.x and later:
   >   - Use “Change Licensed Internal Code for the current release” to update within a release.
   >   - Use “Upgrade Licensed Internal Code to a new release” to upgrade to a new release.

5. Verify that the fix installed successfully.

For a detailed description of each task go to the IBM @server Hardware Information Center and select **Service and support** → **Customer service and support** → **Getting fixes** → **Firmware (Licensed Internal Code) fixes** → **Scenarios: Firmware (Licensed Internal Code) fixes** → **Scenario: Get server firmware fixes using Task an HMC** at:

`http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm`

Note: To view existing levels of server firmware using the **lsmcode** command, you need to have the following service tools installed on your server:

- AIX 5L

  You must have AIX 5L diagnostics installed on your server to perform this task. AIX 5L diagnostics are installed when you install the AIX 5L operating system on your server. However, it is possible to deinstall the diagnostics. Therefore, you need to ensure that the online AIX 5L diagnostics are installed before proceeding with this task.

- Linux

  - Platform Enablement Library - librtas-*xxxxx*.rpm
  - Service Aids - ppc64-utils-*xxxxx*.rpm
  - Hardware Inventory - lsvpd-*xxxxx*.rpm

    Where *xxxxx* represents a specific version of the RPM file.

  If you do not have the service tools on your server, you can download them at the following Web page:

  http://techsupport.services.ibm.com/server/lopdiags

For a detailed description of each task go to the IBM @server Hardware Information Center and select **Service and support** → **Customer service and support** → **Getting fixes** → **Firmware (Licensed Internal Code) fixes** → **Scenarios: Firmware (Licensed Internal Code) fixes** → **Scenario: Get server firmware fixes without an HMC** at:

http://publib.boulder.ibm.com/infocenter/eserver/v1r2s/en_US/index.htm

Capacity on Demand

Through CoD offerings, IBM p5-590 and p5-595 servers can offer either permanent or temporary increases or decreases in processor and memory capacity. CoD is available in four activation configurations, each with specific pricing and availability terms. The four types of CoD activation configurations are discussed within this chapter, from a functional standpoint. Contractual and pricing issues are outside the scope of this document and should be discussed with your IBM Global Financing Representative, IBM Business Partner, or IBM Sales Representative.

Capacity on Demand is supported by the following operating systems:

- AIX 5L Version 5.2 and Version 5.3
- i5/OS V5R3 or later
- SUSE LINUX Enterprise Server 9 for POWER or later
- Red Hat Enterprise Linux AS 3 for POWER (update 4) or later

For additional information about Capacity on Demand, see:

http://www.ibm.com/servers/eserver/pseries/ondemand/cod/

3.1 Types of Capacity on Demand

Capacity on Demand for the @server p5 systems with dynamic logical partitioning (dynamic LPAR) offers system owners the ability to non-disruptively activate processors and memory without rebooting partitions. CoD also gives @server p5 systems owners the option to temporarily activate processors to meet varying performance needs and to activate additional capacity on a trial basis.

IBM has established four types of CoD offerings on the p5-590 and p5-595 systems, each with a specific activation plan. Providing different types of CoD offerings gives clients flexibility when determining their resource needs and establishing their IT budgets. IBM Global Financing can help match individual payments with capacity usage and competitive financing for fixed and variable costs related to IBM Capacity on Demand offerings. By financing Capacity on Demand costs and associated charges together with a base lease, spikes in demand need not become spikes in a budget.

After a system with CoD features is delivered, it can be activated in the following ways:

- Capacity Upgrade on Demand (CUoD) for processors and memory
- On/Off Capacity on Demand (CoD) for processors and memory
- Reserve Capacity on Demand (CoD) for processors only
- Trial Capacity on Demand (CoD) for processors and memory

The p5-590 and p5-595 servers use specific feature codes to enable CoD capabilities. All types of CoD transactions for processors are in whole numbers of processors, not in fractions of processors. All types of CoD transactions for memory are in 1 GB increments.

Table 3-1 on page 69 provides a brief description of the four types of CoD offerings, identifies the proper name of the associated activation plan, and indicates the default type of payment offering and scope of enablement resources. The payment offering information is intended for reference only; all pricing agreements and service contracts should be handled by your IBM representative. A functional overview of each CoD offering is provided in the subsequent sections.

Table 3-1 Types of Capacity on Demand (functional categories)

Activation plan	Functional category	Applicable system resources	Type of payment offering	Description
Capacity Upgrade on Demand	Permanent capacity for nondisruptive growth	Processor and memory resources	Pay when purchased	Provides a means of planned growth for clients who know they will need increased capacity but are not sure when
On/Off Capacity on Demand (CoD)	Temporary capacity for fluctuating workloads	Processor and memory resources	Pay after activation	Provides for planned and unplanned short-term growth driven by temporary processing requirements such as seasonal activity, period-end requirements, or special promotions
Reserve Capacity on Demand		Processor resources only	Pay before activation	
Trial Capacity on Demand	Temporary capacity for workload testing or any one-time need	Processor and memory resources	One-time, no-cost activation for a maximum period of 30 consecutive days	Provides the flexibility to evaluate how additional resources will affect existing workloads, or to test new applications by activating additional processing power or memory capacity (up to the limit installed on the server) for up to 30 contiguous days
Capacity Backup	Disaster recovery	Offsite machine	Pay when purchased	Provides a means to purchase a machine for use when off-site computing is required, such as during disaster recovery

3.1.1 Capacity Upgrade on Demand (CUoD) for processors

Capacity Upgrade on Demand (CUoD) for processors is available for the p5-590 and p5-595 servers. CoD for processors allows inactive processors to be installed in the p5-590 and p5-595 server and can be permanently activated by the customer as required.

All processor books available on the p5-590 and p5-595 are initially implemented as 16-way CoD offerings with zero active processors.

A minimum of 8 or 16 permanently activated processors are required on the p5-590 or p5-595 server, respectively.

The number of permanently activated processors is based on the number of processor books installed as follows:

- One processor book installed requires 8 (p5-590) or 16 (p5-595) permanently activated processors.
- Two processor books installed requires 16 permanently activated processors.
- Three processor books installed requires 24 permanently activated processors.
- Four processor books installed requires 32 permanently activated processors.

Additional processors on the CoD books are activated in increments of one by ordering the appropriate activation feature number. If more than one processor is to be activated at the same time, the activation feature should be ordered in multiples.

After receiving an order for a CUoD for the processors activation feature, IBM will provide the customer with a 34-character encrypted key. This key is entered into the system via the HMC to activate the desired number of additional processors.

CUoD processors that have not been activated are available to the p5-595 server for dynamic processor sparing when running the AIX 5L operating system. If the server detects the impending failure of an active processor, it will attempt to activate one of the unused CoD processors and add it to the system configuration. This helps to keep the server's processing power at full strength until a repair action can be scheduled.

3.1.2 Capacity Upgrade on Demand for memory

Capacity Upgrade on Demand (CUoD) for memory is available for p5-590 and p5-595 servers. CUoD for memory allows inactive memory to be installed in the p5-590 or p5-595 server and can be permanently activated by the customer as required.

CUoD for memory may be used in any available memory position.

Additional CoD memory cards are activated in increments of 1 GB by ordering the appropriate activation feature number. If more than one 1 GB memory increment is to be activated at the same time, the activation code should be ordered in multiples.

After receiving an order for a CUoD for memory activation feature, IBM will provide the customer with a 34-character encrypted key. This key is entered into the system to activate the desired number of additional 1 GB memory increments.

Memory configuration rules for the p5-590 and p5-595 servers apply to CUoD for memory cards as well as conventional memory cards. The memory configuration rules are applied based upon the maximum capacity of the memory card.

- Apply 4 GB configuration rules for 4 GB CoD for memory cards with less than 4 GB of active memory.
- Apply 8 GB configuration rules for 8 GB CoD for memory cards with less than 8 GB of active memory.

3.1.3 On/Off Capacity on Demand (On/Off CoD)

On/Off Capacity on Demand (On/Off CoD) is available for p5-590 and p5-595 servers. On/Off CoD allows customers to temporarily activate installed CUoD processors and memory resources and later deactivate the resources as desired.

On/Off processor and memory resources are implemented on a *pay-as-you-go* basis using:

- On/Off Processor and Memory Enablement features - Signing an On/Off Capacity on Demand contract is required. An enablement code will be supplied to activate the enablement feature.
- After the On/Off Enablement feature is ordered and the associated enablement code is entered into the system, the customer must report on/off usage to IBM at least monthly. This information, which is used to compute the billing data on a quarterly basis, is provided to the sales channel, which will place an order for the quantity of On/Off Processor Day and Memory Day billing features used and invoice the customer.

Each On/Off CoD (Capacity on Demand) enablement feature provides 360 processor days of available usage under On/Off CoD for processors. When the user is near this total, a new enablement feature should be ordered at no charge. Enablement features cannot be added; each new feature resets the available amount to 360 days.

- On/Off CoD billing is based on processor days and memory GB days. A processor day is charged at the time of activation and entails the use of the processor for the next 24-hour period. Note that if a user de-activates a processor and re-activates it a second time, they will be billed for a second processor day even though the re-activation may fall within the initial 24-hour window of entitlement. The same method of charging applies to memory activations with On/Off CoD.

- Each time processors are activated starts a new measurement day. If a customer activates four processors for a two-hour test and later in the same 24-hour period activates two processors for two hours to meet a peak workload, the result is six processor days of usage.

3.1.4 Reserve Capacity on Demand (Reserve CoD)

Reserve Capacity on Demand (Reserve CoD) is available for p5-590 and p5-595 servers. Reserve CoD is an innovative offering allowing clients to temporarily activate in an automated manner installed CoD processors used within a shared processor pool. Charges for the temporary activation of Reserve CoD processors are only incurred when processing needs exceed the fully entitled level.

Reserve CoD is a pre-pay method of temporary activation. It is ordered by purchasing the quantity of Reserve CoD features appropriated for the model and speed of installed processors. Each feature includes 30 days of temporary usage time. When Reserve CoD is ordered, the user will receive a 34-digit activation code to be entered at the HMC. The activation code will establish the Reserve CoD balance of available usage time. Inactive CoD processors can then be assigned to the shared processor pool, which will be available for workload processing. Charges for the inactive processors will only be incurred when the workload in the shared pool exceeds 100 percent of the entitled (permanently activated) level of performance. Charges are made against the Reserve CoD account balance in increments of processor days and Advanced Power Virtualization must be activated in order to use Reserve CoD.

3.1.5 Trial Capacity on Demand

Trial Capacity on Demand (Trial CoD) is a function delivered with all pSeries servers supporting CUoD resources beginning May 30, 2003. Those servers with standby CoD processors or memory will be capable of using a one-time, no-cost activation for a maximum period of 30 consecutive days. This enhancement allows for benchmarking of CoD resources or can be used to provide immediate access to standby resources when the purchase of a permanent activation is pending.

Trial CoD is a complimentary service offered by IBM. Although IBM intends to continue it for the foreseeable future, IBM reserves the right to withdraw Trial CoD at any time, with or without notice.

3.1.6 Capacity on Demand feature codes

The CoD feature codes used to order CoD capabilities on the p5-590 and p5-595 are summarized in Table 3-2.

Table 3-2 p5-590 and p5-595 CoD feature codes

Inactive resource feature		CoD feature codes			
Description	**Feature code (FC)**	**Permanent activation CUoD**	**Reserve CoD**	**Processor**	**Memory**
				On/Off Enablement / Billing	
p5-590					
0/16 Processors (1.65 GHz POWER5)	FC 7981	FC 7925	FC 7926	FC 7839 / FC 7993	
2/4 GB DDR1 Memory	FC 7816	FC 7970			FC 7973 / FC 7974
4/8 GB DDR1 Memory	FC 7835	FC 7970			FC 7973 / FC 7974
p5-595					
0/16 Processors (1.65 GHz POWER5)	FC 7988	FC 7990	FC 7991	FC 7994 / FC 7996	
0/16 Processors (1.9 GHz POWER5)	FC 7813	FC 7815	FC 7975	FC 7971 / FC 7972	
2/4 GB DDR1 Memory	FC 7816	FC 7970			FC 7973 / FC 7974
4/8 GB DDR1 Memory	FC 7835	FC 7970 FC 7799*			FC 7973 / FC 7974
* FC 7799 enables 256 1 GB memory activations (at one time) for FC 7835 on p5-595 only.					

3.1.7 Capacity BackUp

Also available are three new Capacity BackUp features for configuring systems used for disaster recovery. Capacity BackUp for IBM @server p5 590 and 595 systems offers an offsite, disaster recovery machine at an affordable price. This disaster recovery machine has primarily inactive Capacity on Demand (CoD) processors that can be activated in the event of a disaster. Capacity BackUp for IBM @server p5 offering includes:

- Four processors that are permanently activated and can be used for any workload.
- Either 28 or 60 standby processors to be used in the event of a disaster.
- Either 900 (4/32-way) or 1800 (4/64-way) of On/Off CoD processor days available for testing or for use in the event of a disaster.

Capacity BackUp systems can be turned on at any time by using the On/Off CoD activation procedure for the needed performance during an unplanned system outage. Each Capacity

BackUp configuration is limited to 450 On/Off CoD credit days per processor book. For clients who require additional capacity or processor days, additional processor capacity can be purchased under IBM CoD at regular On/Off CoD activation prices. IBM HACMP™ V5 and HACMP/XD software (5765-F62), when installed, can automatically activate Capacity BackUp resources upon failover. When needed, HACMP can also activate dynamic LPAR and CoD resources.

Capacity BackUp for p5 servers is offered in configurations limited to the following:

- p5-590:
  - 4/32 standard processor (1.65 GHz) CBU system: Must configure 2 x FC 7730 and 4 x FC 7925
- p5-595:
  - 4/32 standard processor (1.65 GHz) CBU system: Must configure 2 x FC 7732 and 4 x FC 7990
  - 4/32 turbo processor (1.9 GHz) CBU system: Must configure 2 x FC 7731 and 4 x FC 7815
  - 4/64 standard processor (1.65 GHz) CBU system: Must configure 4 x FC 7732 and 4 x FC 7990
  - 4/64 turbo processor (1.9 GHz) CBU system: Must configure 4 x FC 7731 and 4 x FC 7815

I/O and memory minimums and maximums are the same as the IBM p5-590 and p5-595 offerings.

4

Reliability, availability, and serviceability

IBM's reliability, availability, and serviceability (RAS) philosophy employs a well thought out and organized architectural approach to:

- Avoid problems, where possible with a well engineered design.
- Should a problem occur, attempt to recover or retry the operation.
- Diagnose the problem and reconfigure the system as needed
- Automatically initiate a repair and call for service.

As a result, IBM servers are designed for reliable, robust operation in a wide variety of demanding environments.

The following chapter provides more detailed information about IBM eServer p5 590 and p5 595 reliability, availability, and serviceability features. It includes several features on the benefits available when using AIX 5L. Support of these features using Linux can vary.

4.1 Reliability, fault tolerance, and data integrity

The base reliability of a computing system is, at the most fundamental level, dependent upon the intrinsic failure rates of the components that comprise it. Simply stated, highly reliable servers are built with highly reliable components. This basic premise is augmented with a clear design for reliability, architecture, and methodology. For the past decade, IBM RAS engineers have been systematically adding mainframe-inspired RAS technologies to IBM UNIX OS offerings, resulting in dramatically improved system designs.

Normally, servers with fewer components, with fewer interconnects, have fewer chances to fail. Seemingly simple design choices (for example, integrating two processor cores on a single POWER5 chip) can dramatically reduce the opportunity for server failure. In this case, a 64-way server will include half as many processor chips as with a single CPU per processor design. Not only will this reduce the total number of system components, it will reduce the total amount of heat generated in the design, resulting in an additional reduction in required power and cooling components. Finally, the p5-590 and p5-595 use a very high degree of integration by employing IBM MCM (multi-chip module) designs using the same technology as is currently deployed in zSeries® (mainframe) servers. The POWER5 MCM uses a glass ceramic module that holds four POWER5 microprocessors (8-way modules) and four L3 cache modules. All connections between the on-board components are included in wiring embedded in the substrate, and all system connections are routed through the module to the system board. Not only does this result in a high-performance, highly scalable system package that carefully controls the interconnect speeds and manages the heat, but it also reduces the total number of supporting chips per 8-way module by eight chips over its POWER4 predecessor.

As discussed, system packaging can have a significant impact on server reliability. Since the reliability of electronic components is directly related to their thermal environment (large decreases in component reliability can be measured due to relatively small increases in temperature), IBM servers are carefully packaged to insure adequate cooling. Critical system components (POWER5 processor chips, for example) are positioned so that they receive *upstream* or *fresh* air, while less sensitive or lower power components like memory DIMMs are positioned *downstream*. In addition, POWER5 processor-based servers are built with redundant, variable speed fans that are designed to automatically increase their output to compensate for increased heat in the central electronic complex. During the design and development process, subsystems go through rigorous verification and integration testing processes. During system manufacturing, systems go through a thorough testing process designed to help ensure the highest level of product quality.

- The p5-590 and p5-595 server L3 cache and system memory offers ECC (error checking and correcting) fault-tolerant features. ECC is designed to correct environmentally induced, single-bit, intermittent memory failures and single-bit hard failures. With ECC, the likelihood of memory failures will be substantially reduced.
- ECC also provides double-bit memory error detection that helps protect data integrity in the event of a double-bit memory failure.
- System memory also provides 4-bit packet error detection that helps to protect data integrity in the event of a DRAM chip failure.
- The system bus, I/O bus, and PCI buses are designed with parity error detection.
- Disk mirroring and disk controller duplexing are also provided by the AIX 5L operating system. Linux supports disk mirroring (RAID 1). This is supported in software using the md driver. Some of the hardware RAID adapters supported under Linux also support mirroring.

The Journaled File System maintains file system consistency and reduces the likelihood of data loss when the system is abnormally halted due to a power failure.

4.1.1 PCI extended error handling

In the past, PCI bus parity errors caused a global machine check interrupt, which eventually required a system reboot to continue. In the eserver p5 systems, new I/O drawer hardware, system firmware, and AIX 5L interaction have been designed to allow transparent recovery of intermittent PCI bus parity errors and graceful transition to the I/O device available state in the case of a permanent parity error in the PCI bus. This mechanism is called PCI extended error handling (EEH).

EEH-enabled adapters respond to a special data packet generated from the affected PCI slot hardware by calling system firmware, which will examine the affected bus, allow the device driver to reset it, and continue without a system reboot.

Note: This RAS function is not supported under Linux.

IBM has long built servers with redundant physical I/O paths using CRC checking and failover support to protect RIO server connections from the CEC to the I/O drawers. IBM extended data protection in the pSeries servers, enhancing the extended error handling to allow recovery from PCI-bus error conditions. The eserver p5 systems add additional recovery features to handle potential errors in the Processor Host Bridge (PCI bridge), and the GX+ bus adapter. These features provide improved diagnosis, isolation, and management of errors in the server I/O path and new opportunities for concurrent maintenance to allow faster recovery from I/O path errors, often without impact to system operation.

4.1.2 Memory error correction extensions

There are several levels of memory protection implemented on the p5-590 and p5-595 systems. From the internal L1 caches to the main memory, several features are implemented to assure data integrity and data recovery in case of memory failures.

- The p5-590 and p5-595 servers uses Error Checking and Correcting (ECC) circuitry for memory reliability, fault tolerance, and integrity.
- Memory has single-error-correct and double-error-detect ECC circuitry designed to correct single-bit memory failures. The *double-bit* detection is designed to help maintain data integrity by detecting and reporting multiple errors beyond what the ECC circuitry can correct.
- The memory chips are organized such that the failure of any specific memory module only affects a single-bit within an ECC word (*bit-scattering)*, thus allowing for error correction and continued operation in the presence of a complete chip failure (Chipkill™ recovery).
- The memory also utilizes memory scrubbing and thresholding to determine when spare memory modules, within each bank of memory, if available, should be used to replace ones that have exceeded their threshold value (*dynamic bit-steering*). Memory scrubbing is the process of reading the contents of the memory during idle time and checking and correcting any single-bit errors that have accumulated by passing the data through the ECC logic. This function is a hardware function on the memory controller chip and does not influence normal system memory performance.

4.1.3 Redundancy for array self-healing

Although the most likely failure event in a processor is a soft single-bit error in one of its caches, there are other events that can occur, and they need to be distinguished from one another.

- For the L1, L2, and L3 caches and their directories, hardware and firmware keep track of whether permanent errors are being corrected beyond a threshold. If this threshold is exceeded, a deferred repair error log is created. Additional run-time availability actions, such as CPU vary off[1] or L3 cache line delete, are also initiated.
- L1 and L2 caches and L2 and L3 directories on the POWER5 chip are manufactured with spare bits in their arrays that can be accessed using programmable steering logic to replace faulty bits in the respective arrays. This is analogous to the redundant bit-steering employed in main storage as a mechanism that is designed to help avoid physical repair, and is also implemented in POWER5 systems. The steering logic is activated during processor initialization and is initiated by the built-in self-test (BIST) at power-on time.
- L3 cache redundancy is implemented at the cache line level. Exceeding correctable error thresholds while running causes a dynamic L3 cache line delete function to be invoked.

4.1.4 First Failure Data Capture

Diagnosing problems in a computer is a critical requirement for autonomic computing. The first step to producing a computer that truly has the ability to self-heal is to create a highly accurate way to identify and isolate hardware errors. IBM has implemented a server design that builds in hardware error-check stations that capture and help to identify error conditions within the server. Each of these checkers is viewed as a diagnostic probe into the server, and, when coupled with extensive diagnostic firmware routines, allows quick and accurate assessment of hardware error conditions at runtime.

- First Failure Data Capture (FFDC) check stations are carefully positioned within the server logic and data paths to help ensure that potential errors can be quickly identified and accurately tracked to an individual field replaceable unit (FRU).
- These checkers are collected in a series of Fault Isolation Registers, where they can easily be accessed by the service processor.
- All communication between the SP and the FIR is accomplished *out of band*. That is, operation of the error-detection mechanism is transparent to an operating system. This entire structure is *below the architecture* and is not seen, nor accessed, by system-level activities.

4.1.5 Dynamic CPU Deallocation

Dynamic CPU Deallocation has been available since AIX Version 4.3.3 on previous RS/6000 and pSeries systems, and is the ability of a system to automatically deconfigure an error-prone CPU before it causes an unrecoverable system error (unscheduled server outage). It is part of the p5-590 and p5-595 RAS features.

CPU dynamic deconfiguration relies on the service processor’s ability to use FFDC-generated recoverable-error information and to notify the AIX 5L operating system when the CPU reaches its predefined error limit. AIX 5L will then *drain* the run-queue for that CPU, redistribute the work to the remaining CPUs, deallocate the offending CPU, and continue normal operation, although potentially at a lower level of system performance. While AIX Version 4.3.3 precluded the ability for a SMP server to revert to a uniprocessor (for example, a 2-way to a 1-way configuration), this limitation was lifted with AIX 5L Version 5.1.

[1] This RAS function is only available for a Linux operating system running the 2.6 kernel.

AIX 5L Version 5.2 support for dynamic logical partitioning (dynamic LPAR) allowed additional system availability improvements. An IBM @server p5 server that includes an inactive CPU (an unused CPU included in a Capacity Upgrade on Demand (CUoD) system configuration) can be configured for CPU hot-sparing. In this case, as a system option, the inactive CPU can automatically be used to *back-fill* for the deallocated bad processor. In most cases, this operation is transparent to the system administrator and to end users. The spare CPU is logically moved to the target system partition or shared processor pool, AIX 5L and Linux moves the workload, and the failing processor is deallocated. The server continues normal operation with full function and full performance. The system will generate an error message for inclusion in the error logs calling for deferred maintenance of the faulty component.

4.1.6 Service processor

The service processor included in the p5-590 and p5-595 servers is designed for an immediate means to diagnose, check status, and sense operational conditions of a remote system, even when the main processor is inoperable.

- The service processor enables firmware and operating system surveillance, several remote power controls, environmental monitoring (only critical errors are supported under Linux), reset, boot features, remote maintenance, and diagnostic activities, including console mirroring.
- The service processor can place calls to report surveillance failures, critical environmental faults, and critical processing faults.

For more detailed information on the service processor refer to 2.8.2, “Service processor” on page 51.

4.1.7 Fault monitoring functions

The following are a few of the fault monitoring systems included with p5-590 and p5-595 servers.

- Built-in self-test (BIST) and power-on self-test (POST) check the processor, L3 cache, memory, and associated hardware required for proper booting of the operating system every time the system is powered on. If a noncritical error is detected or if the errors occur in the resources that can be removed from the system configuration, the booting process is designed to proceed to completion. The errors are logged in the system nonvolatile RAM (NVRAM).
- Disk drive fault tracking can alert the system administrator of an impending disk failure before it impacts client operation.
- The AIX 5L or Linux log (where hardware and software failures are recorded and analyzed by the Error Log Analysis (ELA) routine) warns the system administrator about the causes of system problems. This also enables IBM service representatives to bring along probable replacement hardware components when a service call is placed, thus minimizing system repair time.

4.1.8 Mutual surveillance

The service processor monitors the operation of the POWER Hypervisor firmware during the boot process and watches for loss of control during system operation. It also allows the POWER Hypervisor firmware to monitor service processor activity.

The service processor can take appropriate action, including calling for service, when it detects the POWER Hypervisor firmware has lost control. Likewise, the POWER Hypervisor firmware can request a service processor repair action if necessary.

4.1.9 Environmental monitoring functions

The following are some of the environmental monitoring functions available for the p5-590 and p5-595 servers.

- Temperature monitoring increases the fan speed rotation when ambient temperature is above the normal operating range.

  Temperature monitoring warns the system administrator of potential environmental related-problems (for example, air conditioning and air circulation around the system) so that appropriate corrective actions can be taken before a critical failure threshold is reached. It also performs an orderly system shutdown when the operating temperature exceeds the critical level.
- Fan speed monitoring provides a warning and an orderly system shutdown when the speed is out of the operational specification.
- Voltage monitoring provides a warning and an orderly system shutdown when the voltages are out of the operational specification.

4.1.10 Error handling and reporting

In the unlikely event of system hardware or environmentally induced failure, the system run-time error capture capability systematically analyzes the hardware error signature to determine the cause of failure.

- The analysis will be stored in the system NVRAM. When the system can be successfully rebooted either manually or automatically, the error will be reported to the AIX 5L or Linux operating system.
- Error Log Analysis (ELA) can be used to display the failure cause and the physical location of failing hardware.
- A hardware fault will also turn on the Attention Indicator (one is located on the front of the system unit and one is on each light strip) to alert the user of an internal hardware problem. The indicator may also be turned on by the operator as a tool to allow system identification. For identification, the indicators will flash, whereas the indicator will be on solid when an error condition occurs.

4.1.11 Availability enhancement functions

A number of availability improvements have been included in the service processor in the IBM IBM @server p5 servers. Separate copies of service processor microcode and the POWER Hypervisor code are stored in discrete Flash memory storage areas. Code access is CRC protected. Maintaining two copies insures that the Service Processor can run even if a Flash memory copy becomes corrupted, and allows for redundancy in the event of a problem during the upgrade of the firmware.

The service processor performs low-level hardware initialization and configuration of all processors. The POWER Hypervisor firmware performs higher-level configuration for features such as the virtualization support required to run up to 254 partitions concurrently on the p5-590 and p5-595 servers.

In addition, if the service processor encounters an error during runtime, it can reboot itself while the server system stays up and running. There will be no server application impact for service processor transient errors.

The system auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable software error, software hang, hardware failure, or environmentally induced (ac power) failure.

Two service processors are required in all p5-590 and p5-595 configurations.

4.2 Serviceability

The p5-590 and p5-595 servers are designed for IBM service representative setup of the machine and for subsequent addition of most features (adapters/devices).

- The p5-590 and p5-595 server service processor enables the analysis of a system that will not boot.
- The diagnostics consist of Stand-alone Diagnostics, which are loaded from the DVD-ROM drive, and Online Diagnostics.
- Online Diagnostics, when installed, are resident with AIX 5L on the disk or system. They can be booted in single-user mode (service mode), run in maintenance mode, or run concurrently (concurrent mode) with other applications. They have access to the AIX 5L Error Log and the AIX 5L Configuration Data.
  - Service mode allows checking of system devices and features.
  - Concurrent mode allows the normal system functions to continue while selected resources are being checked.
  - Maintenance mode allows checking of most system resources.
- The System Management Services (SMS) error log is accessible from the SMS menu for tests performed through SMS programs. For results of service processor tests, access the error log from the service processor menu.

4.2.1 Service Agent

Service Agent is available at no additional charge. When installed on an IBM @server system, the Service Agent can enhance IBM's ability to provide the system with maintenance service.

The Service Agent:

- Monitors and analyzes system errors and, if needed, can automatically place a service call to IBM without client intervention
- Can help reduce the effect of business disruptions due to unplanned system outages and failures
- Performs problem analysis on a subset of hardware-related problems and, with client authorization, can report automatically the results to IBM customer service.

Note: Because the p5-595 and p5-595 systems do not have integrated media bays, we have the option to order the DVD-ROM (FC 2634 or FC 1106) and DVD-RAM (FC 5752 or FC 1103) as a component of an internal media drawer (FC 5795) or IBM 7112-102 Storage Device Enclosure. Alternate methods for maintaining and servicing the system need to be available if the DVD-ROM or DVD-RAM is not ordered.

4.2.2 Online customer support

Online customer support (OCS) for hardware problem reporting may be performed via remote login by IBM @server specialists. The Electronic Service Agent™ software can also be used for this capability.

AIX 5L support offerings will be under AIXSERV and Electronic Service Agent.

> **Note:** This RAS function is not supported under Linux.

4.3 IBM @server Cluster 1600

Today's IT infrastructure requires that systems meet increasing demands, while offering the flexibility and manageability to rapidly develop and deploy new services. IBM clustering hardware and software provide the building blocks, with availability, scalability, security, and single-point-of-management control, to satisfy these needs. The advantages of a clusters are:

- Large-capacity data and transaction volumes, including support of mixed workloads
- Scale-up (add processors) or scale-out (add servers) without downtime
- Single point-of-control for distributed and clustered server management
- Simplified use of IT resources
- Designed for 24x7 access to data applications
- Business continuity in the event of disaster

IBM @server Cluster 1600 is a POWER processor-based AIX 5L and Linux cluster targeting scientific and technical computing, large-scale databases, and workload consolidation. IBM Cluster Systems Management (CSM) is designed to provide a robust, powerful, and centralized way to manage a large number of POWER5 processor-based systems all from one single point-of-control. CSM can help lower the overall cost of IT ownership by helping to simplify the tasks of installing, operating, and maintaining clusters of servers. CSM can provide one consistent interface for managing both AIX 5L and Linux nodes (physical systems or logical partitions), with capabilities for remote parallel network install, remote hardware control, and distributed command execution.

Cluster Systems Management (CSM) V1.4 for AIX 5L and Linux on POWER is supported on the p5-590 and p5-595 servers. For hardware control, a Hardware Management Console (HMC) is required. Additionally, the p5-590 and p5-595 servers are added to the hardware models supported with the pSeries cluster 1600 running CSM.

See Table 4-1 for information on cluster 1600 scalability limits.

Table 4-1 Cluster 1600 scalability limits

Server	Machine type	Maximum servers per cluster	LPARs supported	Maximum LPARs per server
p5-590	9119	16	Yes	32
p5-595	9119	16	Yes	64

Information regarding the IBM @server Cluster 1600, HMC control, cluster building block servers, and cluster software available can be found at:

http://techsupport.services.ibm.com/server/cluster/

Servicing an IBM @server p5 system

POWER5 servers may be designated:

- Customer Set-Up (CSU) with Customer Installable Features (CIF) and Customer Replaceable Units (CRU)
- Authorized service representative setup, upgraded, and maintained

A number of Web-based resources are available to assist customers and service providers wiht planning, installing, and maintaining p5 servers.

Note: This section is not specific to p5-590 and p5-595, and deals with IBM @server p5 in general.

Resource link

Resource Link™ is a customized Web-based solution, providing access to information for planning, installing, and maintaining IBM eserver p5 and associated software. It also includes similar information about other selected IBM servers. Access to the site is by IBM registration ID and password, which are available free of charge. Resource Link screens can vary by user authorization level, and are continually being updated; the detail that you see when accessing Resource Link may not exactly match that mentioned here.

Resource link contains links to:

- Education
  - Resource Link highlights
  - eserver HW Info Center education
  - Customer Course for Servicing the IBM eserver i5 and p5
- Planning
- Forums
- Fixes

Resource Link is available at:

https://www-1.ibm.com/servers/resourcelink

IBM eServer Hardware Information Center

The IBM eserver Hardware Information Center is a source for both hardware and software technical information for eserver p5 systems. It has information to help perform a variety of tasks, including:

- Preparing a site to accommodate eserver hardware.
- Installing the server, console, features, and options, and other hardware.
- Installing and using a Hardware Management Console.
- Partitioning the server and installing the operating systems.
- Enabling and managing capacity on demand.
- Troubleshooting problems and servicing the server. Included here are component removal and replacement procedures, as well as the Start of Call procedure.
  - Physical components of a system are generally considered either a Customer Replaceable Unit (CRU) or a Field Replaceable Unit (FRU). CRUs are further categorized as either Tier 1 CRUs or Tier 2 CRUs. Definitions are as follows:
    - Tier 1 CRU - Very easy to replace
    - Tier 2 CRU - More complicated to replace
    - FRU - Replaced by the service provider

Removal and replacement procedures may be documented in the information center accompanied by graphics, such as in Figure 1 on page 87, and video clips.

Alternatively, they may take the form of guided procedures using the HMC: **Service Applications** → **Service Focal Point** → **Exchange Parts**.

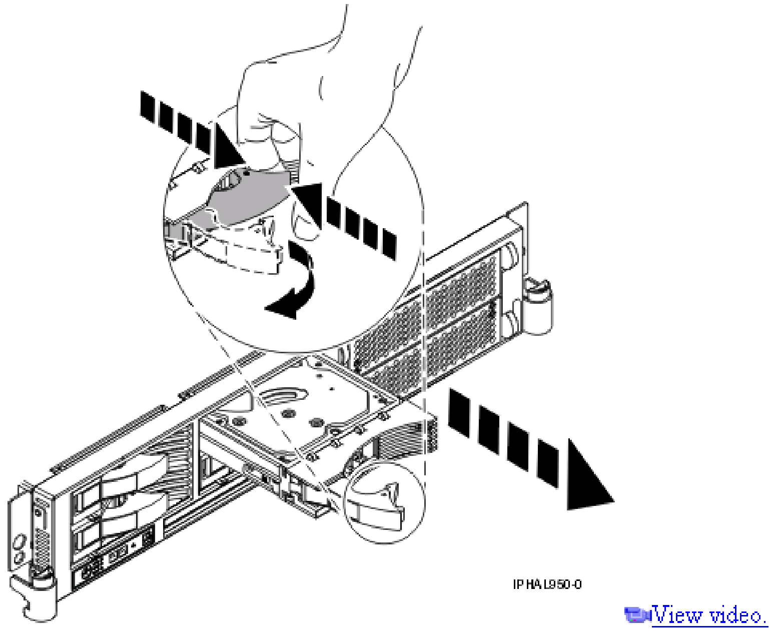


Figure 1 Removing a disk drive

> **Note:** Part classification, contractual agreements, and implementation in specific geographies all affect how CRUs/FRUs are determined.

IBM @server Hardware Information Center is available:

- On the Internet

  `http://www.ibm.com/servers/library/infocenter`

- On the HMC
  - Click **Information Center** and **Setup Wizard** → **Launch the Information Center**.
- On CD-ROM
  - Shipped with the hardware (English: SK3T-8159)
  - Also available to order from IBM Publications Center

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this Redpaper.

IBM Redbooks

For information on ordering these publications, see “How to get IBM Redbooks” on page 91. Note that some of the documents referenced here may be available in softcopy only.

- *Advance POWER Virtualization on IBM ⓔserver p5 Servers,* SG24-7940
- *IBM ⓔserver p5 510 Technical Overview and Introduction*, REDP-4001
- *IBM ⓔserver p5 520 Technical Overview and Introduction*, REDP-9111
- *IBM ⓔserver p5 550 Technical Overview and Introduction*, REDP-9113
- *IBM ⓔserver p5 570 Technical Overview and Introduction*, REDP-9117
- *IBM ⓔserver p5 590 and 595 System Handbook*, SG24-9119
- *Managing AIX Server Farms*, SG24-6606
- *Partitioning Implementations for IBM ⓔserver p5 Servers*, SG24-7039
- *Practical Guide for SAN with pSeries*, SG24-6050
- *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496
- *Understanding IBM ⓔserver pSeries Performance and Sizing*, SG24-4810

Other publications

These publications are also relevant as further information sources:

- *7014 Series Model T00 and T42 Rack Installation and Service Guide*, SA38-0577, contains information regarding the 7014 Model T00 and T42 Rack, in which this server can be installed.
- *7316-TF3 17-Inch Flat Panel Rack-Mounted Monitor and Keyboard Installation and Maintenance Guide*, SA38-0643, contains information regarding the 7316-TF3 Flat Panel Display, which can be installed in your rack to manage your system units.
- *RS/6000 and ⓔserver pSeries Adapters, Devices, and Cable Information for Multiple Bus Systems*, SA38-0516, contains information about adapters, devices, and cables for your system. This manual is intended to supplement the service information found in the *Diagnostic Information for Multiple Bus Systems* documentation.
- *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538, contains information regarding slot restrictions for adapters that can be used in this system.
- *System Unit Safety Information*, SA23-2652, contains translations of safety information used throughout the system documentation.
- *IBM ⓔserver Planning*, SA38-0508, contains site and planning information, including power and environment specification.

Online resources

These Web sites and URLs are also relevant as further information sources:

- AIX 5L operating system maintenance package downloads

  http://www.ibm.com/servers/eserver/support/pseries/aixfixes.html
- IBM @server p5, pSeries, OpenPower and IBM RS/6000 Performance Report

  http://www.ibm.com/servers/eserver/pseries/hardware/system_perf.html
- IBM TotalStorage Expandable Storage Plus

  http://www.ibm.com/servers/storage/disk/expplus/index.html
- IBM TotalStorage Mid-range Disk Systems

  http://www.ibm.com/servers/storage/disk/ds4000/index.html
- IBM TotalStorage Enterprise disk storage

  http://www.ibm.com/servers/storage/disk/enterprise/ds_family.html
- IBM Virtualization Engine

  http://www.ibm.com/servers/eserver/about/virtualization/
- Advanced POWER Virtualization on IBM @server p5

  http://www.ibm.com/servers/eserver/pseries/ondemand/ve/resources.html
- Virtual I/O Server supported environments

  http://techsupport.services.ibm.com/server/virtualization/vios/documentation/datasheet.html
- Hardware Management Console support information

  https://techsupport.services.ibm.com/server/hmc/power5
- The LVT is a PC based tool intended assist you in logical partitioning

  http://www.ibm.com/servers/eserver/iseries/lpar/systemdesign.htm
- Customer Specified Placement and LPAR delivery

  http://www.ibm.com/servers/eserver/power/csp/index.html
- SUMA on AIX 5L

  http://techsupport.services.ibm.com/server/fixget
- Linux on IBM @server p5 and pSeries

  http://www.ibm.com/servers/eserver/pseries/linux/
- SUSE LINUX Enterprise Server 9

  http://www.novell.com/products/linuxenterpriseserver/
- Red Hat Enterprise Linux details

  http://www.redhat.com/software/rhel/details/
- IBM @server Linux on POWER overview

  http://www.ibm.com/servers/eserver/linux/power/whitepapers/linux_overview.html
- IBM @server Cluster 1600

  http://www.ibm.com/servers/eserver/clusters/hardware/1600.html
- Autonomic computing on IBM @server pSeries servers

  http://www.ibm.com/autonomic/index.shtml

- Copper circuitry

  http://www.ibm.com/chips/technology/technologies/copper/

- IBM @server p5 AIX 5L Support for Micro-Partitioning technology and Simultaneous Multithreading White Paper

  http://www.ibm.com/servers/aix/whitepapers/aix_support.pdf

- Hardware documentation

  http://publib16.boulder.ibm.com/pseries/en_US/infocenter/base/

- IBM @server Information Center

  http://publib.boulder.ibm.com/eserver/

- IBM @server pSeries support

  http://www.ibm.com/servers/eserver/support/pseries/index.html

- IBM @server support: Tips for AIX administrators

  http://techsupport.services.ibm.com/server/aix.srchBroker

- IBM Linux news: Subscribe to the Linux Line

  https://www.software.ibm.com/reg/linux/linuxline-i

- IBM online sales manual

  http://www.ibmlink.ibm.com

- Linux for IBM @server pSeries

  http://www.ibm.com/servers/eserver/pseries/linux/

- Microcode Discovery Service

  http://techsupport.services.ibm.com/server/aix.invscoutMDS

- POWER4 system micro architecture, comprehensively described in the IBM Journal of Research and Development, Vol 46 No.1 January 2002

  http://www.research.ibm.com/journal/rd46-1.html

- SCSI T10 Technical Committee

  http://www.t10.org

- Silicon-on-insulator (SOI) technology

  http://www.ibm.com/chips/technology/technologies/soi/

- Microcode Downloads for IBM @server i5, OpenPower, p5, pSeries, and RS/6000 Systems

  http://techsupport.services.ibm.com/server/mdownload

- Additional information about Capacity on Demand

  http://www.ibm.com/servers/eserver/pseries/ondemand/cod/

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM ⓔserver p5 590 and 595
Technical Overview and Introduction

Redpaper

Finer system granularity using Micro-Partitioning technology to help lower TCO

Support for versions of AIX 5L, Linux, and i5/OS operating systems

Enterprise class features for applications that require a robust environment

This IBM Redpaper is a comprehensive guide covering the IBM ⓔserver p5 590 and p5 595 AIX 5L and Linux operating system servers. We introduce major hardware offerings and discuss their prominent functions.

Professionals wishing to acquire a better understanding of IBM ⓔserver p5 products should consider reading this document. The intended audience includes:

-Clients
-Sales and marketing professionals
-Technical support professionals
-IBM Business Partners
-Independent software vendors

This document expands the current set of IBM ⓔserver documentation by providing a desktop reference that offers a detailed technical description of the p5-590 and p5-595 servers.

This publication does not replace the latest IBM ⓔserver marketing materials and tools. It is intended as an additional source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.