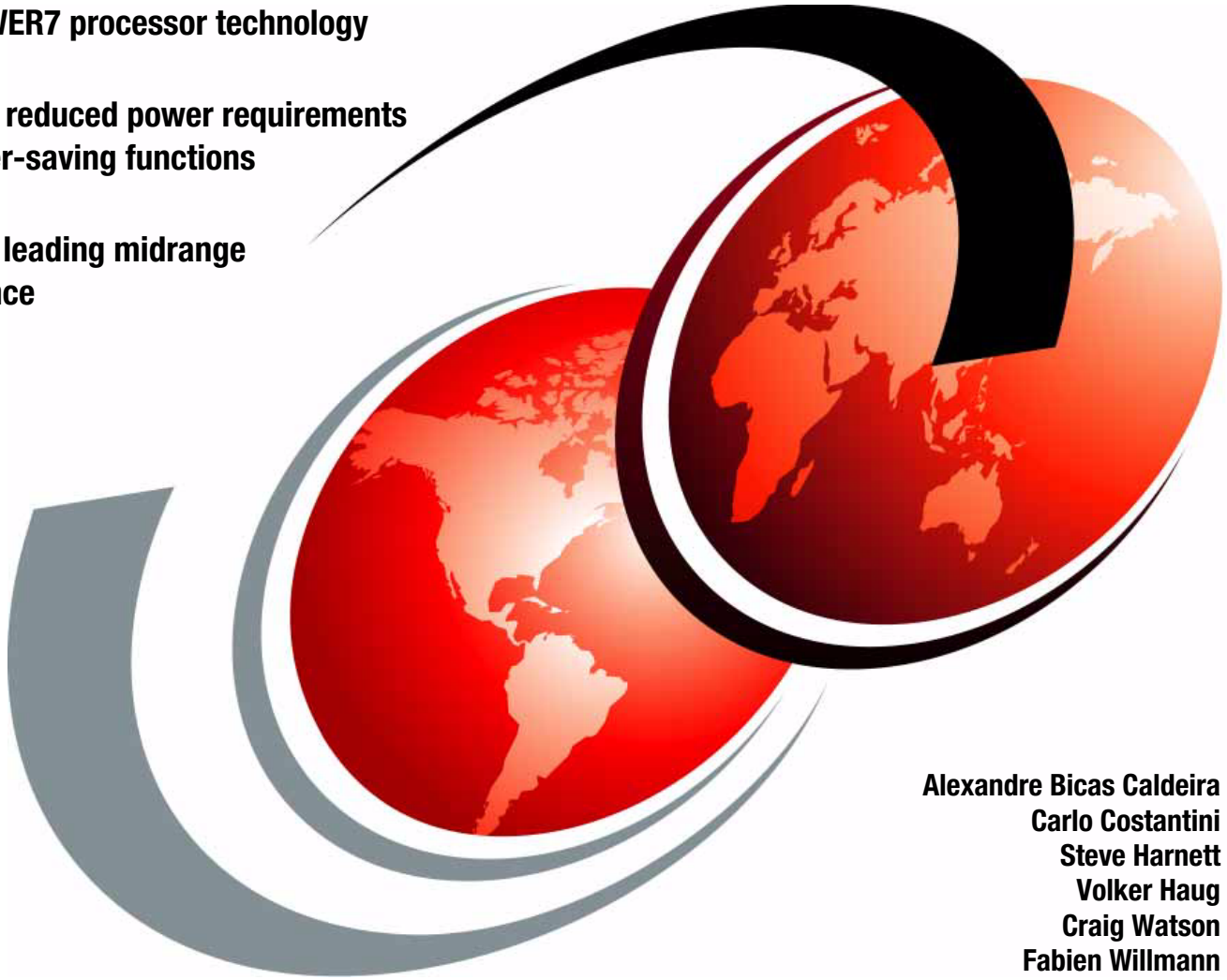
IBM

IBM Power 720 and 740 Technical Overview and Introduction

- Features the 8202-E4C and 8205-E6C based on the latest POWER7 processor technology
- Discusses reduced power requirements with power-saving functions
- Describes leading midrange performance



Alexandre Bicas Caldeira
Carlo Costantini
Steve Harnett
Volker Haug
Craig Watson
Fabien Willmann

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM Power 720 and 740 Technical Overview and Introduction

November 2011

REDP-4797-00

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

First Edition (November 2011)

This edition applies to the IBM Power 720 (8202-E4C) and Power 740 (8205-E6C) Power Systems servers.

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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites 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.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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 illustrate 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.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

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

Active Memory™
AIX®
Electronic Service Agent™
EnergyScale™
Focal Point™
IBM Systems Director Active Energy Manager™
IBM®
Micro-Partitioning®
POWER Hypervisor™
Power Systems™
POWER6+™
POWER6®
POWER7®
PowerHA®
PowerVM®
Power®
POWER®
pSeries®
Redbooks®
Redpaper™
Redbooks (logo) ®
System Storage®
System x®
System z®
Tivoli®

The following terms are trademarks of other companies:

Intel Xeon, Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

LTO, Ultrium, the LTO Logo and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries 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, or service names may be trademarks or service marks of others.

Preface

This IBM® Redpaper™ publication is a comprehensive guide covering the IBM Power® 720 and Power 740 servers supporting AIX®, IBM i, and Linux operating systems. The goal of this paper is to introduce the innovative Power 720 and Power 740 offerings and their major functions, including these:

- The IBM POWER7® processor available at frequencies of 3.0 GHz, 3.55 GHz, and 3.7 GHz.
- The specialized POWER7 Level 3 cache that provides greater bandwidth, capacity, and reliability.
- The 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter included in the base configuration and installed in a PCIe Gen2 x4 slot.
- The integrated SAS/SATA controller for HDD, SSD, tape, and DVD. This controller supports built-in hardware RAID 0, 1, and 10.
- The latest IBM PowerVM® virtualization, including PowerVM Live Partition Mobility and PowerVM IBM Active Memory™ Sharing.
- Active Memory Expansion technology that provides more usable memory than is physically installed in the system.
- IBM EnergyScale™ technology that provides features such as power trending, power-saving, capping of power, and thermal measurement

Professionals who want to acquire a better understanding of IBM Power Systems™ products can benefit from reading this Redpaper publication. The intended audience includes these roles:

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

This paper complements the available set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power 720 and Power 740 systems.

This paper does not replace the latest marketing materials and configuration 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 who wrote this paper

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

Alexandre Bicas Caldeira works on the Power Systems Field Technical Sales Support team for IBM Brazil. He holds a degree in computer science from the Universidade Estadual Paulista (UNESP). Alexandre has more than 11 years of experience working on IBM and Business Partner on Power Systems hardware, AIX, and PowerVM virtualization products.

He is also skilled on IBM System Storage®, IBM Tivoli® Storage Manager, IBM System x®, and VMware.

Carlo Costantini is a Certified IT Specialist for IBM and has over 33 years of experience with IBM and IBM Business Partners. He currently works in Italy Power Systems Platforms as Presales Field Technical Sales Support for IBM Sales Representatives and IBM Business Partners. Carlo has broad marketing experience, and his current major areas of focus are competition, sales, and technical sales support. He is a Certified Specialist for Power Systems servers. He holds a master's degree in electronic engineering from Rome University.

Steve Harnett is a Senior Accredited Professional, Chartered IT Professional, and member of the British Computing Society. He currently works as a pre-sales Technical Consultant in the IBM Server and Technology Group in the UK. Steve has over 16 years of experience working in post sales supporting Power Systems. He is a product Topgun and a recognized SME in Electronic Service Agent™, Hardware Management Console, and High end Power Systems. He also has several years of experience in developing and delivering education to clients, IBM Business Partners, and IBMers.

Volker Haug is a certified Consulting IT Specialist within IBM Systems and Technology Group, based in Ehningen, Germany. He holds a bachelor's degree in business management from the University of Applied Studies in Stuttgart. His career has included more than 24 years working in the IBM PLM and Power Systems divisions as a RISC and AIX Systems Engineer. Volker is an expert in Power Systems hardware, AIX, and PowerVM virtualization. He is POWER7 Champion and also a member of the German Technical Expert Council, a affiliate of the IBM Academy of Technology. He has written several books and white papers about AIX, workstations, servers, and PowerVM virtualization.

Craig Watson has 15 years of experience working with UNIX-based systems in roles including field support, systems administration, and technical sales. He has worked in the IBM Systems and Technology Group since 2003. Craig is currently working as a Systems Architect, designing complex solutions for customers that include Power Systems, System x, and Systems Storage. He holds a master's degree in electrical and electronic engineering from the University of Auckland.

Fabien Willmann is an IT Specialist working with Techline Power Europe in France. He has 10 years of experience with Power Systems, AIX, and PowerVM virtualization. After teaching hardware courses on Power Systems servers, he joined ITS as an AIX consultant where he developed his compentencies in AIX, HMC management, and PowerVM virtualization. Building new Power Systems configurations for STG pre-sales is his major area of expertise today. Recently he gave a workshop on the econfig configuration tool, focused on POWER7 processor-based BladeCenters during the symposium for French Business Partners in Montpellier.

The project that produced this publication was managed by:

Scott Vetter, IBM Certified Project Manager and PMP.

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

Larry Amy, Gary Anderson, Sue Beck, Terry Brennan, Pat Buckland, Paul D. Carey, Pete Heyrman, John Hilburn, Dan Hurlimann, Kevin Kehne, James Keniston, Jay Kruemcke, Robert Lowden, Hilary Melville, Thoi Nguyen, Denis C. Nizinski, Pat O'Rourke, Jan Palmer, Ed Prosser, Robb Romans, Audrey Romonosky, Todd Rosedahl, Melanie Steckham, Ken Trusits, Al Yanes
IBM U.S.A.

Stephen Lutz
IBM Germany

Tamikia Barrow
International Technical Support Organization, Poughkeepsie Center

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

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 paper or other IBM Redbooks® publications in one of the following ways:

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

  ibm.com/redbooks
- Send your comments in an email to:

  redbooks@us.ibm.com
- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:

  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

General description

The IBM Power 720 (8202-E4C) and IBM Power 740 (8205-E6C) servers utilize the latest POWER7 processor technology designed to deliver unprecedented performance, scalability, reliability, and manageability for demanding commercial workloads. The new Power 720 and Power 740 servers provide enhancements that can be particularly beneficial to customers running applications driving very high I/O or memory requirements.

The performance, availability, and flexibility of the Power 720 server can enable companies to spend more time running their business utilizing a proven solution from thousands of ISVs that support the AIX, IBM i, and Linux operating systems. The Power 720 server is a high-performance, energy efficient, reliable, and secure infrastructure and application server in a dense form factor. As a high-performance infrastructure or application server, the Power 720 contains innovative workload-optimizing technologies that maximize performance based on client computing needs and Intelligent Energy features that help maximize performance and optimize energy efficiency, resulting in one of the most cost-efficient solutions for UNIX, IBM i, and Linux deployments.

As a distributed application server, the IBM Power 720 is designed with capabilities to deliver leading-edge application availability and enable more work to be processed with less operational disruption for branch office and in-store applications. As a consolidation server, PowerVM Editions provide the flexibility to use leading-edge AIX, IBM i, Linux applications and offer comprehensive virtualization technologies designed to aggregate and manage resources while helping to simplify and optimize your IT infrastructure and deliver one of the most cost-efficient solutions for UNIX, IBM i, and Linux deployments.

The Power 740 offers the performance, capacity, and configuration flexibility to meet the most demanding growth requirements, and combined with industrial-strength PowerVM virtualization for AIX, IBM i, and Linux, it can fully utilize the capability of the system. These capabilities are designed to satisfy even the most demanding processing environments and can deliver business advantages and higher client satisfaction.

The Power 740 is designed with innovative workload-optimizing and energy management technologies to help clients get the most out of their systems (that is, running applications rapidly and energy efficiently to conserve energy and reduce infrastructure costs). It is fueled by the outstanding performance of the POWER7 processor, making it possible for applications to run faster with fewer processors, resulting in lower per-core software licensing costs.

1.1 Systems overview

You can find detailed information about the Power 720 and Power 740 systems within the following sections.

1.1.1 The Power 720 server

The Power 720 offers a choice of a 4-core, 6-core, or 8-core configuration, available in a 4U rack-mount or a tower form factor. The POWER7 processor chips in this server are 64-bit, 4-core, 6-core, and 8-core modules with 4 MB of L3 cache per core and 256 KB of L2 cache per core.

The Power 720 server supports a maximum of 16 DDR3 DIMM slots, with eight DIMM slots included in the base configuration and eight DIMM slots available with an optional memory riser card. A system with the optional memory riser card installed has a maximum memory of 256 GB.

The Power 720 system comes with an integrated SAS controller, offering RAID 0, 1, and 10 support, and two storage backplanes are available. The base configuration supports up to six SFF SAS HDDs/SSDs, an SATA DVD, and a half-high tape drive. A higher-function backplane is available as an option. This supports up to eight SFF SAS HDDs/SSDs, an SATA DVD, a half-high tape drive, Dual 175 MB Write Cache RAID with RAID 5 and 6 support, and an external SAS port.

All HDDs/SSDs are hot-swap and front accessible. If the internal storage capacity is not sufficient, there are also four disk-only I/O drawers supported, providing large storage capacity and multiple partition support.

The Power 720 comes with five PCI Express (PCIe) Gen2 full-height profile slots for installing adapters in the system. Optionally, an additional riser card with four PCIe Gen2 low-profile (LP) slots can be installed in a GX++ slot available on the backplane. This extends the number of slots to nine. The system also comes with a PCIe x4 Gen2 slot containing a 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter.

If additional slots are required, the Power 720 supports external I/O drawers in place of the riser card, allowing for a maximum of two feature code #5802/#5877 PCIe drawers or four #5796 PCI-X drawers. This increases the number of available slots to 20 PCIe slots or 24 PCI-X slots. Note that only the 6-core and 8-core systems support external I/O slots.

Note: The Integrated Virtual Ethernet (IVE) adapter is not available for the Power 720.

The Power 720 also implements Light Path diagnostics, which provides an obvious and intuitive means to positively identify failing components. Light Path diagnostics allow system engineers and administrators to easily and quickly diagnose hardware problems.

There is an upgrade available from a POWER6® processor-based IBM Power 520 server (8203-E4A) into the Power 720 (8202-E4C). A Power 520 (9408-M25) can be converted to an Power 520 (8203-E4A) and then be upgraded to a Power 720 (8202-E4C). You can also directly upgrade from an Power 520 (8203-E4A) to the Power 720 (8202-E4C), preserving the existing serial number.

The Power 720 system's Capacity Backup (CBU) designation can help meet your requirements for a second system to use for backup, high availability, and disaster recovery. It enables you to temporarily transfer IBM i processor license entitlements and IBM i

user license entitlements purchased for a primary machine to a secondary CBU-designated system. Temporarily transferring these resources instead of purchasing them for your secondary system might result in significant savings. Processor activations cannot be transferred.

Note: The Power 720 Capacity Backup capability is available for IBM i. For information about registration and other topics, visit:

`http://www.ibm.com/systems/power/hardware/cbu`

Figure 1-1 shows the Power 720 rack and tower models.

Figure 1-1 Power 720 rack and tower models

1.1.2 The Power 740 server

The IBM Power 740 server is a 4U rack-mount with two processor sockets offering 4-core 3.3 GHz and 3.7 GHz, 6-core 3.7 GHz, and 8-core 3.55 GHz processor options. The POWER7 processor chips in this server are 64-bit, 4-core, 6-core, and 8-core modules with 4 MB of L3 cache per core and 256 KB of L2 cache per core.

The Power 740 server supports a maximum of 32 DDR3 DIMM slots, with eight DIMM slots included in the base configuration and 24 DIMM slots available with three optional memory riser cards. A system with three optional memory riser cards installed has a maximum memory of 512 GB.

The Power 740 system comes with an integrated SAS controller, offering RAID 0, 1, and 10 support, and two storage backplanes are available. The base configuration supports up to six SFF SAS HDDs/SSDs, an SATA DVD, and a half-high tape drive. A higher-function backplane is available as an option. This supports up to eight SFF SAS HDDs/SSDs, an SATA DVD, a half-high tape drive, Dual 175 MB Write Cache RAID with RAID 5 and 6 support, and an external SAS port.

All HDDs/SSDs are hot-swap and front accessible. If the internal storage capacity is not sufficient, there are also four disk-only I/O drawers supported, providing large storage capacity and multiple partition support.

The Power 740 comes with five PCI Express (PCIe) Gen2 full-height profile slots for installing adapters in the system. Optionally, an additional riser card with four PCIe Gen2 low-profile slots can be installed in a GX++ slot available on the backplane. This extends the number of slots to nine. The system also comes with a PCIe x4 Gen2 slot containing a 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter.

If additional slots are required, the Power 740 supports external I/O drawers, allowing for a maximum of four feature code #5802/#5877 PCIe drawers or four #5796 PCI-X drawers. This makes the server capable of 20 PCIe slots or 24 PCI-X slots.

Note: The Integrated Virtual Ethernet (IVE) adapter is not available for the Power 740.

The Power 740 also implements Light Path diagnostics, which provides an obvious and intuitive means to positively identify failing components. Light Path diagnostics allows system engineers and administrators to easily and quickly diagnose hardware problems.

Note: The Power 740 Capacity Backup capability is available for IBM i. For information about registration and other topics, visit:

`http://www.ibm.com/systems/power/hardware/cbu`

Figure 1-2 shows the Power 740 rack model.

Figure 1-2 Power 740 rack model

1.2 Operating environment

Table 1-1 lists the operating environment specifications for the servers.

Table 1-1 Operating environment for Power 720 and Power 740

Power 720 and Power 740 operating environment				
Description	**Operating**		**Non-operating**	
	Power 720	**Power 740**	**Power 720**	**Power 740**
Temperature	5 - 35 degrees C (41 - 95 degrees F) Recommended: 18 - 27 degrees C (64 - 80 degrees F)		5 - 45 degrees C (41 to 113 degrees F)	
Relative humidity	8 - 80%		8 - 80%	
Maximum dew point	28 degrees C (84 degrees F)		28 degrees C (82 degrees F)	
Operating voltage	100 - 127 V ac or 200 - 240 V ac	200 - 240 V ac	N/A	
Operating frequency	47 - 63 Hz		N/A	
Power consumption	840 watts maximum	1400 watts maximum	N/A	
Power source loading	0.857 kVa maximum	1.428 kVa maximum	N/A	
Thermal output	2867 Btu/hour maximum	4778 Btu/hour maximum	N/A	
Maximum altitude	3050 m (10,000 ft)		N/A	
Noise level reference point:	Tower system: 5.6 bels (operating) 5.5 bels (idle) Rack system: 5.6 bels (operating) 5.5 bels (idle)	Rack system: 6.0 bels (operating 5.9 bels (idle)	N/A	

Note: The maximum measured value is expected from a fully populated server under an intensive workload. The maximum measured value also accounts for component tolerance and non ideal operating conditions. Power consumption and heat load vary greatly by server configuration and utilization. Use the IBM Systems Energy Estimator to obtain a heat output estimate based on a specific configuration:

http://www-912.ibm.com/see/EnergyEstimator

1.3 Physical package

The Power 720 is available in both rack-mount and tower form factors. The Power 740 is available in rack-mounted form factor only. The major physical attributes for each are discussed in the following sections.

1.3.1 Tower model

The Power 720 can be configured as tower models by selecting the features shown in Table 1-2.

Table 1-2 Features for selecting tower models

Cover set	Power 720 (8202-E4C)
IBM Tower Cover Set	#7567
OEM Tower Cover Set	#7568

Table 1-3 shows the physical dimensions of the tower models.

Table 1-3 Physical dimensions of the Power 720 tower chassis

Dimension	Power 720 (8202-E4C)
Width without tip plate	183 mm (7.2 in)
Width with tip plate	328.5 mm (12.9 in)
Depth	688 mm (27.1 in)
Height	541 mm (21.3 in)
Weight without tip plate	53.7 kg (118.1 lb)
Weight with tip plate	57.2 kg (125.8 lb)

1.3.2 Rack-mount model

The Power 720 and Power 740 can be configured as 4U (4EIA) rack-mount models by selecting the features shown in Table 1-4.

Table 1-4 Features for selecting rack-mount models

Cover set	Power 720 (8202-E4C)	Power 740 (8205-E6C)
IBM Rack-mount Drawer Bezel and Hardware	#7134	#7131
OEM Rack-mount Drawer Bezel and Hardware	#7135	#7132

Table 1-5 shows the physical dimensions of the rack-mount models.

Table 1-5 Physical dimensions of the Power 720 and Power 740 rack-mount chassis

Dimension	Power 720 (8202-E4C)	Power 740 (8205-E6C)
Width	440 mm (17.3 in)	440 mm (17.3 in)
Depth	610 mm (24.0 in)	610 mm (24.0 in)

Dimension	Power 720 (8202-E4C)	Power 740 (8205-E6C)
Height	173 mm (6.81 in)	173 mm (6.81 in)
Weight	48.7 kg (107.4 lb)	48.7 kg (107.4 lb)

Figure 1-3 shows the rear view of a Power 740 with the optional PCIe expansion.

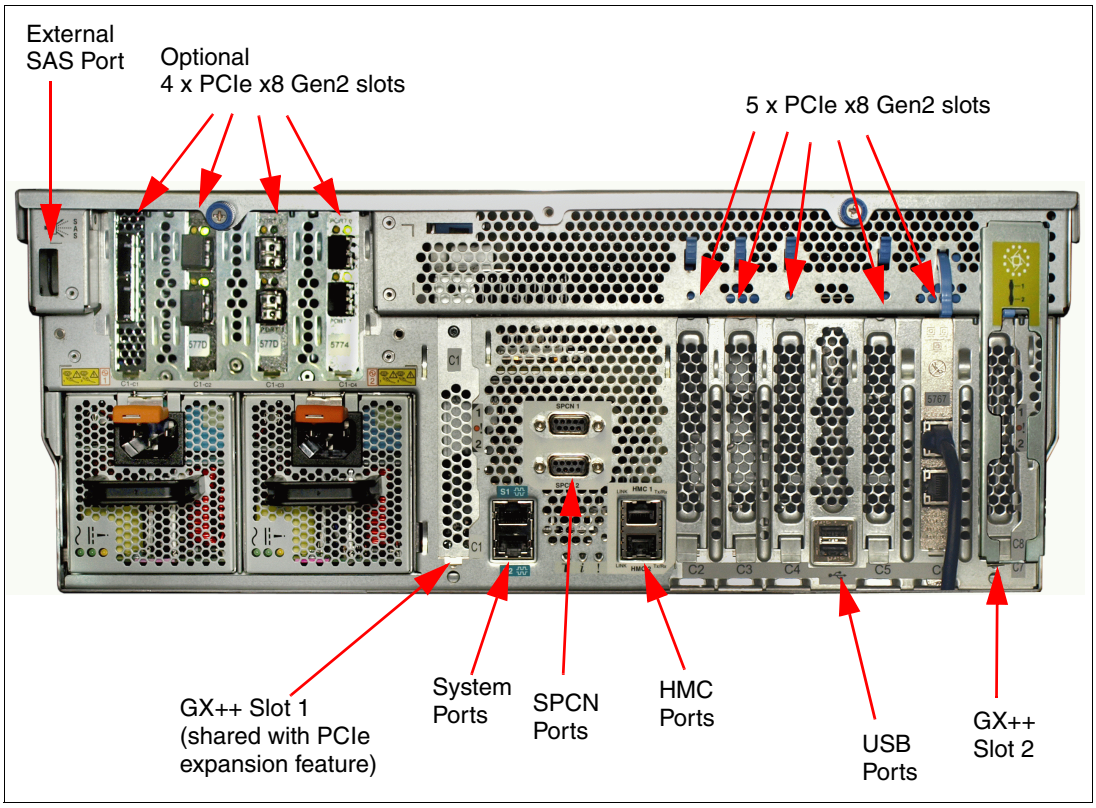


Figure 1-3 Rear view of a rack-mount Power 740 server

1.4 System features

The system chassis contains one processor module (Power 720) or up to two processor modules (Power 740). Each POWER7 processor module has either 4-cores, 6-cores, or 8-cores. Each of the POWER7 processor chips in the server has a 64-bit architecture, up to 2 MB of L2 cache (256 KB per core) and up to 32 MB of L3 cache (4 MB per core).

1.4.1 Power 720 system features

This summary describes the standard features:

- Tower or rack-mount (4U) chassis
- Four-core, 6-core, or 8-core configuration with one 3.0 GHz processor module
- Up to 256 GB of 1066 MHz DDR3 ECC memory
- An integrated SAS controller, offering RAID 0, 1, and 10 support

- Choice of two disk/media backplanes:
  - Six 2.5-inch HDD/SSD/Media backplane with one tape drive bay and one DVD bay
  - Eight 2.5-inch HDD/SSD/Media backplane with one tape drive bay, one DVD bay, Dual 175 MB Write Cache RAID with RAID 5 and 6 support, and one external SAS port
- A PCIe x4 Gen2 slot containing 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter
- A maximum of nine PCIe Gen2 slots:
  - Five PCIe x8 full-height short card slots
  - Optional four PCIe x8 low-profile short card slots
- One GX++ slot
- Integrated:
  - Service Processor
  - EnergyScale technology
  - Hot-swap and redundant cooling
  - Three USB ports and two system ports
  - Two HMC ports and two SPCN ports
- Optional redundant, 1725 Watt ac hot-swap power supplies

1.4.2 Power 740 system features

This summary describes the standard features:

- Tower (4U) chassis
- Processors:
  - Four-core or 8-core configuration with one or two 3.3 GHz or 3.7 GHz 4-core processor modules
  - Six-core or 12-core configuration with one or two 3.7 GHz 6-core processor modules
  - Eight or 16-core configuration with two 3.55 GHz 8-core processor modules
- Up to 512 GB of 1066 MHz DDR3 ECC memory
- An integrated SAS controller, offering RAID 0, 1, and 10 support
- Choice of two disk/media backplanes:
  - Six 2.5-inch HDD/SSD/Media backplanes with one tape drive bay and one DVD bay
  - Eight 2.5-inch HDD/SSD/Media backplanes with one tape drive bay, one DVD bay, Dual 175 MB Write Cache RAID with RAID 5 and 6 support, and one external SAS port
- A PCIe x4 Gen2 slot containing a 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter
- A maximum of nine PCIe Gen2 slots:
  - Five PCIe x8 full-height short card slots
  - Optional four PCIe x8 low-profile short card slots
- Two GX++ slots

- Integrated:
  - Service Processor
  - EnergyScale technology
  - Hot-swap and redundant cooling
  - Three USB ports and two system ports
  - Two HMC ports and two SPCN ports
- Redundant, 1725 watt ac hot-swap power supplies

1.4.3 Minimum features

Each system has a minimum feature set in order to be valid.

The minimum initial order must include a processor, processor activations, memory, a power supply, a power cord (two power supplies and two power cords for the Power 740), one HDD/SSD, a storage backplane, an operating system indicator, and a Language Group Specify.

If IBM i is the Primary Operating System (#2145), the initial order must also include one additional HDD/SSD, a Mirrored System Disk Level Specify Code, and a System Console on HMC Indicator. A DVD is defaulted on every order but can be de-selected. A DVD-ROM or DVD-RAM must be accessible by the system.

Note: If AIX, IBM i, or Linux is the primary operating system, no internal HDD or SSD is required if feature #0837 (Boot from SAN) is selected. A Fibre Channel or FCoE adapter must be ordered if #0837 is selected.

1.4.4 Power supply features

One system 1725 watt ac power supply (#5603) is required for the Power 720. A second power supply is optional. Two system 1725 watt ac power supplies (#5603) are required for the Power 740. The second power supply provides redundant power for enhanced system availability. To provide full redundancy, the two power supplies must be connected to separate PDUs.

The server will continue to function with one working power supply. A failed power supply can be hot swapped but must remain in the system until the replacement power supply is available for exchange.

1.4.5 Processor module features

Each processor module in the system houses a single POWER7 processor chip. The processor has either 4 cores, 6 cores, or 8 cores. The Power 720 supports one processor module. The Power 740 supports a second processor module that must be identical to the first.

The amount of processor activation code features must be equal to the amount of installed processor cores.

Table 1-6 summarizes the processor features available for the Power 720.

Table 1-6 Processor features for the Power 720

Feature code	Processor module description
#EPC5	4-core 3.0 GHz POWER7 processor module
#EPC6	6-core 3.0 GHz POWER7 processor module
#EPC7	8-core 3.0 GHz POWER7 processor module

The Power 740 requires that one or two processor modules be installed. If two processor modules are installed, they have to be identical. Table 1-7 lists the available processor features.

Table 1-7 Processor features for the Power 740

Feature code	Processor module description	Min/max modules
#EPC9	4-core 3.3 GHz POWER7 processor module	1/2
#EPC8	4-core 3.7 GHz POWER7 processor module	1/2
#EPCA	6-core 3.7 GHz POWER7 processor module	1/2
#EPCB	8-core 3.55 GHz POWER7 processor module	1/2

1.4.6 Memory features

In POWER7 processor-based systems, DDR3 memory is used throughout. The POWER7 DDR3 memory uses a new memory architecture to provide greater bandwidth and capacity. This enables operating at a higher data rate for larger memory configurations.

Memory in the Power 720 and 740 systems is installed into memory riser cards. One memory riser card is included in the base system. The base memory riser card does not appear as a feature code in the configurator. Additional memory riser cards, feature #EM01, can be installed up to a maximum of two per processor module. Each memory riser card provides eight DDR3 DIMM slots. DIMMs are available in capacities of 2 GB, 4 GB, 8 GB, and 16 GB at 1066 MHz and are installed in pairs.

Table 1-8 shows memory features available on the systems.

Table 1-8 Summary of memory features

Feature code	Feature capacity	Access rate	DIMMs
#EM04	4 GB	1066 MHz	2 x 2 GB DIMMs
#EM08	8 GB	1066 MHz	2 x 4 GB DIMMs
#EM16[a]	16 GB	1066 MHz	2 x 8 GB DIMMs
#EM32[a]	32 GB	1066 MHz	2 x 16 GB DIMMs

a. A Power 720 system with 4-core processor module feature #EPC5 cannot be ordered with the 16 GB memory feature #EM16 or 32 GB memory feature #EM32.

It is generally best that memory be installed evenly across all memory riser cards in the system. Balancing memory across the installed memory riser cards allows memory access in a consistent manner and typically results in the best possible performance for your

configuration. However, balancing memory fairly evenly across multiple memory riser cards, compared to balancing memory exactly evenly typically has a very small performance difference.

1.5 Disk and media features

The Power 720 and 740 systems feature an integrated SAS controller, offering RAID 0, 1, and 10 support with two storage backplane options:

- The first option (#5618) supports up to six SFF SAS HDDs/SSDs, a SATA DVD, and a half-high tape drive for either a tape drive or USB removable disk. This feature does not provide RAID 5, RAID 6, a write cache, or an external SAS port. Split backplane functionality (3x3) is supported with the additional feature #EJ02.

  Remember: Note this information:

  - No additional PCIe SAS adapter is required for Split Backplane functionality.
  - Feature #5618 is not supported with IBM i.

- The second option (#EJ01) is a higher-function backplane that supports up to eight SFF SAS HDDs/SSDs, a SATA DVD, a half-high tape drive for either a tape drive or USB removable disk, Dual 175 MB Write Cache RAID, and one external SAS port. The #EJ01 supports RAID 5 and RAID 6 and there is no split backplane available for this feature.

All HDDs/SSDs are hot-swap and front accessible.

Table 1-9 shows the available storage configurations available for the Power 720 and Power 740.

Table 1-9 Available storage configurations for Power 720 and Power 740

Feature code	Split backplane	JBOD	RAID 0, 1, and 10	RAID 5 and 6	External SAS port
#5618	No	Yes	Yes	No	No
#5618 and #EJ02	Yes	Yes	Yes	No	No
#EJ01	No	No	Yes	Yes	Yes

Table 1-10 shows the available disk drive feature codes available for the installation a Power 720 and Power 740 server.

Table 1-10 Disk drive feature code description

Feature code	Description	OS support
#1890	69 GB SFF SAS SSD	AIX, Linux
#1886	146 GB 15 K RPM SFF SAS Disk Drive	AIX, Linux
#1917	146 GB 15 K RPM SAS SFF-2 Disk Drive	AIX, Linux
#1775	177 GB SFF-1 SSD with eMLC	AIX, Linux
#1793	177 GB SFF-2 SSD with eMLC	AIX, Linux
#1995	177 GB SSD Module with eMLC	AIX, Linux
#1925	300 GB 10 K RPM SAS SFF-2 Disk Drive	AIX, Linux

Feature code	Description	OS support
#1885	300 GB 10 K RPM SFF SAS Disk Drive	AIX, Linux
#1880	300 GB 15 K RPM SFF SAS Disk Drive	AIX, Linux
#1953	300 GB 10 K RPM SAS SFF-2 Disk Drive	AIX, Linux
#1790	600 GB 10 K RPM SAS SFF Disk Drive	AIX, Linux
#1964	600 GB 10 K RPM SAS SFF-2 Disk Drive	AIX, Linux
#1909	69 GB SFF SAS SSD	IBM i
#1888	139.5 GB 15 K RPM SFF SAS Disk Drive	IBM i
#1947	139 GB 15 K RPM SAS SFF-2 Disk Drive	IBM i
#1787	177 GB SFF-1 SSD with eMLC	IBM i
#1794	177 GB SFF-2 SSD with eMLC	IBM i
#1996	177 GB SSD Module with eMLC	IBM i
#1956	283 GB 10 K RPM SAS SFF-2 Disk Drive	IBM i
#1911	283 GB 10 K RPM SFF SAS Disk Drive	IBM i
#1879	283 GB 15 K RPM SAS SFF Disk Drive	IBM i
#1948	283 GB 15 K RPM SAS SFF-2 Disk Drive	IBM i
#1916	571 GB 10 K RPM SAS SFF Disk Drive	IBM i
#1962	571 GB 10 K RPM SAS SFF-2 Disk Drive	IBM i

Table 1-11 shows the available disk drive feature codes available for the installation in an I/O enclosure external to a Power 720 and Power 740 server.

Table 1-11 Non CEC Disk drive feature code description

Feature code	Description	OS support
#3586	69 GB 3.5" SAS SSD	AIX, Linux
#3647	146 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3648	300 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3649	450 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3587	69 GB 3.5" SAS SSD	IBM i
#3677	139.5 GB 15 K RPM SAS Disk Drive	IBM i
#3678	283.7 GB 15 K RPM SAS Disk Drive	IBM i
#3658	428 GB 15 K RPM SAS Disk Drive	IBM i

Certain adapters are available for order in large quantities. Table 1-12 lists the Gen2 disk drives available in a quantity of 150.

Table 1-12 Available Gen2 disk drives in quantity of 150

Feature code	Description
#1817	Quantity 150 of #1962 (571 GB 10 K RPM SAS SFF-2 Disk Drive)
#1818	Quantity 150 of #1964 (600 GB 10 K RPM SAS SFF-2 Disk Drive)
#1844	Quantity 150 of #1956 (283 GB 10 K RPM SAS SFF-2 Disk Drive)
#1866	Quantity 150 of #1917 (146 GB 15 K RPM SAS SFF-2 Disk Drive)
#1868	Quantity 150 of #1947 (139 GB 15 K RPM SAS SFF-2 Disk Drive)
#1869	Quantity 150 of #1925 (300 GB 10 K RPM SAS SFF-2 Disk Drive)
#1887	Quantity 150 of #1793 (177 GB SFF-2 SSD with eMLC)
#1958	Quantity 150 of #1794 (177 GB SFF-2 SSD with eMLC)

Additional considerations for SAS-bay-based SSDs: Be aware of the following considerations for SAS-bay-based SSDs (#1775, #1787, #1793, #1794, #1890, #1909, #3586, and #3587):

- Feature codes #1775, #1787, #1793, #1794, #1890, and #1909 are supported in the Power 720 and Power 740 CEC.
- The 3.5-inch feature codes #3586 and #3587 are not supported in the Power 720 and Power 740 CEC.
- SSDs and HDDs are not allowed to mirror each other.
- SSDs are not supported by feature codes #5278, #5900, #5901, #5902, and #5912.
- When an SSD is placed in higher-function backplane (#EJ01), no EXP 12S Expansion Drawer (#5886) or EXP24S SFF Gen2-bay Drawer (#5887) is supported to connect to the system's external SAS port.
- When an SSD is placed in a EXP 12S Expansion Drawer (#5886) or EXP24S SFF Gen2-bay Drawer (#5887), the drawer is not allowed to connect to the system's external SAS port.
- A maximum of eight SSDs per EXP 12S Expansion Drawer (#5886) is allowed. No mixing of SSDs and HDDs is allowed in the EXP 12S Expansion Drawer (#5886). A maximum of one feature code #5886 EXP12S drawer containing SSDs attached to a single controller or pair of controllers is allowed. A EXP 12S Expansion Drawer (#5886) containing SSD drives cannot be connected to other feature code 5886's.
- In a Power 720 or Power 740 with a split backplane (3x3), SSDs and HDDs can be placed in either "split," but no mixing of SSDs and HDDs within a split is allowed. IBM i does not support split backplane.
- In a Power 720 or Power 740 without a split backplane, SSDs and HDDs can be mixed in any combination. However, they cannot be in the same RAID array.
- HDD/SSD Data Protection: If IBM i (#2145) is selected, one of the following is required:
  - Disk mirroring (default), which requires feature code #0040, #0043, or #0308
  - SAN boot (#0837)
  - RAID, which requires feature code #5630
  - Mixed Data Protection (#0296)

If you need more disks than available with the internal disk bays, you can attach additional external disk subsystems.

SCSI disks are not supported in the Power 720 and 740 disk bays. However, if you want to use SCSI disks, you can attach existing SCSI disk subsystems.

For more detailed information about the available external disk subsystems, see 2.11, “External disk subsystems” on page 74.

The Power 720 and 740 have a slim media bay that can contain an optional DVD-RAM (#5762) and a half-high bay that can contain a tape drive or removable disk drive.

Table 1-13 shows the available media device feature codes for Power 720 and 740.

Table 1-13 Media device feature code description for Power 720 and 740

Feature code	Description
#1103	USB Internal Docking Station for Removable Disk Drive
#1104	USB External Docking Station for Removable Disk Drive
#5619	80/160 GB DAT160 Tape-SAS
#5638	1.5 TB/3.0 TB LTO-5 Tape-SAS
#5746	800 GB/1.6 TB LTO4 Tape-SAS
#5762	SATA Slimline DVD-RAM Drive

Additional considerations: Take notice of these considerations for tape drives and USB disk drives:

- If tape device feature #5619, #5638, or #5746 is installed in the half-high media bay, feature #3656 must be also selected.
- A half-high tape feature and a feature #1103 Removable USB Disk Drive Docking Station are mutually exclusive. One or the other can be in the half-high bay in the system but not both. As for the tape drive, the #3656 is not required with #1103.

1.6 I/O drawers for Power 720 and Power 740 servers

The Power 720 and Power 740 servers support the following 12X attached I/O drawers, providing extensive capability to expand the overall server expandability and connectivity:

- Feature #5802 provides 10 PCIe slots and 18 SFF SAS disk slots.
- Feature #5877 provides 10 PCIe slots.
- Feature #5796 provides six PCI-X slots (supported but not orderable).
- The 7314-G30 drawer provides six PCI-X slots (supported but not orderable).

Three disk-only I/O drawers are also supported, providing large storage capacity and multiple partition support:

- The feature #5886 EXP 12S SAS drawer holds a 3.5-inch SAS disk or SSD.
- The feature #5887 EXP24S SFF Gen2-bay drawer for high-density storage holds SAS hard disk drives.

- The feature #5786 Totalstorage EXP24 disk drawer and #5787 Totalstorage EXP24 disk tower holds a 3.5-inch SCSI disk (used for migrating existing SCSI drives supported but not orderable).
- The 7031-D24 holds a 3.5-inch SCSI disk (supported but not orderable).

The Power 720 provides one GX++ slot, offering one connection loop. The Power 740 has one GX++ slot if one processor module is installed, and two GX++ slots when two processor modules are installed. Thus, the Power 740 provides one or two connection loops.

1.6.1 12X I/O Drawer PCIe expansion units

The 12X I/O Drawer PCIe, SFF disk (#5802) and 12X I/O Drawer PCIe, no disk (#5877) expansion units are 19-inch, rack-mountable, I/O expansion drawers that are designed to be attached to the system using 12x double date rate (DDR) cables. The expansion units can accommodate 10 generation 3 blind-swap cassettes. These cassettes can be installed and removed without removing the drawer from the rack.

The #5802 I/O drawer has the following attributes:

- Eighteen SAS hot-swap SFF disk bays
- Ten PCIe based I/O adapter slots (blind-swap)
- Redundant hot-swappable power and cooling units

The #5877 drawer is the same as #5802 except that it does not support any disk bays.

A maximum of two #5802 or #5877 drawers can be placed on the same 12X loop. The #5877 I/O drawer can be on the same loop as the #5802 I/O drawer. A #5877 drawer cannot be upgraded to a #5802 drawer.

Note: Mixing #5802/#5877 and #5796 on the same loop is not supported. Mixing #5802 and #5877 on the same loop is supported with a maximum of two drawers total per loop.

1.6.2 PCI-X DDR 12X Expansion Drawer

The PCI-X DDR 12X Expansion Drawer (#5796) and 7314-G30 are a 4 EIA unit tall drawer and mounts in a 19-inch rack. Feature #5796 takes up half the width of the 4 EIA rack space and requires the use of a #7314 drawer mounting enclosure. The 4 EIA tall enclosure can hold up to two #5796 drawers mounted side by side in the enclosure. A maximum of four #5796 drawers can be placed on the same 12X loop.

The I/O drawer has the following attributes:

- One or two #5796 drawers are held by the 4 EIA unit rack-mount enclosure (#7314).
- Six PCI-X DDR slots, 64-bit, 3.3 V, 266 MHz that use blind swap-cassettes.
- Redundant hot-swappable power and cooling units.

The 7314-G30 is equivalent to the #5796 I/O drawer described before. It provides the same six PCI-X DDR slots per unit and has the same configuration rules and considerations as the #5796 drawer.

Note: Mixing #5802/#5877 and #5796 on the same loop is not supported. Mixing #5796 and the 7314-G30 on the same loop is supported with a maximum of four drawers total per loop.

IBM i does not support the 7314-G30 I/O drawer.

1.6.3 I/O drawers and usable PCI slot

The various I/O drawer model types can be intermixed on a single server within the appropriate I/O loop. Depending on the system configuration, the maximum number of I/O drawers supported can vary.

Table 1-14 summarizes the maximum number of I/O drawers supported and the total number of PCI slots available when expansion consists of a single drawer type.

Table 1-14 Maximum number of I/O drawers supported and total number of PCI slots

Server	Processor cards	Max #5796 drawers	Max #5802 and #5877 drawers[a]	Total number of slots			
				#5796		#5802 and #5877	
				PCI-X	PCIe	PCI-X	PCIe
Power 720	One	4	2	24	8[a]	0	30[a]
Power 740	One	4	2	24	8[a]	0	30[a]
Power 740	Two	8	4	48	8[a]	0	30[a]

a. Four additional slots are low-profile PCIe Gen2 slots only.

Table 1-15 summarizes the maximum number of disk-only I/O drawers supported.

Table 1-15 Maximum number of disk-only I/O drawers supported

Server	Processor cards	Max #5886 drawers	Max #5887 drawers	Max #7314-G30 drawers
Power 720	One	28	14	4
Power 740	One	28	14	4
Power 740	Two	28	14	8

Note: The 4-core Power 720 does not support the attachment of 12X I/O drawers or the attachment of disk drawers such as the #5886 EXP 12S SAS drawer, #5887 EXP24S SFF Gen2-bay drawer, #5786 Totalstorage EXP24 disk drawer, or #5787 Totalstorage EXP24 disk tower.

1.6.4 EXP 12S SAS drawer

The EXP 12S SAS drawer (#5886) is a 2 EIA drawer and mounts in a 19-inch rack. The drawer can hold either SAS disk drives or SSD. The EXP 12S SAS drawer has twelve 3.5-inch SAS disk bays with redundant data paths to each bay. The SAS disk drives or SSDs contained in the EXP 12S are controlled by one or two PCIe or PCI-X SAS adapters connected to the EXP 12S via SAS cables.

Feature #5886 can also be directly attached to the SAS port on the rear of the Power 720 and Power 740, providing a low-cost disk storage solution. When used this way, the imbedded SAS controllers in the system unit drive the disk drives in EXP12S. A second unit cannot be cascaded to a feature #5886 attached in this way.

1.6.5 EXP24S SFF Gen2-bay drawer

The EXP24S SFF Gen2-bay drawer is an expansion drawer supporting up to twenty-four 2.5-inch hot-swap SFF SAS HDDs on POWER6 or POWER7 servers in 2U of 19-inch rack space. The EXP24S bays are controlled by SAS adapters/controllers attached to the I/O drawer by SAS X or Y cables.

The SFF bays of the EXP24S are different from the SFF bays of the POWER7 system units or 12X PCIe I/O drawers (#5802 and #5803). The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the Gen1 or SFF-1 bays of the POWER7 system unit or 12X PCIe I/O Drawers or vice versa.

The EXP24S includes redundant AC power supplies and two power cords.

1.7 Build to Order

You can perform a Build to Order or *a la carte* configuration using the IBM Configurator for e-business (e-config), where you specify each configuration feature that you want on the system.

Preferably, begin with one of the available starting configurations, such as the IBM Editions. These solutions are available at initial system order time with a starting configuration that is ready to run as is.

1.8 IBM Editions

IBM Editions are available only as an initial order. If you order a server IBM Edition as defined next, you can qualify for half the initial configuration's processor core activations at no additional charge.

The total memory (based on the number of cores) and the quantity/size of disk, SSD, Fibre Channel adapters, or Fibre Channel over Ethernet (FCoE) adapters shipped with the server are the only features that determine whether a customer is entitled to a processor activation at no additional charge.

When you purchase an IBM Edition, you can purchase an AIX, IBM i, or Linux operating system license, or you can choose to purchase the system with no operating system. The AIX, IBM i, or Linux operating system is processed by means of a feature code on one of these:

- AIX 5.3, 6.1, or 7.1
- IBM i 6.1.1 or IBM i 7.1
- SUSE Linux Enterprise Server or Red Hat Enterprise Linux

If you choose AIX 5.3, 6.1, or 7.1 for your primary operating system, you can also order IBM i 6.1.1 or IBM i 7.1 and SUSE Linux Enterprise Server or Red Hat Enterprise Linux.

The converse is true if you choose an IBM i or Linux subscription as your primary operating system.

These sample configurations can be changed as needed and still qualify for processor entitlements at no additional charge. However, selection of total memory or HDD or SSD/Fibre Channel/FCoE adapter quantities smaller than the totals defined as the minimums disqualifies the order as an IBM Edition, and the no-charge processor activations are then removed.

Consider these minimum definitions for IBM Editions:

- For Power 720, a minimum of 2 GB of memory per core is needed to qualify for the IBM Edition. There can be different valid memory configurations that meet the minimum requirement.
- For the Power 740, a minimum of 4 GB of memory per core is needed to qualify for the IBM Edition. For example, a 4-core minimum is 16 GB, a 6-core minimum is 24 GB, and an 8-core minimum is 32 GB. There can be different valid memory configurations that meet the minimum requirement.

Also, a minimum of two HDDs, or two SSDs, or two Fibre Channel adapters, or two FCoE adapters is required. You only need to meet one of these disk/SSD/FC/FCoE criteria. Partial criteria cannot be combined.

1.8.1 Express editions for IBM i

Express editions for IBM i enable initial ease of ordering and feature a lower price than if you ordered them *a la carte* or build-to-order. Taking advantage of the edition is the only way that you can use no-charge features for processor activations and IBM i user license entitlements. The Express editions are available only during the initial system order and cannot be ordered after your system is shipped.

The IBM configurator offers these easy-to-order Express editions that include no-charge activations or no-charge IBM i user entitlements. You can modify the Express Edition configurations to match your exact requirements for your initial shipment by increasing or decreasing the configuration. If you create a configuration that falls below any of the defined minimums, the IBM configurator replaces the no-charge features with equivalent function regular charge features.

1.8.2 Express editions for Power 720

To configure a Power 720 4-core Power 720 Express Edition (#0777) and use the no-charge features on your initial order, you must order the following components:

- 3.0 GHz 4-core processor module (#EPC5)
- IBM i Primary Operating System Indicator (#2145)
- 8 GB minimum memory: 2 x 4 GB (#EM04) or 1 x 8 GB (#EM08)

> **Note:** Memory features #EM16 and #EM32 are not supported with the 4-core processor module.

- Minimum of two HDDs, or two SSDs, or two Fibre Channel adapters, or two FCoE adapters. You only need to meet one of these disk/SSD/FC/FCoE criteria. Partial criteria cannot be combined.

If the above requirements are met, the following are included:

- Two no-charge activations (2 x #EPE5)
- IBM i user entitlements (no-charge)
- One IBM i Access Family license with unlimited users (57xx-XW1)
- Reduced price on 57xx-WDS and 5733-SOA

To use the no-charge features on your initial order of 6-core and 8-core Power 720 Express Editions (#0779), you must order:

- 3.0 GHz 6-core processor module (#EPC6) or 3.0 GHz 8-core processor module (#EPC7)
- IBM i Primary Operating System Indicator (#2145)
- 16 GB minimum memory: 4 x 4 GB (#EM04), or 2 x 8 GB (#EM08), or 1 x 16 GB (#EM16), or 1 x 32 GB (#EM32)
- Minimum of two HDD, or two SSD, or two Fibre Channel adapters, or two FCoE adapters. You only need to meet one of these disk/SSD/FC/FCoE criteria. Partial criteria cannot be combined.

If the above requirements are met, the following are included:

- Three no-charge activations (3 x #EPE6) with feature #EPC6 or four no-charge activations (4 x #EPE7) with feature #EPC7
- Thirty IBM i user entitlements (charged)
- One IBM i Access Family license with unlimited users (57xx-XW1)
- Reduced price on 57xx-WDS and 5733-SOA

Note: The Power 740 does not have an Express Edition for the IBM i feature code available.

1.9 IBM i Solution Edition for Power 720 and Power 740

The IBM i Solution Editions for Power 720 and Power 740 are designed to help you take advantage of the combined experience and expertise of IBM and independent software vendors (ISVs) in building business value with your IT investments. A qualifying purchase of software, maintenance, services or training for a participating ISV solution is required when purchasing an IBM i Solution Edition.

The Power 720 IBM i Solution Edition feature code #4928 supports the 4-core configuration and feature code (#4927) supports both 6-core and 8-core configurations. The Power 720 Solution Edition includes no-charge features resulting in a lower initial list price for qualifying clients. Also included is an IBM Service voucher to help speed implementation of the ISV solution.

The Power 740 IBM i Solution Edition (#4929) supports 4-core to 16-core configurations. The Power 740 Solution Edition includes no-charge features resulting in a lower initial list price for qualifying clients. Also included is an IBM Service voucher to help speed implementation of the ISV solution.

For a list of participating ISVs, a registration form, and additional details, visit the Solution Edition website:

http://www-03.ibm.com/systems/power/hardware/editions/solutions.html

These are the requirements to be eligible to purchase a Solution Edition order:

- The offering must include new and/or upgrade software licenses and/or software maintenance from the ISV for the qualifying IBM server. Services and training for the qualifying server can also be provided.
- Proof of purchase of the solution with a participating ISV must be provided to IBM on request. The proof must be dated within 90 days before or after the date of order of the qualifying server.

1.10 IBM i for Business Intelligence

Business Intelligence remains top priority of mid-market companies, but budgets and staff/skills to support enterprise BI solutions are very small in comparison to enterprise accounts.

Table 1-16 lists the three new orderable hardware features that generate a configuration of defaults/minimums for the Power 720.

Table 1-16 List of available hardware features for IBM i for Business Intelligence

Feature	Feature code
#4934	IBM i for BI - Small configuration
#4935	IBM i for BI - Medium configuration
#4936	IBM i for BI - Large configuration

Note: The IBM i for Business Intelligence solution is not available for the Power 740.

1.11 Model upgrade

A model upgrade from a Power 520 to the Power 720, preserving the existing serial number, is available. You can upgrade the 2-core or 4-core Power 520 (8203-E4A) with IBM POWER6 or POWER6+™ processors to the 6-core or 8-core IBM Power 720 (8202-E4C) with POWER7 processors. For upgrades from POWER6 or POWER6+ processor-based systems, IBM will install new CEC enclosures to replace the existing enclosures. The current CEC enclosures will be returned to IBM.

Note: The model upgrade is being performed from a system (8203-E4A) with a 1-year warranty into a system (8202-E4C) with a 3-year warranty.

1.11.1 Upgrade considerations

Feature conversions have been set up for the IBM POWER6 and IBM POWER6+ processors to POWER7 processors.

Table 1-17 shows the supported conversions for the processors.

Table 1-17 Processor conversions

Power 520	Power 720
#5634 2-core 4.2 GHz Processor Card	#EPC6 6-core 3.0 GHz POWER7 Processor Module
#5577 2-core 4.7 GHz Processor Card	#EPC6 6-core 3.0 GHz POWER7 Processor Module
#5635 4-core 4.2 GHz Processor Card	#EPC6 6-core 3.0 GHz POWER7 Processor Module
#5587 4-core 4.7 GHz Processor Card	#EPC6 6-core 3.0 GHz POWER7 Processor Module
#5634 2-core 4.2 GHz Processor Card	#EPC7 8-core 3.0 GHz POWER7 Processor Module
#5577 2-core 4.7 GHz Processor Card	#EPC7 8-core 3.0 GHz POWER7 Processor Module
#5635 4-core 4.2 GHz Processor Card	#EPC7 8-core 3.0 GHz POWER7 Processor Module
#5587 4-core 4.7 GHz Processor Card	#EPC7 8-core 3.0 GHz POWER7 Processor Module

1.11.2 Features

The following features present on the current system can be moved to the new system:

- All PCIe adapters with cables
- All line cords, keyboards, and displays
- PowerVM Express, Standard, or Enterprise Editions (#5225, #5227, and #5228)
- I/O drawers (#5786, #5796, #5802, #5877, #5886, and #5887)
- Racks (#0551, #0553, and #0555)
- Rack doors (#6068, #6069, #6248, and #6249)
- Rack trim kits (#6246 and 6247)
- SATA DVD-ROM (#5743)
- SATA DVD-RAM (#5762)

The Power 720 can support the following 12X drawers and disk-only drawers:

- #5802 and #5877 PCIe 12X I/O drawers
- #5796 and 7413-G30 PCI-X (12X) I/O Drawer
- #5786 and 7031-D24 IBM TotalStorage EXP24 SCSI Disk Drawer
- #5886 EXP12S SAS Disk Drawer
- #5887 EXP 24S SFF Gen2-bay Drawer

Note: In the Power 720 system unit SAS bays, only the SAS SFF hard disks or SFF solid-state drives are supported internally. The 3.5-inch HDD or SSD can be attached to the Power 720 but must be located in a EXP 12S drawer (#5886).

1.12 Server and virtualization management

If you want to implement partitions, a Hardware Management Console (HMC), the Integrated Virtualization Manager (IVM), or the new released IBM Systems Director Management

Console (SDMC) is required to manage the Power 720 and Power 740 servers. Multiple POWER6 and POWER7 processor-based servers can be supported by a single HMC or SDMC.

Note: If you do not use an HMC or IVM or SDMC, the Power 720 and Power 740 runs in full system partition mode, meaning that a single partition owns all the server resources and only one operating system can be installed.

If an HMC is used to manage the Power 720 and Power 740, the HMC must be a rack-mount CR3 or later or a deskside C05 or later.

The IBM Power 720 and IBM Power 740 servers require the Licensed Machine Code Version 7 Revision 740.

Remember: You can download or order the latest HMC code from the Fix Central website:

http://www.ibm.com/support/fixcentral

Existing HMC models 7310 can be upgraded to Licensed Machine Code Version 7 to support environments that can include IBM POWER5, IBM POWER5+, POWER6, and POWER7 processor-based servers. Licensed Machine Code Version 6 (#0961) is not available for 7042 HMCs.

When IBM Systems Director is used to manage an HMC, or if the HMC manages more than 254 partitions, the HMC must have a minimum of 3 GB RAM and must be a rack-mount CR3 model, or later, or deskside C06 or later.

Future enhancements: At the time of writing, the SDMC is not supported for the Power 720 (8202-E4C) and Power 740 (8205-E6C) models.

IBM intends to enhance the IBM Systems Director Management Console (SDMC) to support the Power 720 (8202-E4C) and Power 740 (8205-E6C). IBM also intends for the current Hardware Management Console (HMC) 7042-CR6 to be upgradable to an IBM SDMC that supports the Power 720 (8202-E4C) and Power 740 (8205-E6C).

1.13 System racks

The Power 720 and Power 740 and their I/O drawers are designed to mount in the 25U 7014-S25 (#0555), 36U 7014-T00 (#0551), or 42U 7014-T42 (#0553) rack. These racks are built to the 19-inch EIA standard.

Remember: A new Power 720 or Power 740 server can be ordered with the appropriate 7014 rack model. The racks are available as features of the Power 720 and Power 740 only when an additional I/O drawer for an existing system (MES order) is ordered. The rack feature number must be used if IBM manufacturing has to integrate the newly ordered I/O drawer in a 19-inch rack before shipping the MES order.

If a system is to be installed in a non-IBM rack or cabinet, ensure that the rack meets the requirements that are described in 1.13.10, “OEM rack” on page 28.

Remember: It is the client's responsibility to ensure that the installation of the drawer in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the drawer requirements for power, cooling, cable management, weight, and rail security.

1.13.1 IBM 7014 Model S25 rack

The 1.3-meter (49-inch) Model S25 rack has the following features:

- Twenty-five EIA units
- Weights:
  - Base empty rack: 100.2 kg (221 lb.)
  - Maximum load limit: 567.5 kg (1250 lb.)

The S25 racks do not have vertical mounting space that will accommodate #7188 PDUs. All PDUs required for application in these racks must be installed horizontally in the rear of the rack. Each horizontally mounted PDU occupies 1U of space in the rack, and therefore reduces the space available for mounting servers and other components.

1.13.2 IBM 7014 Model T00 rack

The 1.8 Meter (71-in.) Model T00 is compatible with past and present IBM Power systems. The T00 rack has these features:

- It has 36 EIA units (36 U) of usable space.
- These optional features are available:
  - Optional removable side panels
  - Optional highly perforated front door
  - Optional side-to-side mounting hardware for joining multiple racks
- You have a choice of standard business black or optional white color in OEM format.
- Increased power distribution and weight capacity is provided.
- Support for both AC and DC configurations is included.
- The rack height is increased to 1926 mm (75.8 in.) if a power distribution panel is fixed to the top of the rack.
- Up to four power distribution units (PDUs) can be mounted in the PDU bays (Figure 1-4 on page 25), but others can fit inside the rack. See 1.13.7, "The AC power distribution unit and rack content" on page 24.
- The weights are as follows:
  - T00 base empty rack: 244 kg (535 lb.)
  - T00 full rack: 816 kg (1795 lb.)

1.13.3 IBM 7014 Model T42 rack

The 2.0-meter (79.3-inch) Model T42 addresses the client requirement for a tall enclosure to house the maximum amount of equipment in the smallest possible floor space. The following features differ in the Model T42 rack from the Model T00:

- There are 42 EIA units (42 U) of usable space (6 U of additional space).
- The Model T42 supports AC only.
- Rack weights are as follows:
  - T42 base empty rack: 261 kg (575 lb.)
  - T42 full rack: 930 kg (2045 lb.)

1.13.4 Feature code 0555 rack

The 1.3-meter rack (#0555) is a 25 EIA unit rack. The rack that is delivered as #0555 is the same rack delivered when you order the 7014-S25 rack. The included features can vary. The #0555 is supported, but it is no longer orderable.

1.13.5 Feature code 0551 rack

The 1.8-meter rack (#0551) is a 36 EIA unit rack. The rack that is delivered as #0551 is the same rack delivered when you order the 7014-T00 rack. The included features can vary. Certain features that are delivered as part of the 7014-T00 must be ordered separately with the #0551.

1.13.6 Feature code 0553 rack

The 2.0-meter rack (#0553) is a 42 EIA unit rack. The rack that is delivered as #0553 is the same rack delivered when you order the 7014-T42 or B42 rack. The included features can vary. Some features that are delivered as part of the 7014-T42 or B42 must be ordered separately with the #0553.

1.13.7 The AC power distribution unit and rack content

For rack models T00 and T42, 12-outlet PDUs are available. These include AC power distribution units #9188 and #7188 and AC Intelligent PDU+ #5889 and #7109.

Four PDUs can be mounted vertically in the back of the T00 and T42 racks. See Figure 1-4 for the placement of the four vertically mounted PDUs. In the rear of the rack, two additional PDUs can be installed horizontally in the T00 rack and three in the T42 rack. The four vertical mounting locations will be filled first in the T00 and T42 racks. Mounting PDUs horizontally consumes 1U per PDU and reduces the space available for other racked components. When mounting PDUs horizontally, it is best to use fillers in the EIA units occupied by these PDUs to facilitate proper air-flow and ventilation in the rack.

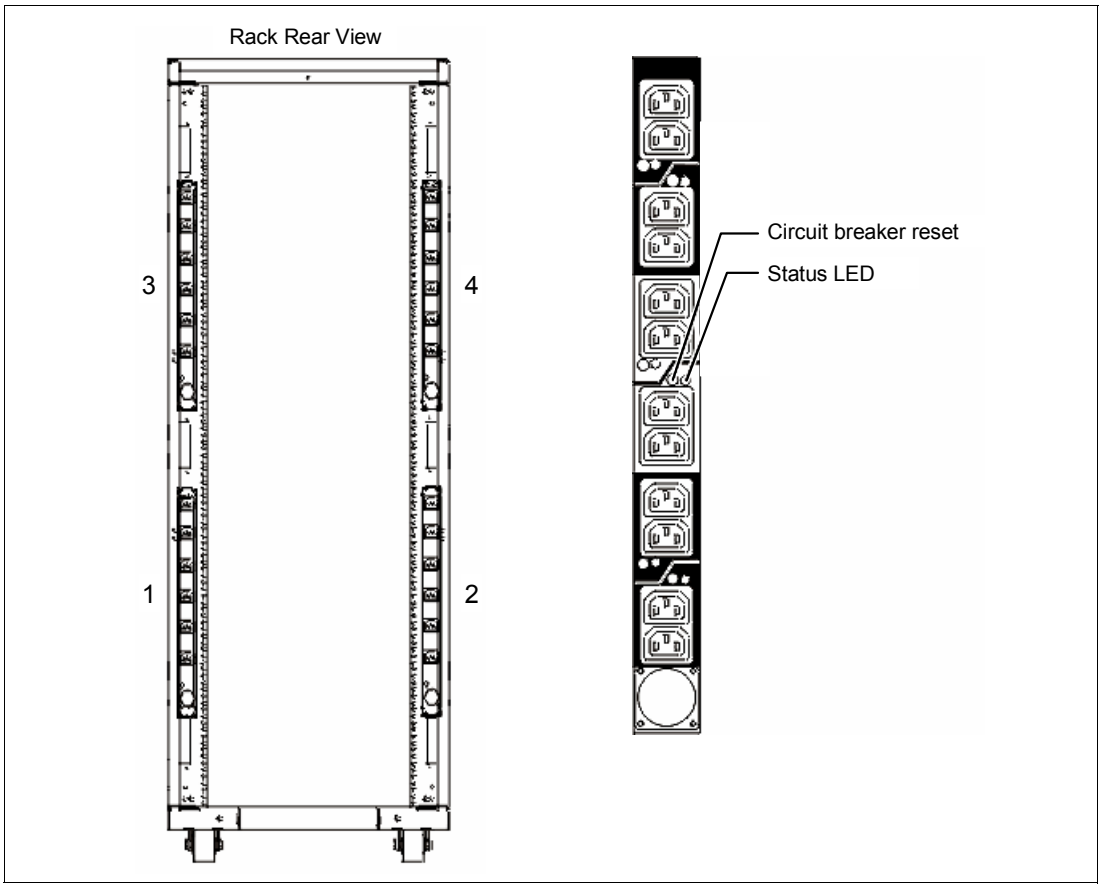


Figure 1-4 PDU placement and PDU view

For detailed power cord requirements and power cord feature codes, see the IBM Power Systems Hardware Information Center at the following website:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

Note: Ensure that the appropriate power cord feature is configured to support the power being supplied.

The Intelligent PDU+, base option, 1 EIA Unit, Universal, UTG0247 Connector (#5889), the Base/Side Mount Universal PDU (#9188) and the optional, additional, Universal PDU (#7188) and the Intelligent PDU+ options (#7109) support a wide range of country requirements and electrical power specifications. The #5889 and #7109 PDUs are identical to #9188 and #7188 PDUs but are equipped with one Ethernet port, one console serial port, and one rs232 serial port for power monitoring.

The PDU receives power through a UTG0247 power line connector. Each PDU requires one PDU-to-wall power cord. Various power cord features are available for various countries and

applications by varying the PDU-to-wall power cord, which must be ordered separately. Each power cord provides the unique design characteristics for the specific power requirements. To match new power requirements and save previous investments, these power cords can be requested with an initial order of the rack or with a later upgrade of the rack features.

The PDU has 12 client-usable IEC 320-C13 outlets. There are six groups of two outlets fed by six circuit breakers. Each outlet is rated up to 10 amps, but each group of two outlets is fed from one 15 amp circuit breaker.

Note: Based on the power cord that is used, the PDU can supply from 4.8 kVA to 19.2 kVA. The power of all the drawers plugged into the PDU must not exceed the power cord limitation.

The Universal PDUs are compatible with previous models.

Note: Each system drawer to be mounted in the rack requires two power cords, which are not included in the base order. For maximum availability, it is highly desirable to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to independent power sources.

1.13.8 Rack-mounting rules

Follow these primary rules when mounting the system into a rack:

- The system is designed to be placed at any location in the rack. For rack stability, it is advisable to start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripherals, provided that the maximum permissible weight of the rack is not exceeded and the installation rules for these devices are followed.
- Before placing the system into the service position, it is essential that the rack manufacturer's safety instructions have been followed regarding rack stability.

1.13.9 Useful rack additions

This section highlights solutions available for IBM Power Systems rack-based systems.

IBM System Storage 7214 Tape and DVD Enclosure

The IBM System Storage 7214 Tape and DVD Enclosure is designed to mount in one EIA unit of a standard IBM Power Systems 19-inch rack and can be configured with one or two tape drives, or either one or two Slim DVD-RAM or DVD-ROM drives in the right-side bay.

The two bays of the IBM System Storage 7214 Tape and DVD Enclosure can accommodate the following tape or DVD drives for IBM Power servers:

- DAT72 36 GB Tape Drive - up to two drives
- DAT72 36 GB Tape Drive - up to two drives
- DAT160 80 GB Tape Drive - up to two drives
- Half-high LTO Ultrium 4 800 GB Tape Drive - up to two drives
- DVD-RAM Optical Drive - up to two drives
- DVD-ROM Optical Drive - up to two drives

IBM System Storage 7216 Multi-Media Enclosure

The IBM System Storage 7216 Multi-Media Enclosure (Model 1U2) is designed to attach to the Power 720 and the Power 740 through a USB port on the server, or through a PCIe SAS adapter. The 7216 has two bays to accommodate external tape, removable disk drive, or DVD-RAM drive options.

These optional drive technologies are available for the 7216-1U2:

- DAT160 80 GB SAS Tape Drive (#5619
- DAT320 160 GB SAS Tape Drive (#1402)
- DAT320 160 GB USB Tape Drive (#5673)
- Half-high LTO Ultrium 5 1.5 TB SAS Tape Drive (#8247)
- DVD-RAM - 9.4 GB SAS Slim Optical Drive (#1420 and 1422
- RDX Removable Disk Drive Docking Station (#1103)

Note: The DAT320 160 GB SAS Tape Drive (#1402) and the DAT320 160 GB USB Tape Drive (#5673) are no longer available as of July 15, 2011.

To attach a 7216 Multi-Media Enclosure to the Power 720 and Power 740, consider the following cabling procedures:

- Attachment by an SAS adapter

  A PCIe Dual-X4 SAS adapter (#5901) or a PCIe LP 2-x4-port SAS Adapter 3 Gb (#5278) must be installed in the Power 720 and Power 740 server in order to attach to a 7216 Model 1U2 Multi-Media Storage Enclosure. Attaching a 7216 to a Power 720 and Power 740 through the integrated SAS adapter is not supported.

  For each SAS tape drive and DVD-RAM drive feature installed in the 7216, the appropriate external SAS cable will be included.

  An optional Quad External SAS cable is available by specifying (#5544) with each 7216 order. The Quad External Cable allows up to four 7216 SAS tape or DVD-RAM features to attach to a single System SAS adapter.

  Up to two 7216 storage enclosure SAS features can be attached per PCIe Dual-X4 SAS adapter (#5901) or the PCIe LP 2-x4-port SAS Adapter 3 Gb (#5278).

- Attachment by an USB adapter

  The Removable RDX HDD Docking Station features on 7216 only support the USB cable that is provided as part of the feature code. Additional USB hubs, add-on USB cables, or USB cable extenders are not supported.

  For each RDX Docking Station feature installed in the 7216, the appropriate external USB cable will be included. The 7216 RDX Docking Station feature can be connected to the external, integrated USB ports on the Power 710 and Power 730 or to the USB ports on 4-Port USB PCI Express Adapter (# 2728).

  The 7216 DAT320 USB tape drive or RDX Docking Station features can be connected to the external, integrated USB ports on the Power 710 and Power 730.

The two drive slots of the 7216 enclosure can hold the following drive combinations:

- One tape drive (DAT160 SAS or Half-high LTO Ultrium 5 SAS) with second bay empty
- Two tape drives (DAT160 SAS or Half-high LTO Ultrium 5 SAS) in any combination
- One tape drive (DAT160 SAS or Half-high LTO Ultrium 5 SAS) and one DVD-RAM SAS drive sled with one or two DVD-RAM SAS drives
- Up to four DVD-RAM drives

- One tape drive (DAT160 SAS or Half-high LTO Ultrium 5 SAS) in one bay, and one RDX Removable HDD Docking Station in the other drive bay
- One RDX Removable HDD Docking Station and one DVD-RAM SAS drive sled with one or two DVD-RAM SAS drives in the right bay
- Two RDX Removable HDD Docking Stations

Figure 1-5 shows the 7216 Multi-Media Enclosure.

Figure 1-5 7216 Multi-Media Enclosure

In general, the 7216-1U2 is supported by the AIX, IBM i, and Linux operating system. However, the RDX Removable Disk Drive Docking Station and the DAT320 USB Tape Drive are not supported with IBM i.

Flat panel display options

The IBM 7316 Model TF3 is a rack-mountable flat panel console kit consisting of a 17-inch 337.9 mm x 270.3 mm flat panel color monitor, rack keyboard tray, IBM Travel Keyboard, support for IBM Keyboard/Video/Mouse (KVM) switches, and language support. The IBM 7316-TF3 Flat Panel Console Kit offers these features:

- Slim, sleek, lightweight monitor design that occupies only 1 U (1.75 inches) in a 19-inch standard rack
- A 17-inch, flat screen TFT monitor with truly accurate images and virtually no distortion
- Ability to mount the IBM Travel Keyboard in the 7316-TF3 rack keyboard tray
- Support for IBM KVM switches that provide control of as many as 128 servers, and support of both USB and PS/2 server-side keyboard and mouse connections

1.13.10 OEM rack

The system can be installed in a suitable OEM rack, provided that the rack conforms to the EIA-310-D standard for 19-inch racks. This standard is published by the Electrical Industries Alliance. For detailed information see the IBM Power Systems Hardware Information Center:

http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp

The key points mentioned are as follows:

- The front rack opening must be 451 mm wide ± 0.75 mm (17.75 in. ± 0.03 in.), and the rail-mounting holes must be 465 mm ± 0.8 mm (18.3 in. ± 0.03 in.) apart on center (horizontal width between the vertical columns of holes on the two front-mounting flanges and on the two rear-mounting flanges). Figure 1-6 shows a top view showing the specification dimensions.

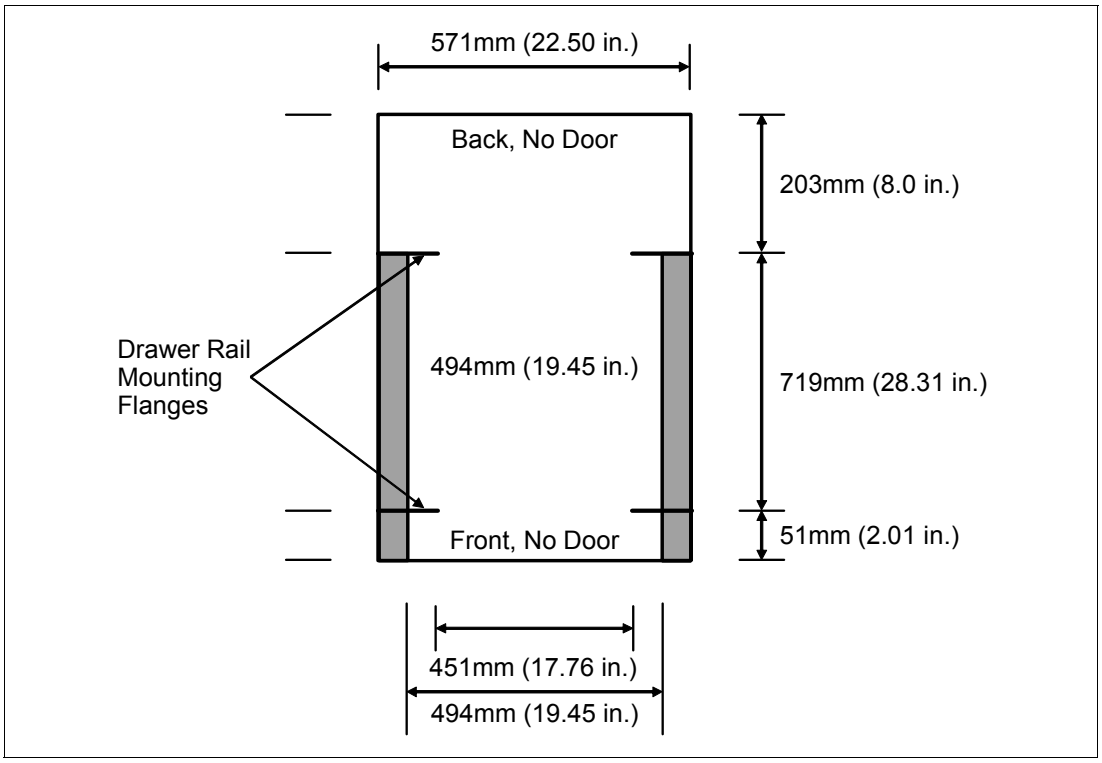


Figure 1-6 Top view of non-IBM rack specification dimensions

- The vertical distance between the mounting holes must consist of sets of three holes spaced (from bottom to top) 15.9 mm (0.625 in.), 15.9 mm (0.625 in.), and 12.67 mm (0.5 in.) on center, making each three-hole set of vertical hole spacing 44.45 mm (1.75 in.) apart on center. Rail-mounting holes must be 7.1 mm ± 0.1 mm (0.28 in. ± 0.004 in.) in diameter. Figure 1-7 shows the top front specification dimensions.

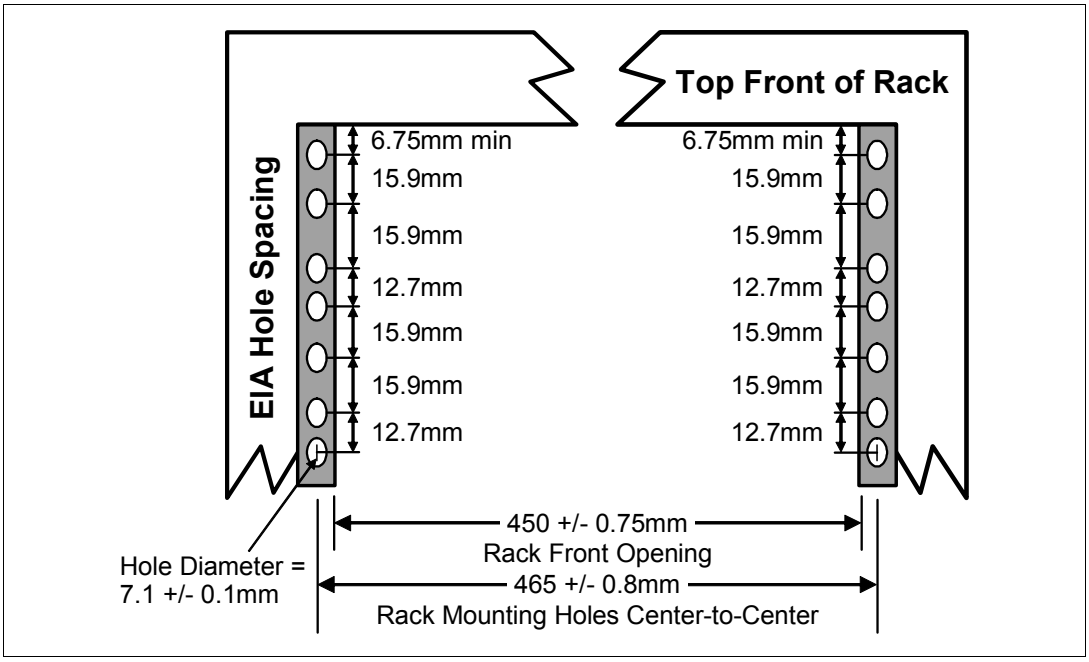


Figure 1-7 Rack specification dimensions, top front view

2

Architecture and technical overview

This chapter discusses the overall system architecture for the IBM Power 720 and Power 740, represented by Figure 2-1 on page 32 and Figure 2-2 on page 33. The bandwidths that are provided throughout the section are theoretical maximums used for reference.

The speeds shown are at an individual component level. Multiple components and application implementation are key to achieving the best performance.

Always do the performance sizing at the application workload environment level and evaluate performance using real-world performance measurements and production workloads.

Figure 2-1 shows the logical system diagram of the Power 720.

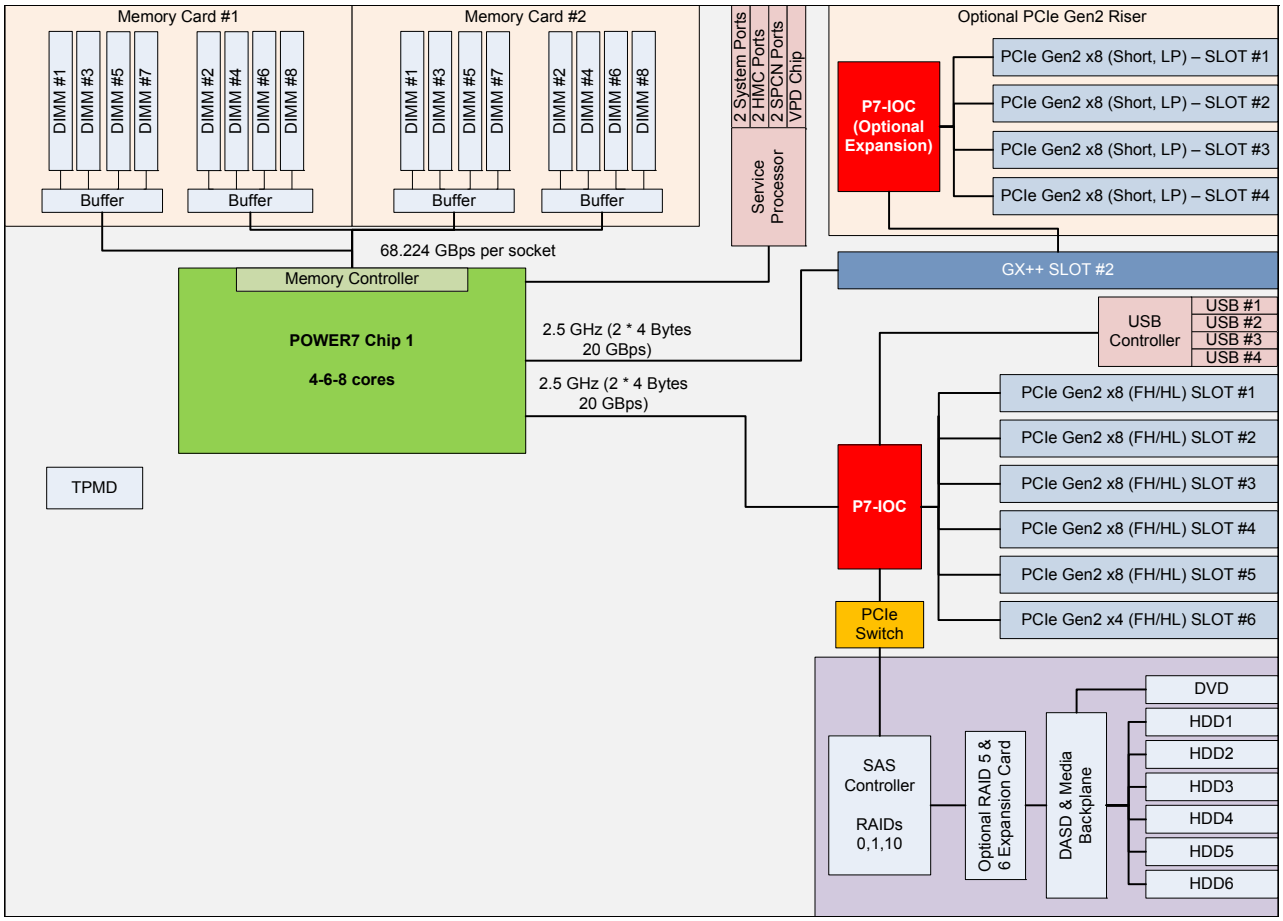


Figure 2-1 IBM Power 720 logical system diagram

Figure 2-2 shows the logical system diagram of the Power 740.

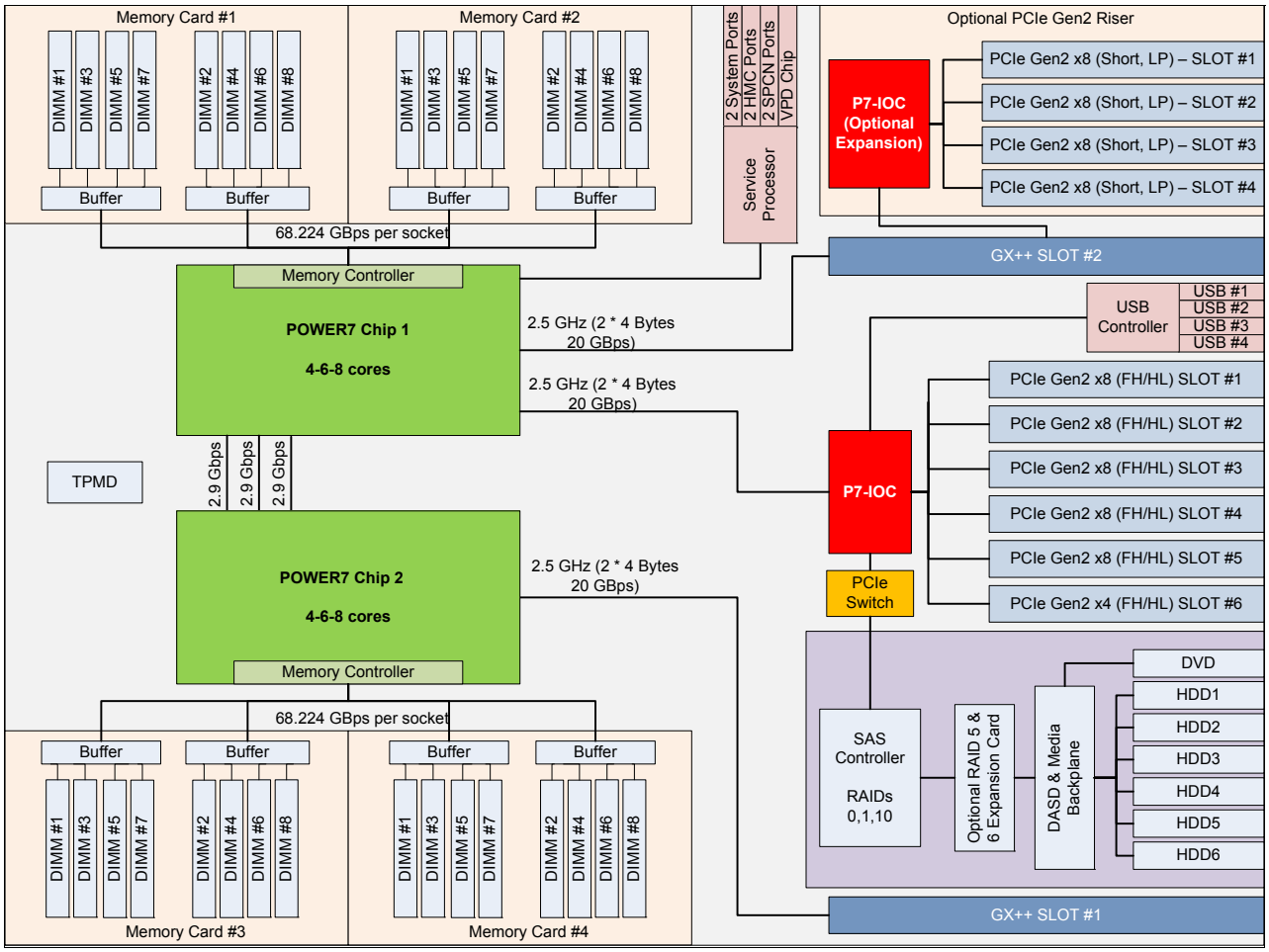


Figure 2-2 IBM Power 740 logical system diagram

2.1 The IBM POWER7 processor

The IBM POWER7 processor represents a leap forward in technology achievement and associated computing capability. The multi-core architecture of the POWER7 processor has been matched with innovation across a wide range of related technologies to deliver leading throughput, efficiency, scalability, and RAS.

Although the processor is an important component in delivering outstanding servers, many elements and facilities have to be balanced on a server to deliver maximum throughput. As with previous generations of systems based on POWER® processors, the design philosophy for POWER7 processor-based systems is one of system-wide balance in which the POWER7 processor plays an important role.

In many cases, IBM has been innovative to achieve required levels of throughput and bandwidth. Areas of innovation for the POWER7 processor and POWER7 processor-based systems include, but are not limited to, these features:

- On-chip L3 cache implemented in embedded dynamic random access memory (eDRAM)
- Cache hierarchy and component innovation
- Advances in memory subsystem
- Advances in off-chip signalling
- Exploitation of long-term investment in coherence innovation

The superscalar POWER7 processor design also provides a variety of other capabilities:

- Binary compatibility with the prior generation of POWER processors
- Support for PowerVM virtualization capabilities, including PowerVM Live Partition Mobility to and from POWER6 and POWER6+ processor-based systems

Figure 2-3 shows the POWER7 processor die layout with the major areas identified:

- Processor cores
- L2 cache
- L3 cache and chip interconnection
- Symmetric Multi Processing (SMP) links
- Memory controllers

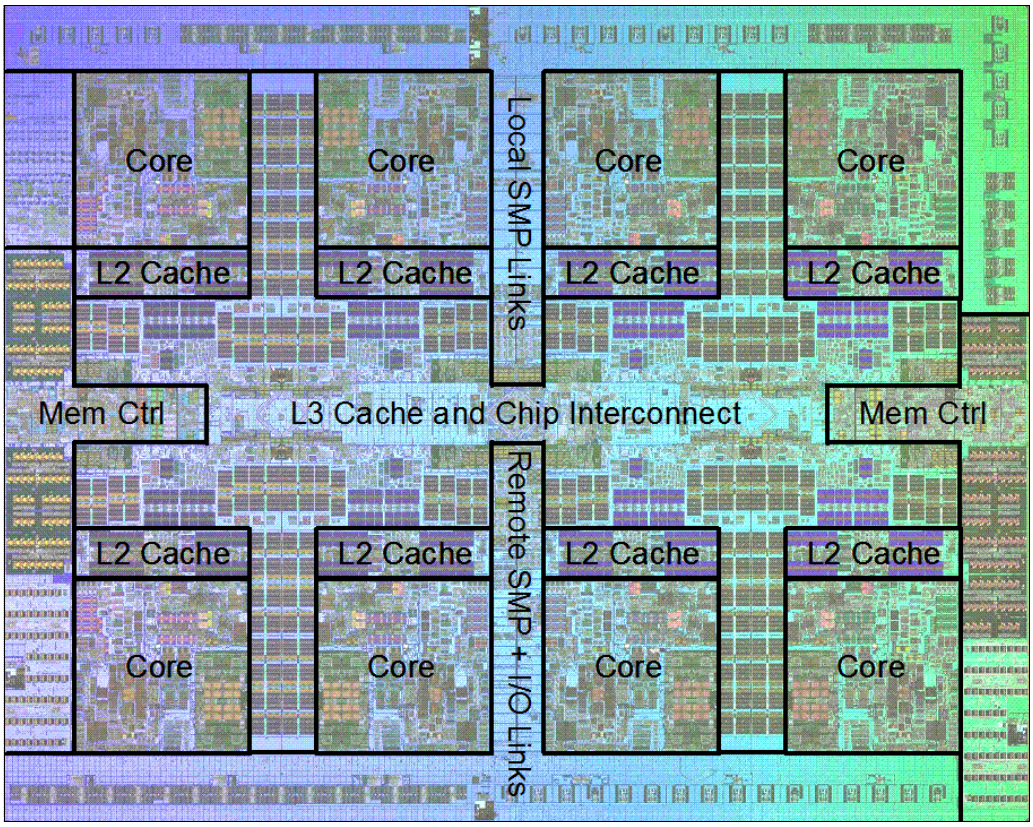


Figure 2-3 POWER7 processor die with key areas indicated

2.1.1 POWER7 processor overview

The POWER7 processor chip is fabricated using the IBM 45 nm Silicon-On-Insulator (SOI) technology using copper interconnect and implements an on-chip L3 cache using eDRAM.

The POWER7 processor chip has an area of 567 mm^2 and is built using 1.2 billion components (transistors). Eight processor cores are on the chip, each with 12 execution units, 256 KB of L2 cache, and access to 32 MB of shared on-chip L3 cache.

For memory access, the POWER7 processor includes two Double Data Rate 3 (DDR3) memory controllers, each with four memory channels. To be able to scale effectively, the POWER7 processor uses a combination of local and global SMP links with very high coherency bandwidth and takes advantage of the IBM dual-scope broadcast coherence protocol.

Table 2-1 summarizes the technology characteristics of the POWER7 processor.

Table 2-1 Summary of POWER7 processor technology

Technology	POWER7 processor
Die size	567 mm^2
Fabrication technology	► 45 nm lithography ► Copper interconnect ► Silicon-on-Insulator ► eDRAM
Components	1.2 billion components/transistors offering the equivalent function of 2.7 billion (For further details, see 2.1.6, “On-chip L3 cache innovation and Intelligent Cache” on page 40.)
Processor cores	4, 6, or 8
Max execution threads per core/chip	4/32
L2 cache per core/chip	256 KB/2 MB
On-chip L3 cache per core/chip	4 MB/32 MB
DDR3 memory controllers	1 or 2
SMP design-point	32 sockets with IBM POWER7 processors
Compatibility	With prior generation of POWER processor

2.1.2 POWER7 processor core

Each POWER7 processor core implements aggressive out-of-order (OoO) instruction execution to drive high efficiency in the use of available execution paths. The POWER7 processor has an *instruction sequence unit* that is capable of dispatching up to six instructions per cycle to a set of queues. Up to eight instructions per cycle can be issued to the *instruction execution units*.

The POWER7 processor has a set of 12 execution units:

- Two fixed point units
- Two load store units
- Four double precision floating point units
- One vector unit
- One branch unit
- One condition register unit
- One decimal floating point unit

These caches are tightly coupled to each POWER7 processor core:

- Instruction cache: 32 KB
- Data cache: 32 KB
- L2 cache: 256 KB, implemented in fast SRAM

2.1.3 Simultaneous multithreading

An enhancement in the POWER7 processor is the addition of the SMT4 mode to enable four instruction threads to execute simultaneously in each POWER7 processor core. Thus, the instruction thread execution modes of the POWER7 processor are as follows:

- SMT1: Single instruction execution thread per core
- SMT2: Two instruction execution threads per core
- SMT4: Four instruction execution threads per core

Maximizing throughput

SMT4 mode enables the POWER7 processor to maximize the throughput of the processor core by offering an increase in core efficiency. SMT4 mode is the latest step in an evolution of multithreading technologies introduced by IBM. Figure 2-4 shows the evolution of simultaneous multithreading.

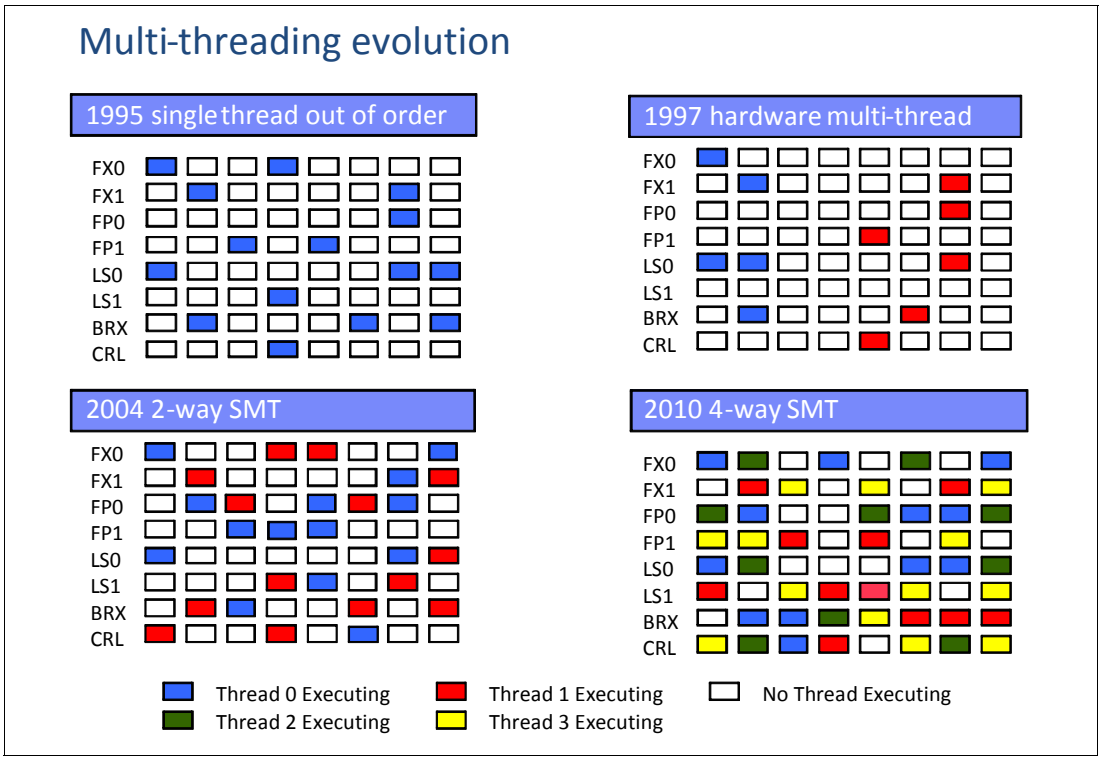


Figure 2-4 Evolution of simultaneous multi-threading

The various SMT modes offered by the POWER7 processor allow flexibility, enabling users to select the threading mode that meets an aggregation of objectives such as performance, throughput, energy use, and workload enablement.

Intelligent Threads

The POWER7 processor features Intelligent Threads that can vary based on the workload demand. The system either automatically selects (or the system administrator can manually select) whether a workload benefits from dedicating as much capability as possible to a single thread of work, or if the workload benefits more from having capability spread across two or four threads of work. With more threads, the POWER7 processor can deliver more total capacity as more tasks are accomplished in parallel. With fewer threads, those workloads that need very fast individual tasks can get the performance that they need for maximum benefit.

2.1.4 Memory access

Each POWER7 processor chip has two DDR3 memory controllers, each with four memory channels (enabling eight memory channels per POWER7 processor chip). Each channel operates at 6.4 GHz and can address up to 32 GB of memory. Thus, each POWER7 processor chip is capable of addressing up to 256 GB of memory.

Figure 2-5 gives a simple overview of the POWER7 processor memory access structure.

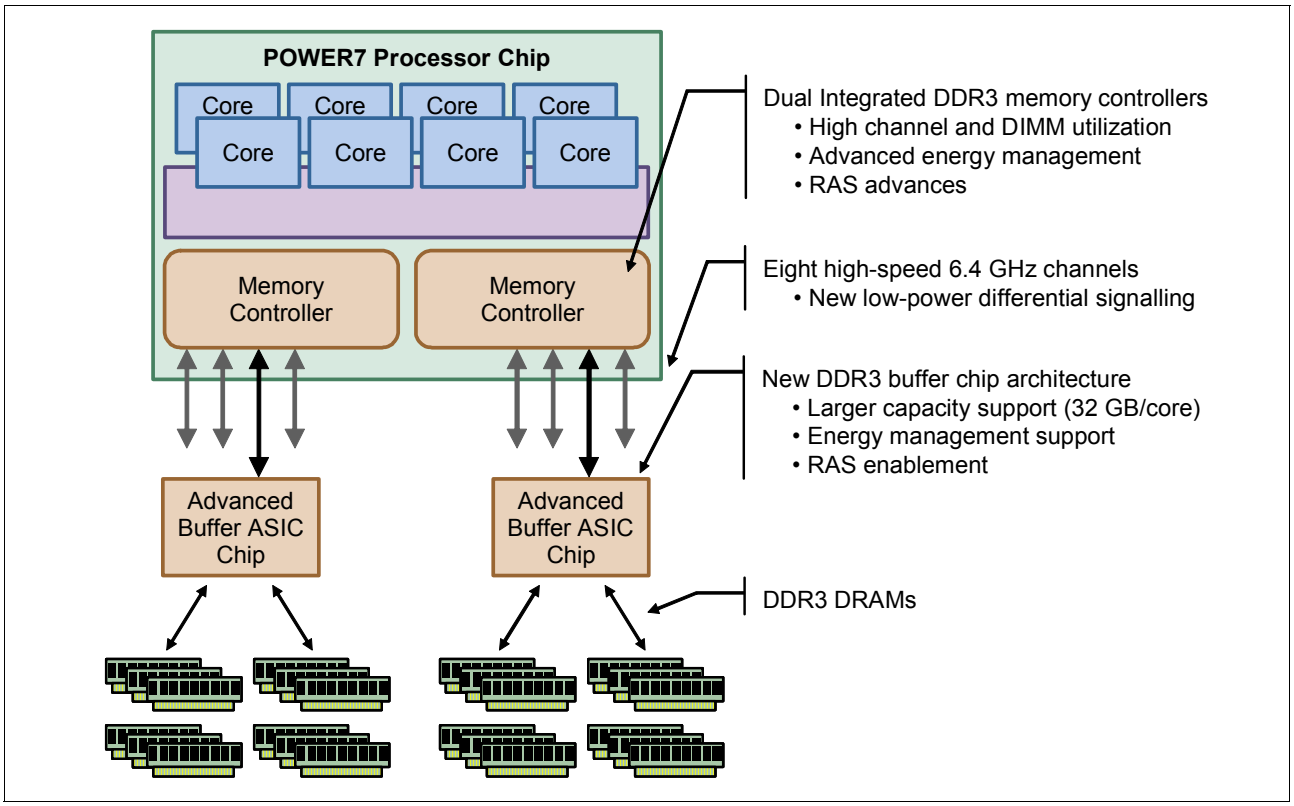


Figure 2-5 Overview of POWER7 memory access structure

2.1.5 Flexible POWER7 processor packaging and offerings

The POWER7 processor forms the basis of a flexible compute platform and can be offered in a number of guises to address differing system requirements.

The POWER7 processor can be offered with a single active memory controller with four channels for servers where higher degrees of memory parallelism are not required.

Similarly, the POWER7 processor can be offered with a variety of SMP bus capacities that are appropriate to the scaling-point of particular server models.

Figure 2-6 outlines the physical packaging options that are supported with POWER7 processors.

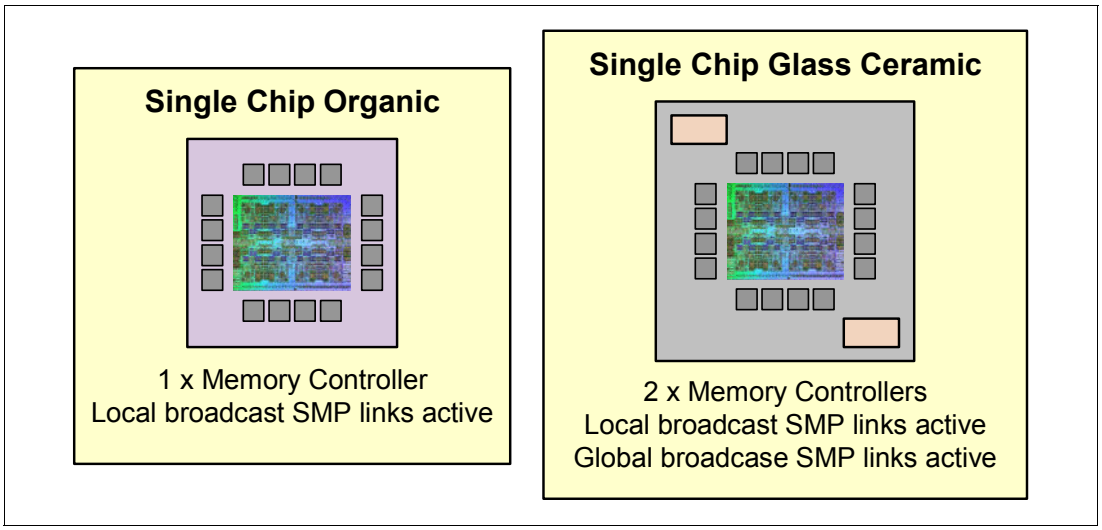


Figure 2-6 Outline of the POWER7 processor physical packaging

POWER7 processors have the unique ability to optimize to various workload types. For example, database workloads typically benefit from very fast processors that handle high transaction rates at high speeds. Web workloads typically benefit more from processors with many threads that allow the breaking down of web requests into many parts and handle them in parallel. POWER7 processors uniquely have the ability to provide leadership performance in either case.

TurboCore mode

Users can choose to run selected servers in TurboCore mode. This mode uses four cores per POWER7 processor chip with access to the entire 32 MB of L3 cache (8 MB per core) and at a faster processor core frequency, which delivers higher performance per core, and might save on software costs for those applications that are licensed per core.

Note: TurboCore is available on the Power 780 and Power 795.

MaxCore mode

MaxCore mode is for workloads that benefit from a higher number of cores and threads handling multiple tasks simultaneously, taking advantage of increased parallelism. MaxCore mode provides up to eight cores and up to 32 threads per POWER7 processor.

POWER7 processor 4-core and 6-core offerings

The base design for the POWER7 processor is an 8-core processor with 32 MB of on-chip L3 cache (4 MB per core). However, the architecture allows for differing numbers of processor cores to be active (4 cores or 6 cores, as well as the full 8-core version).

In most cases (MaxCore mode), the L3 cache associated with the implementation is dependent on the number of active cores. For a 6-core version, this typically means that 6 x 4 MB (24 MB) of L3 cache is available. Similarly, for a 4-core version, the L3 cache available is 16 MB.

2.1.6 On-chip L3 cache innovation and Intelligent Cache

A breakthrough in material engineering and microprocessor fabrication has enabled IBM to implement the L3 cache in eDRAM and place it on the POWER7 processor die. L3 cache is critical to a balanced design, as is the ability to provide good signalling between the L3 cache and other elements of the hierarchy such as the L2 cache or SMP interconnect.

The on-chip L3 cache is organized into separate areas with differing latency characteristics. Each processor core is associated with a Fast Local Region of L3 cache (FLR-L3) but also has access to other L3 cache regions as shared L3 cache. Additionally, each core can negotiate to use the FLR-L3 cache associated with another core, depending on reference patterns. Data can also be cloned to be stored in more than one core's FLR-L3 cache, again depending on reference patterns. This Intelligent Cache management enables the POWER7 processor to optimize the access to L3 cache lines and minimize overall cache latencies.

Figure 2-7 shows the FLR-L3 cache regions for each core on the POWER7 processor die.

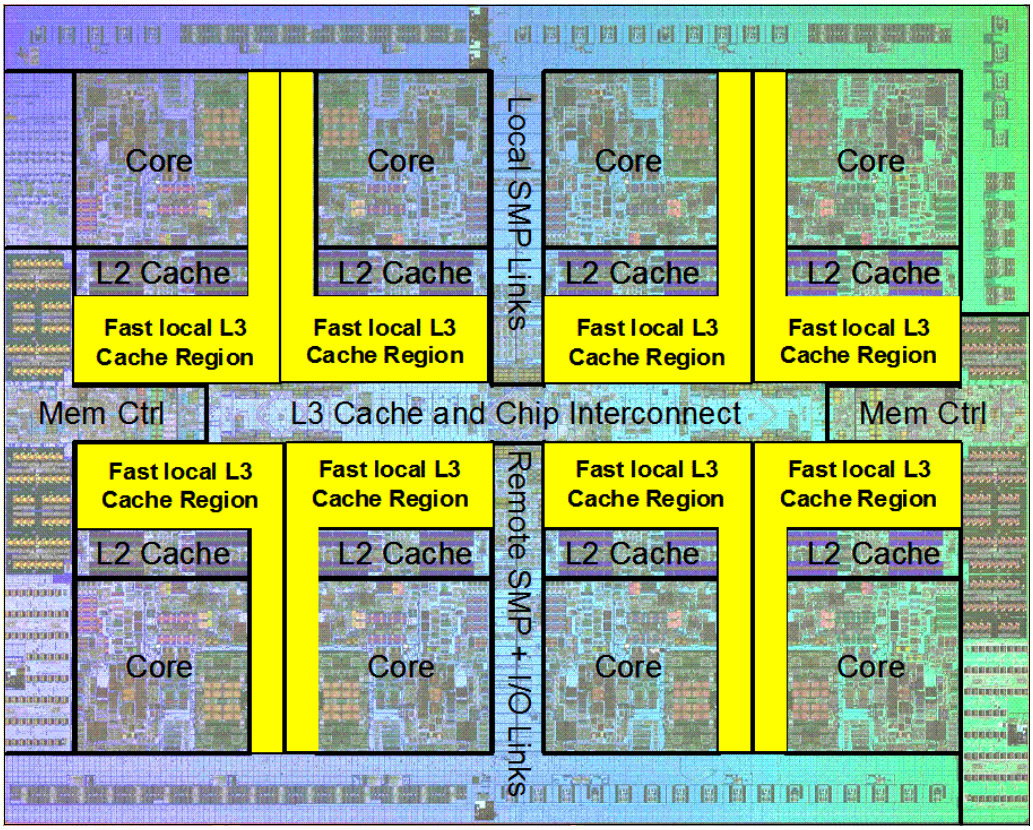


Figure 2-7 Fast local regions of L3 cache on the POWER7 processor

Innovation using eDRAM on the POWER7 processor die is significant for these reasons:

- Latency improvement

  A six-to-one latency improvement occurs by moving the L3 cache on-chip compared to L3 accesses on an external (on-ceramic) ASIC.

- Bandwidth improvement

  A 2x bandwidth improvement occurs with on-chip interconnect. Frequency and bus sizes are increased to and from each core.

- No off-chip driver or receivers

  Removing drivers or receivers from the L3 access path lowers interface requirements, conserves energy, and lowers latency.

- Small physical footprint

  The eDRAM L3 cache requires far less physical space than an equivalent L3 cache implemented with conventional SRAM. IBM on-chip eDRAM uses only a third of the components used in conventional SRAM, which has a minimum of six transistors to implement a 1-bit memory cell.

- Low energy consumption

  The on-chip eDRAM uses only 20% of the standby power of SRAM.

2.1.7 POWER7 processor and Intelligent Energy

Energy consumption is an important area of focus for the design of the POWER7 processor, which includes Intelligent Energy features that help to dynamically optimize energy usage and performance so that the best possible balance is maintained. Intelligent Energy features such as EnergyScale work with IBM Systems Director Active Energy Manager™ to dynamically optimize processor speed based on thermal conditions and system utilization.

2.1.8 Comparison of the POWER7 and POWER6 processors

Table 2-2 shows comparable characteristics between the generations of POWER7 and POWER6 processors.

Table 2-2 Comparison of technology for the POWER7 processor and the prior generation

Feature	POWER7	POWER6
Technology	45 nm	65 nm
Die size	567 mm^2	341 mm^2
Maximum cores	8	2
Maximum SMT threads per core	4 threads	2 threads
Maximum frequency	4.25 GHz	5 GHz
L2 cache	256 KB per core	4 MB per core
L3 cache	4 MB of FLR-L3 cache per core with each core having access to the full 32 MB of L3 cache, on-chip eDRAM	32 MB off-chip eDRAM ASIC
Memory support	DDR3	DDR2
I/O bus	Two GX++	One GX++
Enhanced cache mode (TurboCore)	Yes[a]	No
Sleep and nap modes[b]	Both	Nap only

a. Not supported on the Power 770 and the Power 780 4-socket systems.

b. For more information about sleep and nap modes, see 2.15.1, “IBM EnergyScale technology” on page 94.

2.2 POWER7 processor modules

The Power 720 and Power 740 server chassis house POWER7 processor modules that host POWER7 processor sockets (SCM) and eight DDR3 memory DIMM slots for each processor module.

The Power 720 server houses one processor module offering 4-core 3.0 GHz, 6-core 3.0 GHz, or 8-core 3.0 GHz configurations.

The Power 740 server houses one or two processor modules offering 4-core or 8-core 3.3 GHz and 3.7 GHz, 6-core or 12-core 3.7 GHz, or 8-core and 16-core 3.55 GHz configurations.

All off the installed processors must be activated, unless they are factory deconfigured using feature #2319.

Note: All POWER7 processors in the system must be the same frequency and have the same number of processor cores. POWER7 processor types cannot be mixed within a system.

2.2.1 Modules and cards

Figure 2-8 shows the system planar highlighting the POWER7 processor modules and the memory riser cards.

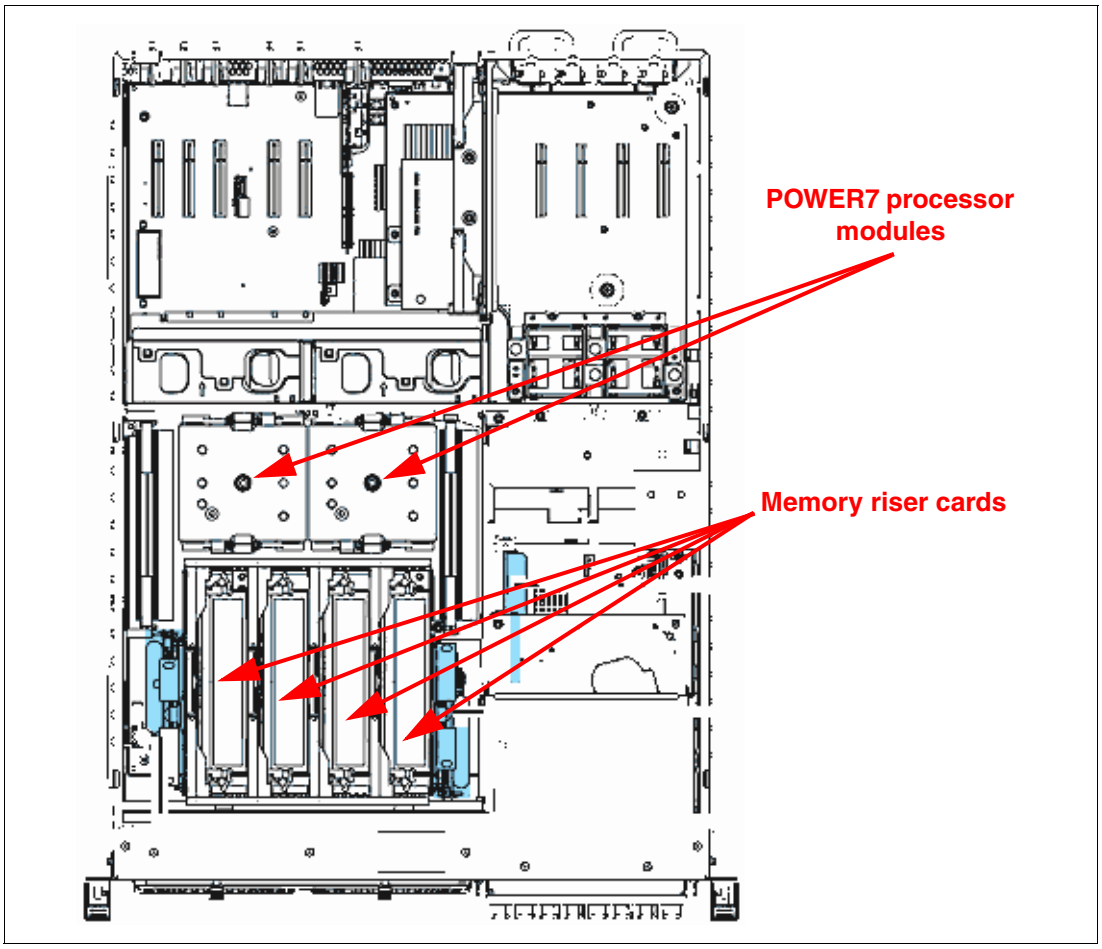


Figure 2-8 POWER7 processor card shown in a Power 740 with processor module

2.2.2 Power 720 and Power 740 systems

Power 720 and Power 740 systems support POWER7 processors with various core-counts. Table 2-3 summarizes the POWER7 processor options for the Power 720 system.

Table 2-3 Summary of POWER7 processor options for the Power 720 system

Feature	Cores per POWER7 processor	Frequency (GHz)	Processor activation	Min/Max cores per system	Min/Max processor module
#EPC5	4	3.0	The 4-core 3.0 GHz requires that four processor activation codes are ordered, available as 4 x #EPD5 or 2 x #EPD5 and 2 x #EPE5.	4/4	1/1

Feature	Cores per POWER7 processor	Frequency (GHz)	Processor activation	Min/Max cores per system	Min/Max processor module
#EPC6	6	3.0	The 6-core 3.0 GHz requires that six processor activation codes be ordered, available as 6 x #EPD6 or 3 x #EPD6 and 3 x #EPE6.	6/6	1/1
#EPE7	8	3.0	The 8-core 3.0 GHz requires that eight processor activation codes be ordered, available as 8 x #EPD7 or 4 x #EPD7 and 4 x #EPE7.	8/8	1/1

Table 2-4 summarizes the POWER7 processor options for the Power 740 system.

Table 2-4 Summary of POWER7 processor options for the Power 740 system

Feature	Cores per POWER7 processor	Frequency (GHz)	Processor activation	Min/Max cores per system	Min/Max processor module
#EPE9	4	3.3	The 4-core 3.3 GHz requires that four processor activation codes are ordered, available as 4 x #EPD9 or 2 x #EPD9 and 2 x #EPE9.	4/8	1/2
#EPE8	4	3.7	The 4-core 3.7 GHz requires that four processor activation codes are ordered, available as 4 x #EPD8 or 2 x #EPD8 and 2 x #EPE8.	4/8	1/2
#EPEA	6	3.7	The 6-core 3.7 GHz requires that six processor activation codes are ordered, available as 6 x #EPDA or 3 x #EPDA and 3 x #EPEA.	6/12	1/2
#EPEB	8	3.55	2 x of the 8 core 3.55 GHz requires that 16 processor activation codes are ordered, available as 16 x #EPDB or 8 x #EPDB and 8 x #EPEB.	16	2/2

2.3 Memory subsystem

The Power 720 is a one-socket system supporting a single POWER7 processor module. The server supports a maximum of 16 DDR3 DIMM slots, with eight DIMM slots included in the base configuration and eight DIMM slots available with an optional memory riser card. Memory features (two memory DIMMs per feature) supported are 4 GB, 8 GB, 16 GB, and 32 GB running at speeds of 1066 MHz. A system with the optional memory riser card installed has a maximum memory of 256 GB.

The Power 740 is a two-socket system supporting up to two POWER7 processor modules. The server supports a maximum of 32 DDR3 DIMM slots, with eight DIMM slots included in the base configuration and 24 DIMM slots available with three optional memory riser cards. Memory features (two memory DIMMs per feature) supported are 4 GB, 8 GB, 16 GB, and 32 GB run at speeds of 1066 MHz. A system with three optional memory riser cards installed has a maximum memory of 512 GB.

2.3.1 Registered DIMM

Industry standard DDR3 Registered DIMM (RDIMM) technology is used to increase reliability, speed, and density of memory subsystems.

2.3.2 Memory placement rules

The following memory options are orderable:

- 4 GB (2 x 2 GB) Memory DIMMs, 1066 MHz (#EM04)
- 8 GB (2 x 4 GB) Memory DIMMs, 1066 MHz (#EM08)
- 16 GB (2 x 8 GB) Memory DIMMs, 1066 MHz (#EM16)
- 32 GB (2 x 16 GB) Memory DIMMs, 1066 MHz (#EM32)

The minimum memory capacity for the Power 720 and Power 740 systems is 4 GB (2 x 2 GB DIMMs). Table 2-5 shows the maximum memory supported on the Power 720.

Table 2-5 Power 720 maximum memory

Processor cores	One memory riser card	Two memory riser cards
4-core	32 GB	64 GB
6-core 8-core	128 GB	256 GB
Note: A system with the 4-core processor module (#EPC5) does not support the 16 GB (#EM16) and 32 GB (#EM32) memory features.		

Table 2-6 shows the maximum memory supported on the Power 740.

Table 2-6 Power 740 maximum memory

Processor cores	One memory riser card	Two memory riser cards	Three memory riser cards	Four memory riser cards
1 x 4-core 1 x 6-core 1 x 8-core	128 GB	256 GB	Not available	Not available
2 x 4-core 2 x 6-core 2 x 8-core	128 GB	256 GB	384 GB	512 GB

Remember: DDR2 memory (used in POWER6 processor-based systems) is not supported in POWER7 processor-based systems.

Figure 2-9 shows the logical memory DIMM topology for the POWER7 processor card.

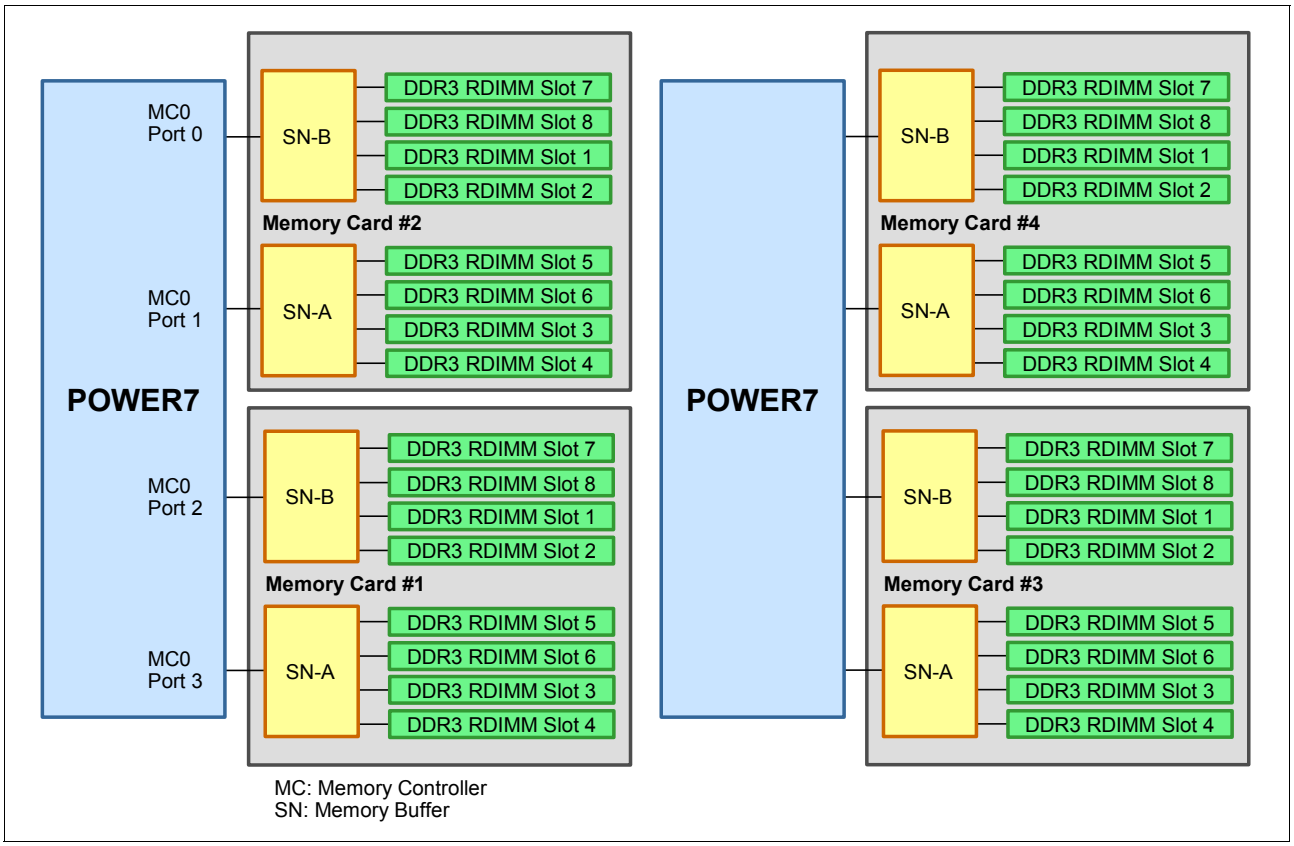


Figure 2-9 Memory DIMM topology for the Power 740

Figure 2-10 shows memory location codes and how the Memory Riser Cards are divided in Quads, each Quad being attached to a memory buffer.

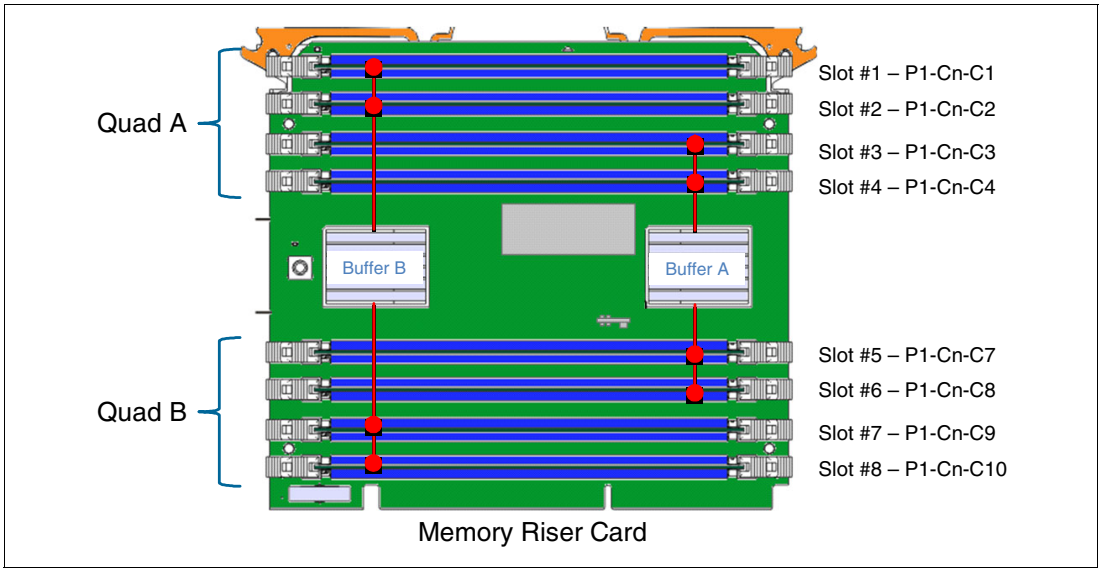


Figure 2-10 Memory Riser Card for Power 720 and Power 740 Systems

The memory-placement rules are as follows:

- The base machine contains one memory riser card with eight DIMM sockets. Memory features occupy two memory DIMM sockets.
- One additional memory riser card feature (1 x #EM01) with an additional eight DIMM sockets is available when one processor module is installed in the system. For the Power 740, three optional memory riser card features (3 x #EM01) with an additional eight DIMM sockets per feature are available when two processor modules are installed in the system.
- Each DIMM within a DIMM quad must be equivalent. However, Quad B DIMMs can be different from the Quad A DIMMs.
- Mixing features #EM04, #EM08, #EM16, or #EM32 is supported on the same memory riser card as long as there is only one type of memory DIMM in the same Quad.

It is generally best to install memory evenly across all memory riser cards in the system. Balancing memory across the installed memory riser cards allows memory access in a consistent manner and typically results in the best possible performance for your configuration. However, balancing memory fairly evenly across multiple memory riser cards, compared to balancing memory exactly evenly, typically has a very small performance difference.

Take into account any plans for future memory upgrades when deciding which memory feature size to use at the time of initial system order.

Figure 2-11 shows the installation order of the DIMMs along with the location codes, based on the physical view of the memory cards.

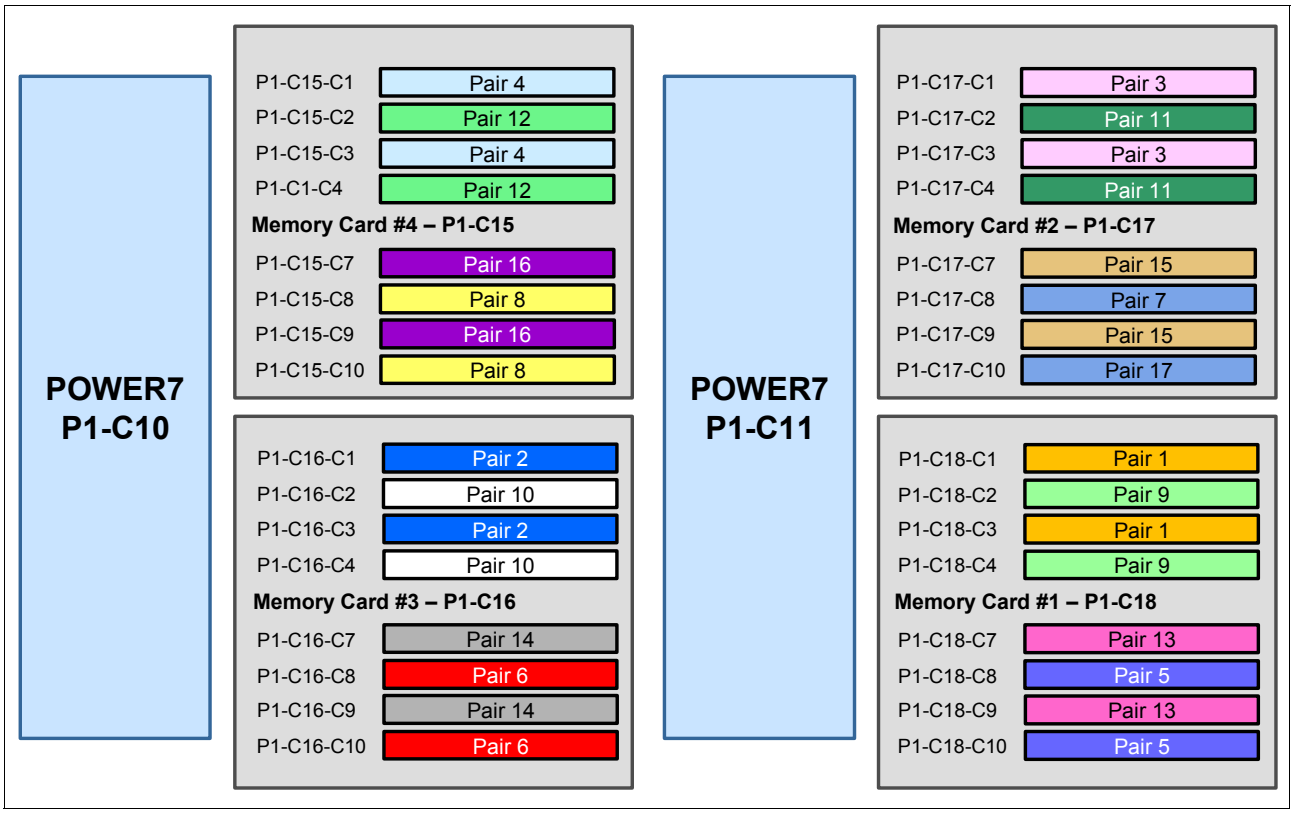


Figure 2-11 Memory DIMM installation sequence and slots

2.3.3 Memory bandwidth

The POWER7 processor has exceptional cache, memory, and interconnect bandwidths. Table 2-7 shows the bandwidth estimates for the Power 720. Table 2-8 covers the Power 740 system.

Table 2-7 Power 720 processor, memory, and I/O bandwidth estimates

Memory	3.00 GHz
L1 (data) cache	144 GBps
L2 cache	144 GBps
L3 cache	96 GBps
System memory	68.22 GBps

Table 2-8 Power 740 processor, memory, and I/O bandwidth estimates

Memory	3.55 GHz
L1 (data) cache	170.4 GBps
L2 cache	170.4 GBps
L3 cache	113.6 GBps
System memory	136.44 GBps total

2.4 Capacity on Demand

Capacity on Demand is not supported on the Power 720 and Power 740 systems.

2.5 Factory deconfiguration of processor cores

The Power 720 and Power 740 servers have the capability to be shipped with some installed cores deconfigured at the factory. The primary use for this feature is to assist with optimization of software licensing. A deconfigured core is unavailable for use in the system and thus does not require software licensing.

Feature #2319 deconfigures one core in a Power 720 or Power 740 system. It is available in a Power 720 with either 4-core, 6-core, or 8-core configurations and in a Power 740 with 12-core or 16-core configurations.

The maximum number of this feature that can be ordered is one less than the number of cores in the system. For example a maximum of 5 x #2319 can be ordered for a 6-core system. Feature #2319 can only be specified at initial order and cannot be applied to an installed machine.

Note: All processor cores *must* have activations ordered for them, even the ones being deconfigured.

2.6 System bus

This section provides additional information related to the internal buses.

The Power 720 and Power 740 systems have internal I/O connectivity through PCIe slots, as well as external connectivity through InfiniBand adapters.

The internal I/O subsystem on the Power 720 and Power 740 is connected to the GX bus on a POWER7 processor in the system. This bus runs at 2.5 GHz and provides 20 GBps of I/O connectivity to the PCIe slots, integrated Ethernet adapter ports, SAS internal adapters, and USB ports.

Additionally, the POWER7 processor chip installed on the Power 720 and each of the processor chips on the Power 740 provide a GX++ bus, which is used to optionally connect to a 12x GX++ adapter. Each bus runs at 2.5 GHz and provides 20 GBps bandwidth.

One GX++ slot is available on the Power 720 and two GX++ slots are available on the Power 740. The GX++ Dual-port 12x Channel Attach Adapter (#EJ0G) can be installed in either GX++ slot. The first GX++ slot can also be used by the optional PCIe Gen2 Adapter Riser Card feature (#5685) to add four short, 8x, PICe Gen2 low-profile slots.

Remember: The GX++ slots are not hot pluggable.

Table 2-9 provides I/O bandwidth of Power 720 and Power 740 processors configuration.

Table 2-9 I/O bandwidth

I/O	I/O Bandwidth (maximum theoretical)
GX++ Bus from the first POWER7 SCM to the IO chip	10 GBps simplex 20 GBps duplex
GX++ Bus (slot 1)	10 GBps simplex 20 GBps duplex
GX++ Bus (slot 2)	10 GBps simplex 20 GBps duplex
Total I/O bandwidth	30 GBps simplex 60 GBps duplex

2.7 Internal I/O subsystem

The internal I/O subsystem resides on the system planar that supports the PCIe slot. PCIe slots on the Power 720 and 740 are not hot pluggable. PCIe and PCI-X slots on the I/O drawers are hot-pluggable.

All PCIe slots support Enhanced Error Handling (EEH). PCI EEH-enabled adapters respond to a special data packet generated from the affected PCIe 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. For Linux, EEH support extends to the majority of frequently used devices, although various third-party PCI devices might not provide native EEH support.

An optional PCIe Adapter Riser Card (#5685) adds four short, 8x PCIe Gen2 low-profile slots and is installed in a GX++ slot 1. All PCIe slots are EEH, but they are not hot pluggable.

2.7.1 Slot configuration

Table 2-10 describes the slot configuration of the Power 720 and Power 740.

Table 2-10 Slot configuration of a Power 720 and Power 740

Slot number	Description	Location code	PCI Host Bridge (PHB)	Max. card size
Slot 1	PCIe Gen2 x8	P1-C2	P7IOC PCIe PHB5	Full height/short
Slot 2	PCIe Gen2 x8	P1-C3	P7IOC PCIe PHB4	Full height/short
Slot 3	PCIe Gen2 x8	P1-C4	P7IOC PCIe PHB3	Full height/short
Slot 4	PCIe Gen2 x8	P1-C5	P7IOC PCIe PHB2	Full height/short
Slot 5	PCIe Gen2 x8	P1-C6	P7IOC PCIe PHB1	Full height/short
Slot 6	PCIe Gen2 x4	P1-C7	P7IOC multiplexer PCIe PHB0	Full height/short
Slot 7	PCIe Gen2 x8	P1-C1-C1	P7IOC PCIe PHB1	Low profile/short
Slot 8	PCIe Gen2 x8	P1-C1-C2	P7IOC PCIe PHB4	Low profile/short
Slot 9	PCIe Gen2 x8	P1-C1-C3	P7IOC PCIe PHB2	Low profile/short
Slot 10	PCIe Gen2 x8	P1-C1-C4	P7IOC PCIe PHB3	Low profile/short

Remember: Full-height PCIe adapters and low-profile PCIe adapters are not interchangeable. Even if the card was designed with low-profile dimensions, the tail stock at the end of the adapter is specific to either low-profile or full-height PCIe slots.

2.7.2 System ports

The system planar has two serial ports that are called system ports. When an HMC is connected to the server, the integrated system ports of the server are rendered non-functional. In this case, you must install an asynchronous adapter, which is described in Table 2-21 on page 61 for serial port usage:

- Integrated system ports are not supported under AIX or Linux when the HMC ports are connected to an HMC. Either the HMC ports or the integrated system ports can be used, but not both.
- The integrated system ports are supported for modem and asynchronous terminal connections. Any other application using serial ports requires a serial port adapter to be installed in a PCI slot. The integrated system ports do not support IBM PowerHA® configurations.
- Configuration of the two integrated system ports, including basic port settings (baud rate, and so on), modem selection, call-home and call-in policy, can be performed with the Advanced Systems Management Interface (ASMI).

Remember: The integrated console/modem port usage just described is for systems configured as a single, system-wide partition. When it is configured with multiple partitions, the integrated console/modem ports are disabled because the TTY console and call-home functions are performed with the HMC.

2.8 PCI adapters

This section covers the different types and funtionalities of the PCI cards supported with the IBM Power 720 and Power 740 systems.

2.8.1 PCIe Gen1 and Gen2

Peripheral Component Interconnect Express (PCIe) uses a serial interface and allows for point-to-point interconnections between devices (using a directly wired interface between these connection points). A single PCIe serial link is a dual-simplex connection that uses two pairs of wires, one pair for transmit and one pair for receive, and can transmit only one bit per cycle. These two pairs of wires are called a *lane*. A PCIe link can consist of multiple lanes. In such configurations, the connection is labeled as x1, x2, x8, x12, x16, or x32, where the number is effectively the number of lanes.

Two generations of PCIe interface are supported in Power 720 and Power 740 models:

- Gen1: Capable of transmitting at the extremely high speed of 2.5 Gbps, which gives a capability of a peak bandwidth of 2 GBps simplex on an x8 interface
- Gen2: Double the speed of the Gen1 interface, which gives a capability of a peak bandwidth of 4 GBps simplex on an x8 interface

PCIe Gen1 slots support Gen1 adapter cards and also most of the Gen2 adapters. In this case, where a Gen2 adapter is used in a Gen1 slot, the adapter will operate at PCIe Gen1 speed. PCIe Gen2 slots support both Gen1 and Gen2 adapters. In this case, where a Gen1 card is installed into a Gen2 slot, it will operate at PCIe Gen1 speed with a slight performance enhancement. When a Gen2 adapter is installed into a Gen2 slot, it will operate at the full PCIe Gen2 speed.

The Power 720 and Power 740 system enclosure is equipped with five PCIe x8 Gen2 full-height slots. There is a sixth PCIe x4 slot dedicated to the PCIe Ethernet card that is standard with the base system. An optional PCIe Gen2 expansion feature is also available that provides an additional four PCIe x8 low profile slots.

IBM offers only PCIe low-profile adapter options for the Power 720 and Power 740 systems. All adapters support Extended Error Handling (EEH). PCIe adapters use a different type of slot than PCI and PCI-X adapters. If you attempt to force an adapter into the wrong type of slot, you might damage the adapter or the slot.

Note: IBM i IOP adapters are not supported in the Power 720 and Power 740 systems.

2.8.2 PCIe adapter form factors

IBM POWER7 processor-based servers are able to support two different form factors of PCIe adapters:

- PCIe low profile (LP) cards, which are used with the Power 710 and Power 730 PCIe slots. Low profile adapters are also used in the PCIe riser card slots of the Power 720 and Power 740 servers.
- PCIe full height and full high cards, which are plugged into the following server slots:
  - Power 720 and Power 740 (Within the base system five PCIe slots half length slots are supported.)
  - Power 750
  - Power 755
  - Power 770
  - Power 780
  - Power 795
  - PCIe slots of the #5802 and #5877 drawers

Low profile PCIe adapter cards are only supported in low profile PCIe slots, and full height cards are only supported in full height slots.

Figure 2-12 lists the PCIe adapter form factors.

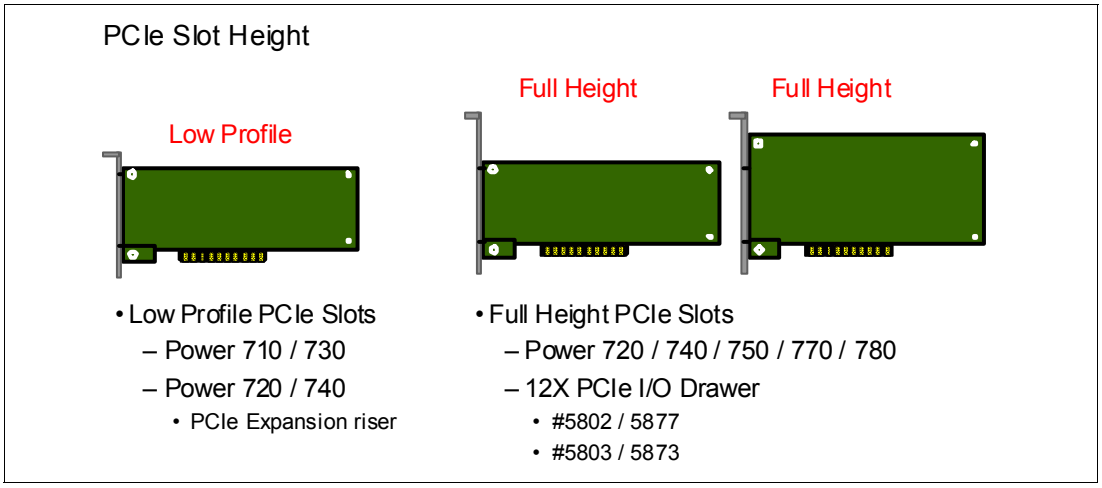


Figure 2-12 PCIe adapter form factors

Many of the full-height card features are available in low-profile format. For example, the #5273 8 Gb dual port Fibre channel adapter is the low-profile adapter equivalent of the #5735 adapter full height. They have equivalent functional characteristics.

Table 2-11 provides a list of low profile adapter cards and their equivalent in full height.

Table 2-11 Equivalent adapter cards

Low profile		Adapter description	Full height	
Feature code	**CCIN**		**Feature code**	**CCIN**
#2053	57CD	PCIe RAID and SSD SAS Adapter 3 Gb Low Profile	#2054 or #2055	57CD or 57CD
#5269		POWER GXT145 PCI Express Graphics Accelerator (LP)	#5748	5748
#5270		10 Gb FCoE PCIe Dual Port adapter (LP)	#5708	2BCB
#5271		4-Port 10/100/1000 Base-TX PCI-Express adapter	#5717	5717
#5272		10 Gigabit Ethernet-CX4 PCI Express adapter (LP)	#5732	5732
#5273		8 Gigabit PCI Express Dual Port Fibre Channel adapter (LP)	#5735	577D
#5274		2-Port Gigabit Ethernet-SX PCI Express adapter (LP)	#5768	5768
#5275		10 Gb ENet Fibre RNIC PCIe 8x	#5769	5769
#5276		4 Gigabit PCI Express Dual Port Fibre Channel adapter (LP)	#5774	5774
#5277		4-Port Async EIA-232 PCIe adapter (LP)	#5785	
#5278		SAS Controller PCIe 8x	#5901	57B3

Before adding or rearranging adapters, you can use the System Planning Tool to validate the new adapter configuration. See the System Planning Tool website:

http://www.ibm.com/systems/support/tools/systemplanningtool/

If you are installing a new feature, ensure that you have the software required to support the new feature and determine whether there are any existing update prerequisites to install. To do this, use the IBM Prerequisite website:

https://www-912.ibm.com/e_dir/eServerPreReq.nsf

The following sections discuss the supported adapters and provide tables of orderable feature numbers. The tables indicate operating support, AIX (A), IBM i (i), and Linux (L), for each of the adapters.

2.8.3 LAN adapters

Table 2-12 shows the local area network (LAN) adapters available for use with Power 720 and Power 740 systems. The adapters listed are supported in the base system PCIe slots, or in I/O enclosures that can be attached to the system using a 12X technology loop.

Table 2-12 Available LAN adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5271		PCIe LP 4-Port 10/100/1000 Base-TX Ethernet adapter	PCIe	Low profile Short	A, L
#5272		PCIe LP 10 GbE CX4 1-port adapter	PCIe	Low profile Short	A, L
#5274		PCIe LP 2-Port 1 GbE SX adapter	PCIe	Low profile Short	A, i, L
#5275		PCIe LP 10 GbE SR 1-port adapter	PCIe	Low profile Short	A, L
#5281		PCIe LP 2-Port 1 GbE TX adapter	PCIe	Low profile, Short	A, i, L
#5284		PCIe LP 2-Port 1 GbE TX adapter	PCIe	Low profile, Short	A, L
#5286		PCIe2 LP 2-Port 10 GbE SFP+ Copper adapter	PCIe	Low profile, Short	A, L
#5287		PCIe2 2-port 10 GbE SR adapter	PCIe	Low profile, Short	A, L
#5288		PCIe2 2-Port 10 GbE SFP+Copper adapter	PCIe	Full height Short	A, L
#5706	5706	IBM 2-Port 10/100/1000 Base-TX Ethernet PCI-X adapter	PCI-X	Full height Short	A, i, L
#5717	5717	4-Port 10/100/1000 Base-TX PCI Express adapter	PCIe	Full height Short	A, L
#5732	5732	10 Gigabit Ethernet-CX4 PCI Express adapter	PCIe	Full height Short	A, L
#5740		4-Port 10/100/1000 Base-TX PCI-X adapter	PCI-X	Full height Short	A, L
#5767	5767	2-Port 10/100/1000 Base-TX Ethernet PCI Express adapter	PCIe	Full height Short	A, i, L
#5768	5768	2-Port Gigabit Ethernet-SX PCI Express adapter	PCIe	Full height Short	A, i, L
#5769	5769	10 Gigabit Ethernet-SR PCI Express adapter	PCIe	Full height Short	A, L
#5772	576E	10 Gigabit Ethernet-LR PCI Express adapter	PCIe	Full height Short	A, i, L
#9055[a]		PCIe LP 2-Port 1GbE TX adapter	PCIe	Full height, Short	A, i, L

a. This adapter is required in the Power 720 and Power 740 systems.

Note: For IBM i OS, Table 2-12 on page 54 shows the native support of the card. All Ethernet cards can be supported by IBM i through the VIOS server.

2.8.4 Graphics accelerator adapters

Table 2-13 lists the available graphics accelerator adapters. They can be configured to operate in either 8-bit or 24-bit color modes. These adapters support both analog and digital monitors, and they are not hot pluggable.

Table 2-13 Available graphics accelerator adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5269	2849	PCIe LP POWER GXT145 Graphics Accelerator	PCIe	Low profile Short	A, L
#5748[a]		POWER GXT145 PCI Express Graphics Accelerator	PCIe	Full height Short	A, L

a. This card card is not supported in slot 6, P1-C7

2.8.5 SCSI and SAS adapters

To connect to external SCSI or SAS devices, the adapters listed in Table 2-14 are available.

Table 2-14 Available SCSI and SAS adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5278		PCIe LP 2-x4-port SAS adapter 3 Gb	PCIe	Low profile Short	A, i, L
#5736	571A	PCI-X DDR Dual Channel Ultra320 SCSI adapter	PCI-X	Full height Short	A, i, L
#5805[bc]	574E	PCIe 380 MB Cache Dual - x4 3 Gb SAS RAID adapter	PCIe	Full height Short	A, i, L
#5900[a]	572A	PCI-X DDR Dual -x4 SAS adapter	PCI-X	Full height Short	A, L
#5901[b]	57B3	PCIe Dual-x4 SAS adapter	PCIe	Full height Short	A, i, L
#5908	575C	PCI-X DDR 1.5 GB Cache SAS RAID adapter (BSC)	PCI-X	Full height Short	A, i, L
#5912	572A	PCI-X DDR Dual - x4 SAS adapter	PCI-X	Full height Short	A, i, L
#5913[c]	57B5	PCIe2 1.8 GB Cache RAID SAS adapter Tri-port 6 Gb	PCIe	Full height Short	A, i, L

a. Supported, but no longer orderable.
b. This card card is not supported in slot 6, P1-C7.
c. A pair of adapters is required to provide mirrored write cache data and adapter redundancy.

For detailed information about SAS cabling of external storage, see the IBM Power Systems Hardware Information Center:

http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp

Table 2-15 shows a comparison between parallel SCSI and SAS.

Table 2-15 Comparison parallel SCSI to SAS

Feature	Parallel SCSI	SAS
Architecture	Parallel, all devices connected to shared bus	Serial, point-to-point, discrete signal paths
Performance	320 MBps (Ultra320 SCSI), performance degrades as devices added to shared bus	3 Gbps, roadmap to 12 Gbps, performance maintained as more devices added
Scalability	15 drives	Over 16,000 drives
Compatibility	Incompatible with all other drive interfaces	Compatible with Serial ATA (SATA)
Max. cable length	12 meters total (must sum lengths of all cables used on bus)	8 meters per discrete connection, total domain cabling hundreds of meters
Cable from factor	Multitude of conductors adds bulk, cost	Compact connectors and cabling save space, cost
Hot pluggability	Yes	Yes
Device identification	Manually set, user must ensure no ID number conflicts on bus	Worldwide unique ID set at time of manufacture
Termination	Manually set, user must ensure proper installation and functionality of terminators	Discrete signal paths enable device to include termination by default

2.8.6 PCIe RAID and SSD SAS Adapter

A new SSD option for selected POWER7 processor-based servers offers a significant price/performance improvement for many client SSD configurations. The new SSD option is packaged differently from those currently available with Power Systems. The new PCIe RAID and SSD SAS adapter has up to four 177 GB SSD modules plugged directly onto the adapter, saving the need for the SAS bays and cabling associated with the current SSD offering. The new PCIe-based SSD offering can save up to 70% of the list price, and reduce up to 65% of the footprint, compared to disk enclosure based SSD, assuming equivalent capacity. This is dependant on the configuration required.

Figure 2-13 shows the double-wide adapter and SSD modules.

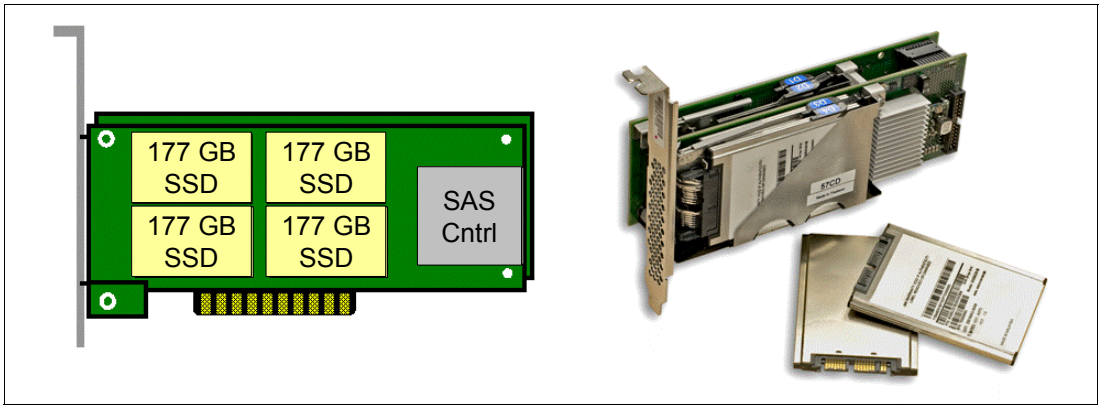


Figure 2-13 The PCIe RAID and SSD SAS Adapter and 177 GB SSD modules

To connect to external SCSI or SAS devices, the adapters listed in Table 2-14 on page 55 are available.

Table 2-16 Available PCIe RAID and SSD SAS adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#2053[a]	57CD	PCIe LP RAID and SSD SAS adapter 3 Gb	PCIe	Low profile Double wide, short	A, i, L
#2054	57CD	PCIe RAID and SSD SAS 3 Gb	PCIe	Double wide, short	A, i, L
#2055[b]	57CD	PCIe RAID and SSD SAS adapter 3 Gb with Blind Swap Cassette	PCIe	Full height inside a Blind Swap Cassette (BSC) Double wide, short	A, i, L

a. Only supported in the Rack-mount configuration. VIOS attachment requires Version 2.2 or later.
b. Only supported in a #5802/#5877 PCIe I/O drawer. Not supported in the Power 720 and Power 740 CEC. If used with the Virtual I/O server, the Virtual I/O server Version 2.2 or later is required.

Note: For a Power 720 tower configuration, it is possible to place PCIe-based SSDs in a #5802/#5877 PCIe I/O drawer.

The 177 GB SSD Module with Enterprise Multi-level Cell (eMLC) uses a new enterprise-class MLC flash technology, which provides enhanced durability, capacity, and performance. One, two, or four modules can be plugged onto a PCIe RAID and SSD SAS adapter, providing up to 708 GB of SSD capacity on one PCIe adapter.

Because the SSD modules are mounted on the adapter, to service either the adapter or one of the modules, the entire adapter must be removed from the system. Although the adapter can be hot plugged when installed in a #5802 or #5877 I/O drawer, removing the adapter also removes all SSD modules. So, to be able to hot plug the adapter and maintain data availability, two adapters must be installed and the data mirrored across the adapters.

Under AIX and Linux, the 177 GB modules can be reformatted as JBOD disks, providing 200 GB of available disk space. This removes RAID error correcting information, so it is best to mirror the data using operating system tools in order to prevent data loss in case of failure.

2.8.7 iSCSI

iSCSI adapters in Power Systems provide the advantage of increased bandwidth through hardware support of the iSCSI protocol. The 1 Gigabit iSCSI TOE (TCP/IP Offload Engine) PCI-X adapters support hardware encapsulation of SCSI commands and data into TCP, and transports them over the Ethernet using IP packets. The adapter operates as an iSCSI TOE. This offload function eliminates host protocol processing and reduces CPU interrupts. The adapter uses a small form factor LC type fiber optic connector or a copper RJ45 connector. Table 2-17 lists the orderable iSCSI adapters.

Table 2-17 Available iSCSI adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5713	573B	1 Gigabit iSCSI TOE PCI-X on Copper Media adapter	PCI-X	Full height Short	A, i, L

2.8.8 Fibre Channel adapters

The systems support direct or SAN connection to devices using Fibre Channel adapters. Table 2-18 provides a summary of the available Fibre Channel adapters.

All of these adapters except #5735 have LC connectors. If you are attaching a device or switch with an SC type fibre connector, an LC-SC 50 Micron Fiber Converter Cable (#2456) or an LC-SC 62.5 Micron Fiber Converter Cable (#2459) is required.

Table 2-18 Available Fibre Channel adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5273		PCIe LP 8 Gb 2-Port Fibre Channel adapter	PCIe	Low profile Short	A, i, L
#5276		PCIe LP 4 Gb 2-Port Fibre Channel adapter	PCIe	Low profile Short	A, i, L
#5735[a]	577D	8 Gigabit PCI Express Dual Port Fibre Channel adapter	PCIe	Full height Short	A, i, L
#5749	576B	4 Gbps Fibre Channel (2-Port) adapter	PCI-X	Full height Short	i
#5759	1910 5759	4 Gb Dual-Port Fibre Channel PCI-X 2.0 DDR adapter	PCI-X	Full height Short	A, L
#5774	5774	4 Gigabit PCI Express Dual Port Fibre Channel adapter	PCIe	Full height Short	A, i, L
#5729[b]		PCIe2 8 Gb 4-port Fibre Channel adapter	PCIe	Full height, Short	A, L[c]

a. At the time of writing the IBM i device driver does not support this card at PCIe slot 6, P1-C7.
b. A Gen2 PCIe slot is required to provide the bandwidth for all four ports to operate at full speed.
c. The usage within IBM i it is not supported. Instead, use it with the Virtual I/O server.

Note: The usage of NPIV through the Virtual I/O server requires 8 Gb Fibre Channel adapters such as the #5273, #5735, and #5729.

2.8.9 Fibre Channel over Ethernet

Fibre Channel over Ethernet (FCoE) allows for the convergence of Fibre Channel and Ethernet traffic onto a single adapter and converged fabric.

Figure 2-14 shows a comparison between existing FC and network connections and FCoE connections.

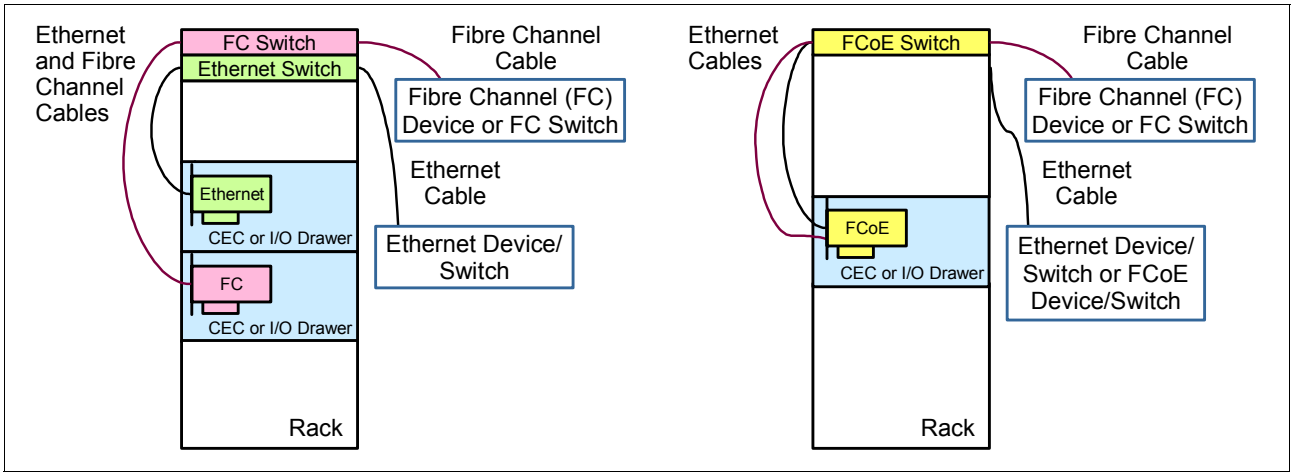


Figure 2-14 Comparison between existing FC and network connection and FCoE connection

Table 2-19 lists the available Fibre Channel over Ethernet Adapters. They are high-performance Converged Network Adapters (CNA) using SR optics. Each port can provide Network Interface Card (NIC) traffic and Fibre Channel functions simultaneously.

Table 2-19 Available FCoE adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5270		PCIe LP 10 Gb FCoE 2-port adapter	PCIe	Low profile Short	A, L
#5708		10 Gb FCoE PCIe Dual Port adapter	PCIe	Full height Short	A, L

For more information about FCoE, see *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493.

2.8.10 InfiniBand Host Channel adapter

The InfiniBand Architecture (IBA) is an industry-standard architecture for server I/O and inter-server communication. It was developed by the InfiniBand Trade Association (IBTA) to provide the levels of reliability, availability, performance, and scalability necessary for present and future server systems with levels significantly better than can be achieved using bus-oriented I/O structures.

InfiniBand (IB) is an open set of interconnect standards and specifications. The main IB specification has been published by the InfiniBand Trade Association and is available at:

http://www.infinibandta.org/

InfiniBand is based on a switched fabric architecture of serial point-to-point links, where these IB links can be connected to either host channel adapters (HCAs), used primarily in servers, or target channel adapters (TCAs), used primarily in storage subsystems.

The InfiniBand physical connection consists of multiple byte lanes. Each individual byte lane is a four-wire, 2.5, 5.0, or 10.0 Gbps bi-directional connection. Combinations of link width and byte lane speed allow for overall link speeds from 2.5 Gbps to 120 Gbps. The architecture defines a layered hardware protocol and a software layer to manage initialization and the communication between devices. Each link can support multiple transport services for reliability and multiple prioritized virtual communication channels.

IBM offers the GX++ 12X DDR Adapter (#EJ04) that plugs into the system backplane (GX++ slot). One GX++ slot is available on the Power 720. One or two GX++ slots are available on the Power 740, if used with one or two processor cards. Detailed information can be found in 2.6, “System bus” on page 49.

By attaching a 12X to 4X converter cable (#1828, #1841,or #1842) to #EJ04, a supported IB switch can be attached. AIX, IBM i, and Linux operating systems are supported.

A new PCIe Gen2 LP 2-Port 4X InfiniBand quad data rates (QDR) Adapter 40 Gb (#5283) is available. The PCIe Gen2 low-profile adapter provides two high-speed 4X InfiniBand connections for IP over IB usage in the Power 720 and Power 740. On the Power 720 and Power 740, this adapter is supported in PCIe Gen2 slots. The following types of QDR IB cables are provided for attachment to the QDR adapter and its QSFP(Quad Small Form-Factor Pluggable) connectors:

- Copper cables provide 1-meter, 3-meter, and 5-meter lengths (#3287, #3288, and 3289).
- Optical cables provide 10-meter and 30-meter lengths (#3290 and #3293). These are QSFP/QSFP cables that also attach to QSFP ports on the switch.

The feature #5283 QDR adapter attaches to the QLogic QDR switches. These switches can be ordered from IBM using the following machine type and model numbering:

- 7874-036 is a QLogic 12200 36-port, 40 Gbps InfiniBand Switch that cost-effectively links workgroup resources into a cluster.
- 7874-072 is a QLogic 12800-040 72-port, 40 Gbps InfiniBand switch that links resources using a scalable, low-latency fabric, supporting up to four 18-port QDR Leaf Modules.
- 7874-324 is a QLogic 12800-180 324-port 40 Gbps InfiniBand switch designed to maintain larger clusters, supporting up to eighteen 18-port QDR Leaf Modules.

Note: The feature #5283 adapter has two 40 Gb ports ,and a PCIe Gen2 slot has the bandwidth to support one port. This means that the benefit of two ports will be for redundancy rather than additional performance.

Table 2-20 lists the available InfiniBand adapters.

Table 2-20 Available InfiniBand adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5283		PCIe2 LP 2-Port 4X IB QDR adapter 40 Gb	PCIe	Low profile Short	A, L
#5285[a]		2-Port 4X IB QDR adapter 40 Gb	PCIe	full high slot	A, L
#EJ04		GX++ Dual-port 12x Channel Attach adapter	GX++		A, L

a. Requires PCIe Gen2 full high slot

For more information about Infiniband, see *HPC Clusters Using InfiniBand on IBM Power Systems Servers*, SG24-7767.

2.8.11 Asynchronous adapters

Asynchronous PCI adapters provide connection of asynchronous EIA-232 or RS-422 devices. If you have a cluster configuration or high-availability configuration and plan to connect the IBM Power Systems using a serial connection, you can use the features listed in Table 2-21.

Table 2-21 Available asynchronous adapters

Feature code	CCIN	Adapter description	Slot	Size	OS support
#5277		PCIe LP 4-Port Async EIA-232 adapter	PCIe	Low profile Short	A, L
#5785[a]		4 Port Async EIA-232 PCIe adapter	PCIe	Full height Short	A, L
#5289		Port Async EIA-232 PCIe adapter	PCIe	Full height, Short	A, L
#5290		PCIe LP 2-Port Async EIA-232 adapter	PCIe	Low profile Short	A, L

a. This card is not supported in slot 6, P1-C7.

2.9 Internal storage

The Power 720 and Power 740 servers use an integrated SAS/SATA controller connected through a PCIe bus to the P7-IOC chip (Figure 2-15 on page 62). The SAS/SATA controller used in the server's enclosure has two sets of four SAS/SATA channels, which give Power 720 and Power 740 the combined total of eight SAS busses. Each channel can support either SAS or SATA operation. The SAS controller is connected to a DASD backplane and supports three or six small form factor (SFF) disk drive bays depending on the backplane option.

One of the following options must be selected as the backplane:

- Feature #5618 provides a backplane that supports up to six SFF SAS HDDs/SSDs, a SATA DVD, and a half-high tape drive for either a tape drive or USB removable disk. This feature does not provide RAID 5, RAID 6, a write cache, or an external SAS port. Split backplane functionality (3x3) is supported with the additional feature #EJ02.

  Remember: Note the following information:

  - No additional PCIe SAS adapter is required for split backplane functionality.
  - Feature #5618 is not supported with IBM i.

- Feature #EJ01 is a higher-function backplane that supports up to eight SFF SAS HDDs/SSDs, a SATA DVD, a half-high tape drive for either a tape drive or USB removable disk, Dual 175 MB Write Cache RAID, and one external SAS port. The #EJ01 supports RAID 5 and RAID 6, and there is no split backplane available for this feature.

  Remember: Feature #EJ01 is required by IBM i to natively use the internal storage (HDDs/SSDs and media) and the system SAS port.

Figure 2-15 details an internal topology overview for the #5618 backplane.

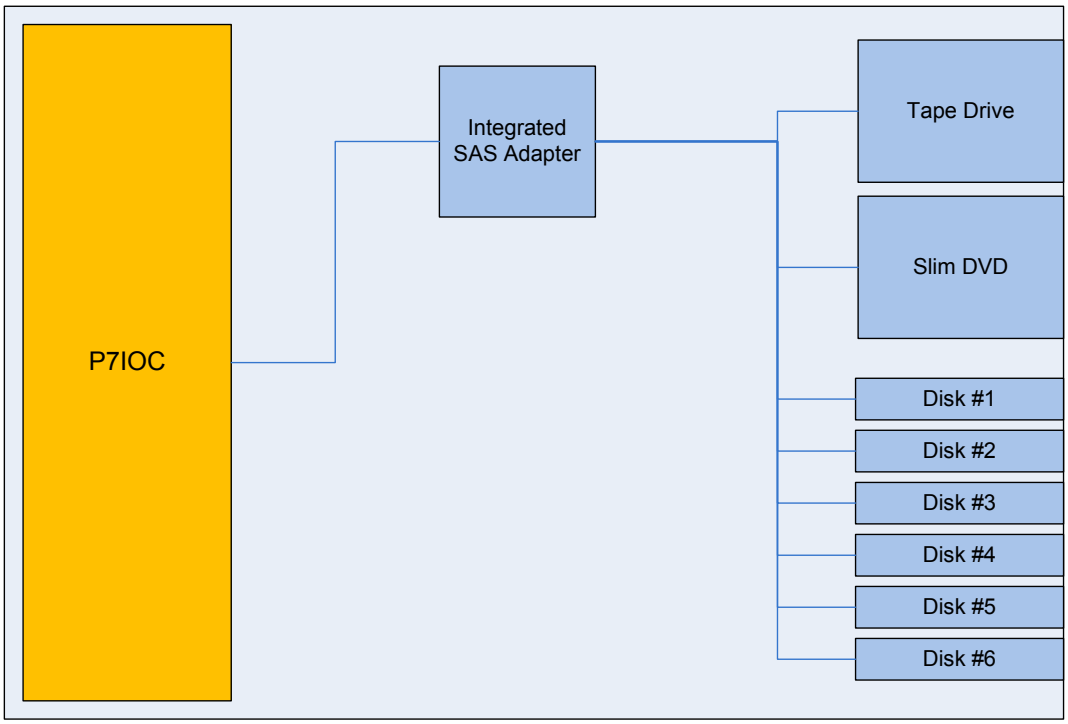


Figure 2-15 Internal topology overview for #5618 DASD backplane

Figure 2-16 shows an internal topology overview for the #EJ01 backplane:

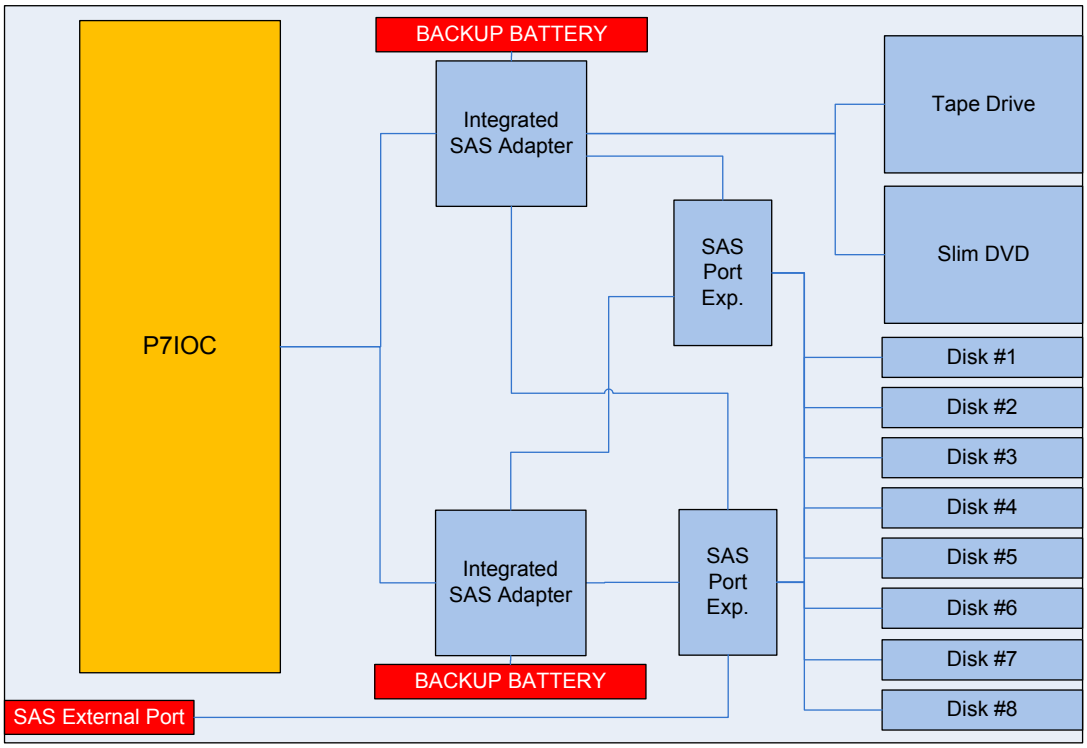


Figure 2-16 #EJ01 DASD backplane - Internal topology overview

2.9.1 RAID support

There are multiple protection options for HDD/SSD drives in the SAS SFF bays in the Power 720 and 740 system unit or drives in 12X attached I/O drawers or drives in disk-only I/O drawers. Although protecting drives is always the best idea, AIX/Linux users can choose to leave a few or all drives unprotected at their own risk, and IBM supports these configurations. IBM i configuration rules differ in this regard, and IBM supports IBM i partition configurations only when HDD/SSD drives are protected.

Drive protection

HDD/SSD drive protection can be provided by AIX, IBM i, and Linux software or by the HDD/SSD hardware controllers. Mirroring of drives is provided by AIX, IBM i, and Linux software. In addition, AIX/Linux supports controllers providing RAID 0, 1, 5, 6, or 10. IBM i integrated storage management already provides striping, so IBM i also supports controllers providing RAID 5 or 6. To further augment HDD/SSD protection, hot spare capability can be used for protected drives. Specific hot spare prerequisites apply.

An integrated SAS controller offering RAID 0, 1, and 10 support is provided in the Power 720 and Power 740 system unit.

It can be optionally augmented by RAID 5 and RAID 6 capability when storage backplane #EJ01 is added to the configuration. In addition to these protection options, mirroring of drives by the operating system is supported. AIX or Linux supports all of these options. IBM i does not use JBOD, and uses imbedded functions instead of RAID 10, but does leverage the RAID 5 or 6 function of the integrated controllers.

Other disk/SSD controllers are provided as PCIe SAS adapters are supported. Different PCI controllers with and without write cache are supported. Also, RAID 5 and 6 on controllers with write cache are supported with one exception: The PCIe RAID and SSD SAS adapter has no write cache, but it supports RAID 5 and RAID 6.

Table 2-22 list the RAID support by backplane.

Table 2-22 RAID configurations for the Power 720 and Power 740

Feature code	Split cackplane	JBOD	RAID 0, 1, and 10	RAID 5 and 6	External SAS Port
#5618	No	Yes	Yes	No	No
#5618 and #EJ02	Yes	Yes	Yes	No	No
#EJ01	No	No	Yes	Yes	Yes

AIX and Linux can use disk drives formatted with 512-byte blocks when being mirrored by the operating system. These disk drives must be reformatted to 528-byte sectors when used in RAID arrays. Although a small percentage of the drive's capacity is lost, additional data protection such as ECC and bad block detection is gained in this reformatting. For example, a 300 GB disk drive, when reformatted, provides around 283 GB. IBM i always uses drives formatted to 528 bytes. Solid state disks are formatted to 528 bytes.

Power 720 and Power 740 support a dual write cache RAID feature that consists of an auxiliary write cache for the RAID card and the optional RAID enablement.

Supported RAID functions

Base hardware supports RAID 0,1, and 10. When additional features are configured, Power 720 and Power 740 support hardware RAID 0, 1, 5, 6, and 10:

- RAID-0 provides striping for performance, but does not offer any fault tolerance.

  The failure of a single drive results in the loss of all data on the array. This increases I/O bandwidth by simultaneously accessing multiple data paths.
- RAID-1 mirrors the contents of the disks, making a form of 1:1 ratio realtime backup. The contents of each disk in the array are identical to that of every other disk in the array.
- RAID-5 uses block-level data striping with distributed parity.

  RAID-5 stripes both data and parity information across three or more drives. Fault tolerance is maintained by ensuring that the parity information for any given block of data is placed on a drive separate from those used to store the data itself.
- RAID-6 uses block-level data striping with dual distributed parity.

  RAID-6 is the same as RAID-5 except that it uses a second level of independently calculated and distributed parity information for additional fault tolerance. RAID-6 configuration requires N+2 drives to accommodate the additional parity data, which makes it less cost effective than RAID-5 for equivalent storage capacity.
- RAID-10 is also known as a striped set of mirrored arrays.

  It is a combination of RAID-0 and RAID-1. A RAID-0 stripe set of the data is created across a 2-disk array for performance benefits. A duplicate of the first stripe set is then mirrored on another 2-disk array for fault tolerance.

2.9.2 External SAS port and split backplane

This section describes the external SAS port and split backplane features.

External SAS port feature

The Power 720 and Power 740 DASD backplane (#EJ01) offers an external SAS port:

- The SAS port connector is located next to the GX++ slot 2 on the rear bulkhead.
- The external SAS port can be used to connect external SAS devices (#5886 or #5887), the IBM System Storage 7214 Tape and DVD Enclosure Express (Model 1U2), or the IBM System Storage 7216 Multi-Media Enclosure (Model 1U2).

Note: Only one SAS drawer is supported from the external SAS port. Additional SAS drawers can be supported by SAS adapters. SSDs are not supported on the SAS drawer connected to the external port.

Split DASD backplane feature

The Power 720 and Power 740 DASD backplane (#5618) supports split drive bay (#EJ02). If #EJ02 is configured, then the six Small Form Factors (SFFs) slots are split into a pair of three drive bay groups (3x3).

Note: In a Power 720 and Power 740 with a split backplane, SSDs and HDDs can be placed in a set of three disks, but no mixing of SSDs and HDDs within a split configuration is allowed. IBM i does not support split backplane.

Figure 2-17 details the split backplane.

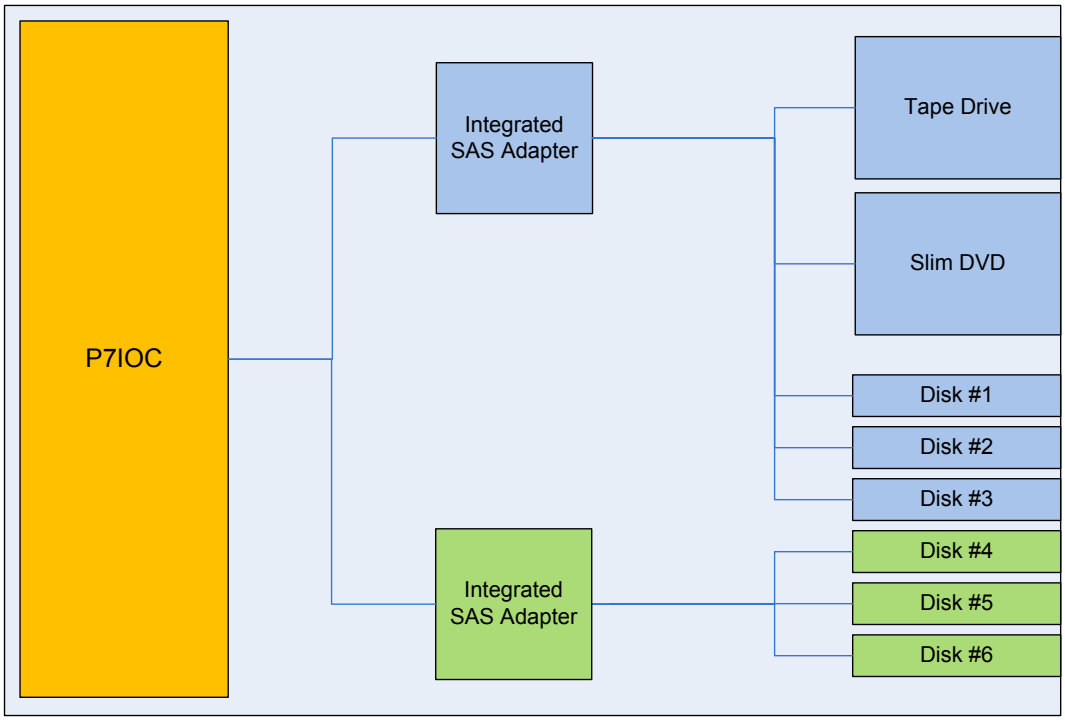


Figure 2-17 Internal topology overview for #5618 DASD backplane with split backplane feature #EJ01

2.9.3 Media bays

The Power 720 and Power 740 each offer a slim media bay to support a slim SATA DVD device. Direct dock and hot-plug of the DVD media device are supported. Also, the half-high bay is available to support an optional SAS tape drive or removable disk drive.

2.10 External I/O subsystems

This section describes the external 12X I/O subsystems that can be attached to the Power 720 and Power 740, listed as follows:

- PCI-DDR 12X Expansion Drawer (#5796) (supported but not orderable)
- 12X I/O Drawer PCIe, SFF disk (#5802)
- 12X I/O Drawer PCIe, no disk (#5877)

Table 2-23 provides an overview of the capabilities of the supported I/O drawers.

Table 2-23 I/O drawer capabilities

Feature code	DASD	PCI slots	Requirements for a 720/740
#5796	None	6 x PCI-X	GX++ Dual-port 12x Channel Attach (#EJ04) and interface attach #6446/#6457
#5802	18 x SFF disk drive bays	10 x PCIe	GX++ Dual-port 12x Channel Attach (#EJ04)
#5877	None	10 x PCIe	GX++ Dual-port 12x Channel Attach (#EJ04)
#5887	24 x SAS disk drive bays	None	SAS PCI-X or PCIe adapter
#5886	12 x SAS disk drive bays	None	SAS PCI-X or PCIe adapter

Each processor card feeds one GX++ adapter slot. On the Power 720, there can be one GX++ slot available, and on the Power 740, there can be one or two, depending on whether one or two processor modules are installed.

Note: The attachment of external I/O drawers is not supported on the 4-core Power 720.

2.10.1 PCI-DDR 12X Expansion Drawer

The PCI-DDR 12X Expansion Drawer (#5796) is a 4U (EIA units) drawer and mounts in a 19-inch rack. Feature #5796 is 224 mm (8.8 in.) wide and takes up half the width of the 4U (EIA units) rack space. The 4U enclosure can hold up to two #5796 drawers mounted side-by-side in the enclosure. The drawer is 800 mm (31.5 in.) deep and can weigh up to 20 kg (44 lb).

The PCI-DDR 12X Expansion Drawer has six 64-bit, 3.3 V, PCI-X DDR slots running at 266 MHz that use blind-swap cassettes and support hot-plugging of adapter cards. The drawer includes redundant hot-plug power and cooling.

Two interface adapters are available for use in the #5796 drawer:

- Dual-Port 12X Channel Attach Adapter Long Run (#6457)
- Dual-Port 12X Channel Attach Adapter Short Run (#6446)

The adapter selection is based on how close the host system or the next I/O drawer in the loop is physically located. Feature #5796 attaches to a host system CEC enclosure with a 12X adapter in a GX++ slot through SDR or DDR cables (or both SDR and DDR cables). A maximum of four #5796 drawers can be placed on the same 12X loop. Mixing #5802/5877 and #5796 on the same loop is not supported.

A minimum configuration of two 12X cables (either SDR or DDR), two ac power cables, and two SPCN cables is required to ensure proper redundancy.

Figure 2-18 shows the back view of the expansion unit.

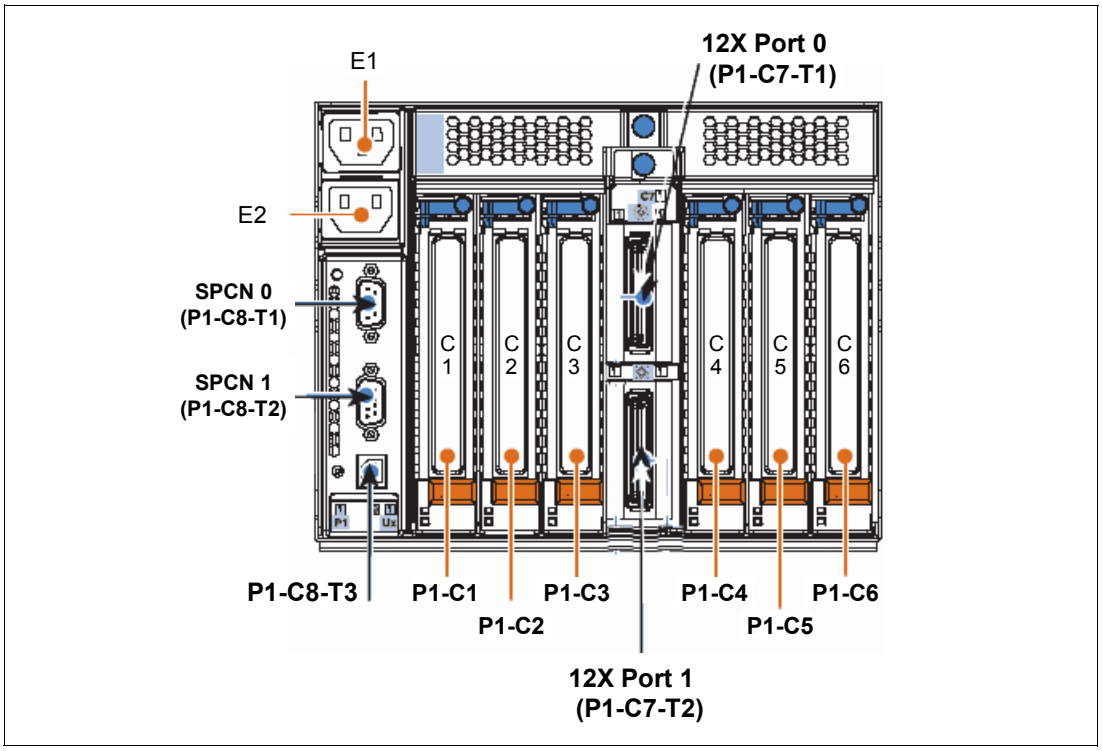


Figure 2-18 PCI-X DDR 12X Expansion Drawer rear side

2.10.2 12X I/O Drawer PCIe

The 12X I/O Drawer PCIe (#5802) is a 19-inch I/O and storage drawer. It provides a 4U-tall (EIA units) drawer containing 10 PCIe-based I/O adapter slots and 18 SAS hot-swap Small Form Factor disk bays, which can be used for either disk drives or SSD. The adapter slots use blind-swap cassettes and support hot-plugging of adapter cards.

A maximum of two #5802 drawers can be placed on the same 12X loop. Feature #5877 is the same as #5802 except that it does not support any disk bays. Feature #5877 can be on the same loop as #5802. Feature #5877 cannot be upgraded to #5802.

The physical dimensions of the drawer measure 444.5 mm (17.5 in.) wide by 177.8 mm (7.0 in.) high by 711.2 mm (28.0 in.) deep for use in a 19-inch rack.

A minimum configuration of two 12X DDR cables, two ac power cables, and two SPCN cables is required to ensure proper redundancy. The drawer attaches to the host CEC enclosure with a 12X adapter in a GX++ slot through 12X DDR cables that are available in various cable lengths:

- 0.6 (#1861)
- 1.5 (#1862)
- 3.0 (#1865)
- 8 meters (#1864)

The 12X SDR cables are not supported.

Figure 2-19 shows the front view of the 12X I/O Drawer PCIe (#5802).

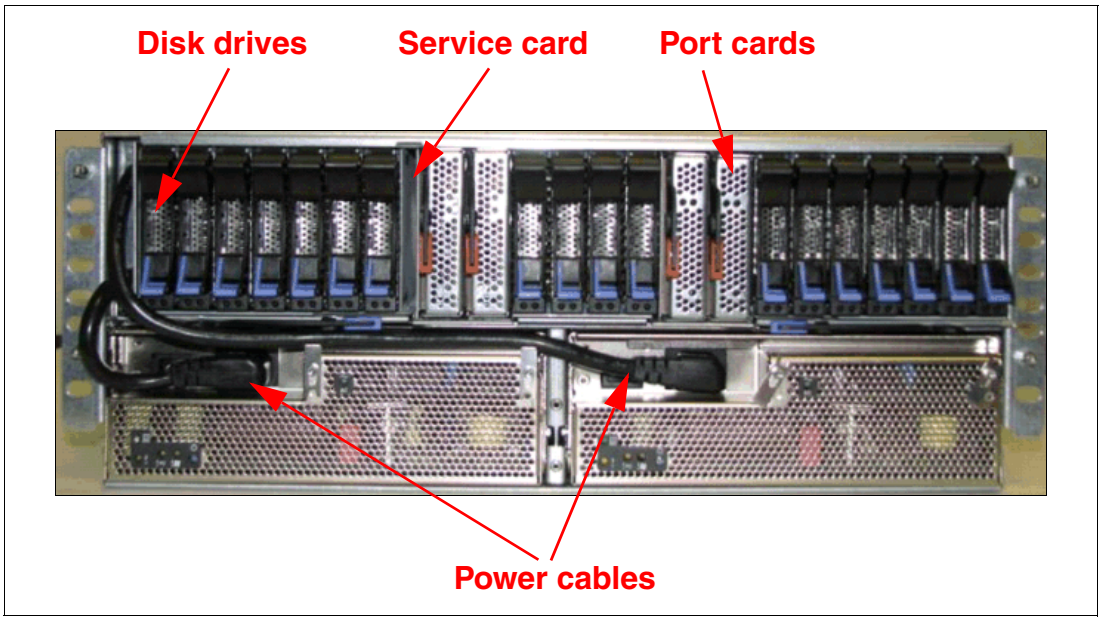


Figure 2-19 Front view of the 12X I/O Drawer PCIe

Figure 2-20 shows the rear view of the 12X I/O Drawer PCIe (#5802).

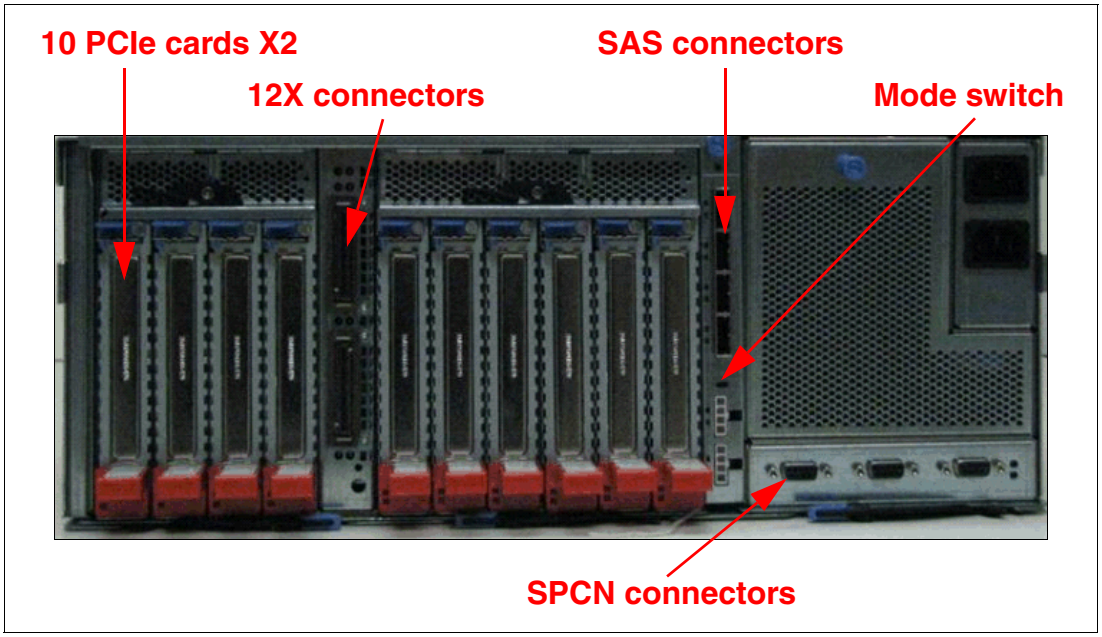


Figure 2-20 Rear view of the 12X I/O Drawer PCIe

2.10.3 Dividing SFF drive bays in a 12X I/O drawer PCIe

Disk drive bays in a 12X I/O drawer PCIe can be configured as a set of one, two, or four, allowing for partitioning of disk bays. Disk bay partitioning configuration can be done by physical mode switch on the I/O drawer.

Note: Mode change using the physical mode switch requires power-off/on of the drawer.

Figure 2-20 on page 68 indicates the Mode Switch in the rear view of the #5802 I/O Drawer.

Each disk bay set can be attached to its own controller or adapter. The #5802 PCIe 12X I/O Drawer has four SAS connections to drive bays. It connects to PCIe SAS adapters or controllers on the host system.

Figure 2-21 shows the configuration rule of disk bay partitioning in the #5802 PCIe 12X I/O Drawer. There is no specific feature code for mode switch setting.

Note: The IBM System Planing Tool supports disk bay partitioning. Also, the IBM configuration tool accepts this configuration from the IBM System Planing Tool and passes it through IBM manufacturing using the Customer Specified Placement (CSP) option.

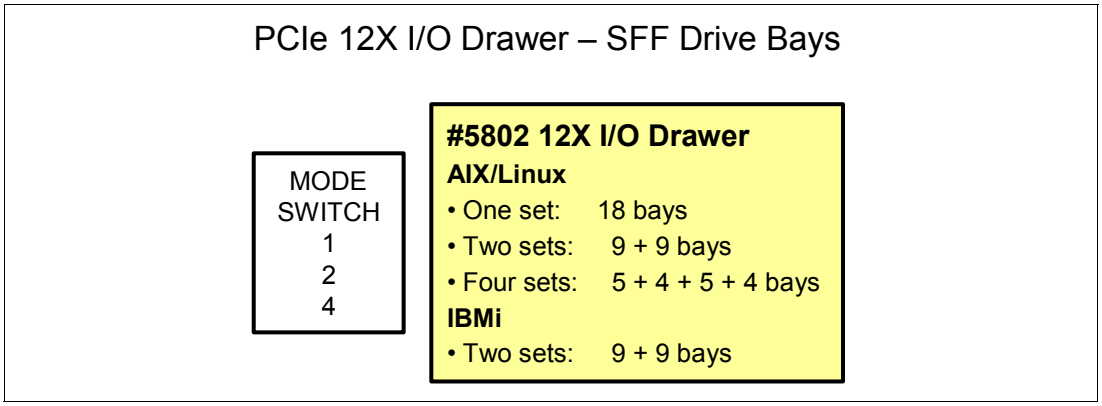


Figure 2-21 Disk bay partitioning in #5802 PCIe 12X I/O drawer

The SAS ports, as associated with the mode selector switch map to the disk bays, have the mappings shown in Table 2-24.

Table 2-24 SAS connection mappings

Location code	Mappings	Number of bays
P4-T1	P3-D1 to P3-D5	5 bays
P4-T2	P3-D6 to P3-D9	4 bays
P4-T3	P3-D10 to P3-D14	5 bays
P4-T3	P3-D15 to P3-D18	4 bays

Location codes for #5802 I/O drawer

Figure 2-22 and Figure 2-23 provide the location codes for the front and rear views of the #5802 I/O drawer.

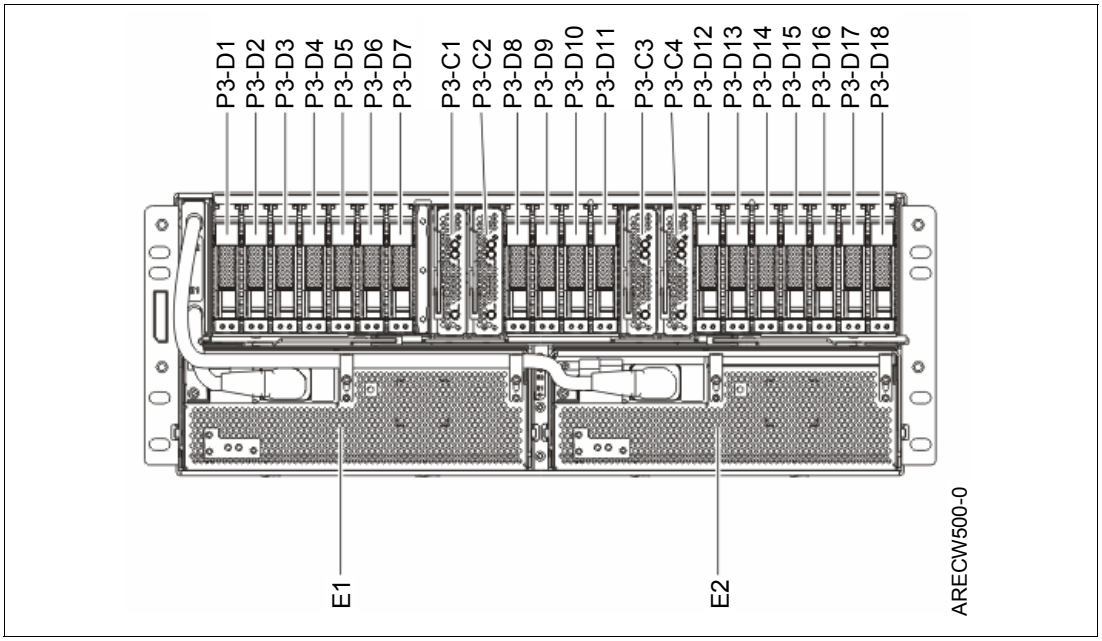


Figure 2-22 5802 I/O drawer from view location codes

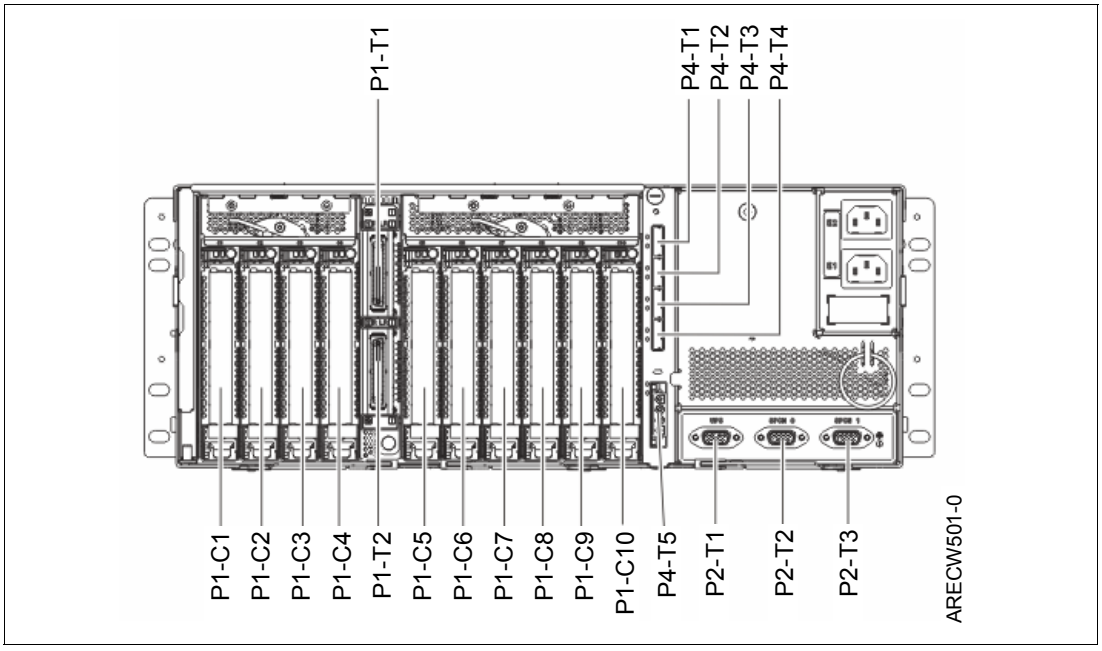


Figure 2-23 5802 I/O drawer rear view location codes

Configuring the #5802 disk drive subsystem

The #5802 SAS disk drive enclosure can hold up 18 disk drives. The disks in this enclosure can be organized in various configurations depending on the operating system used, the type of SAS adapter card, and the position of the mode switch.

Each disk bay set can be attached to its own controller or adapter. The Feature #5802 PCIe 12X I/O Drawer has four SAS connections to drive bays. It connects to PCIe SAS adapters or controllers on the host systems.

For detailed information about how to configure this, see the IBM Power Systems Hardware Information Center at:

http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp

2.10.4 12X I/O drawer PCIe and PCI-DDR 12X Expansion Drawer 12X cabling

I/O drawers are connected to the adapters in the CEC enclosure with data transfer cables:

- 12X DDR cables for the #5802 and #5877 I/O drawers
- 12X SDR and/or DDR cables for the #5796 I/O drawers

The first 12X I/O Drawer that is attached in any I/O drawer loop requires two data transfer cables. Each additional drawer, up to the maximum allowed in the loop, requires one additional data transfer cable. Note the following information:

- A 12X I/O loop starts at a CEC bus adapter port 0 and attaches to port 0 of an I/O drawer.
- The I/O drawer attaches from port 1 of the current unit to port 0 of the next I/O drawer.
- Port 1 of the last I/O drawer on the 12X I/O loop connects to port 1 of the same CEC bus adapter to complete the loop.

Figure 2-24 shows typical 12X I/O loop port connections.

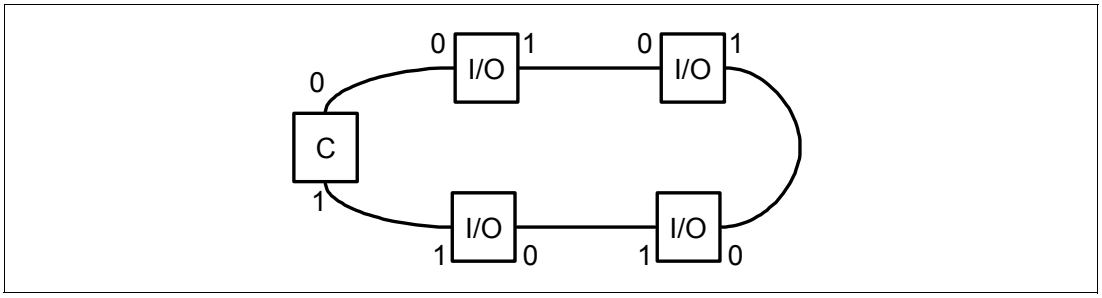


Figure 2-24 Typical 12X I/O loop port connections

Table 2-25 shows various 12X cables to satisfy the various length requirements.

Table 2-25 12X connection cables

Feature code	Description
#1861	0.6 meter 12X DDR cable
#1862	1.5 meter 12X DDR cable
#1865	3.0 meter 12X DDR cable
#1864	8.0 meter 12X DDR cable

General rules for 12X IO Drawer configuration

If you have two processor cards, spread the I/O drawers across two busses for better performance. Figure 2-25 shows the configuration to attach 12X IO Drawers to a Power 720.

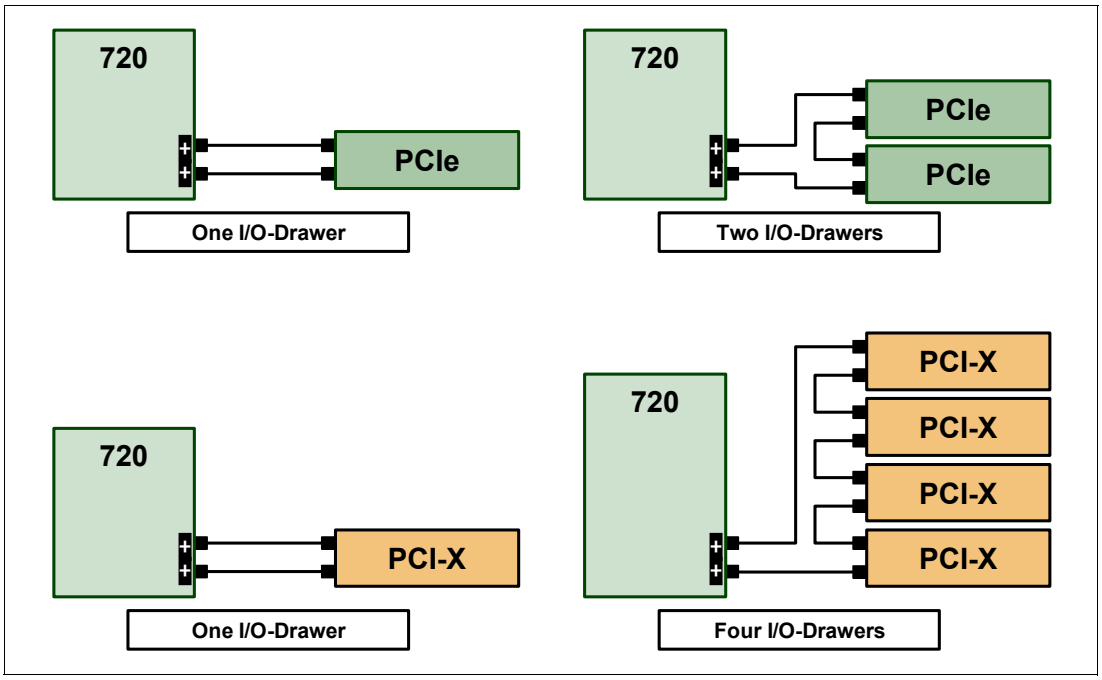


Figure 2-25 12X I/O Drawer configuration for a Power 720 with one GX++ slot

The configuration rules are the same for the Power 740. However, because the Power 740 can have up to two GX++ slots, you have various options available to attach 12X IO drawers. Figure 2-26 shows four of them, but there are more options available.

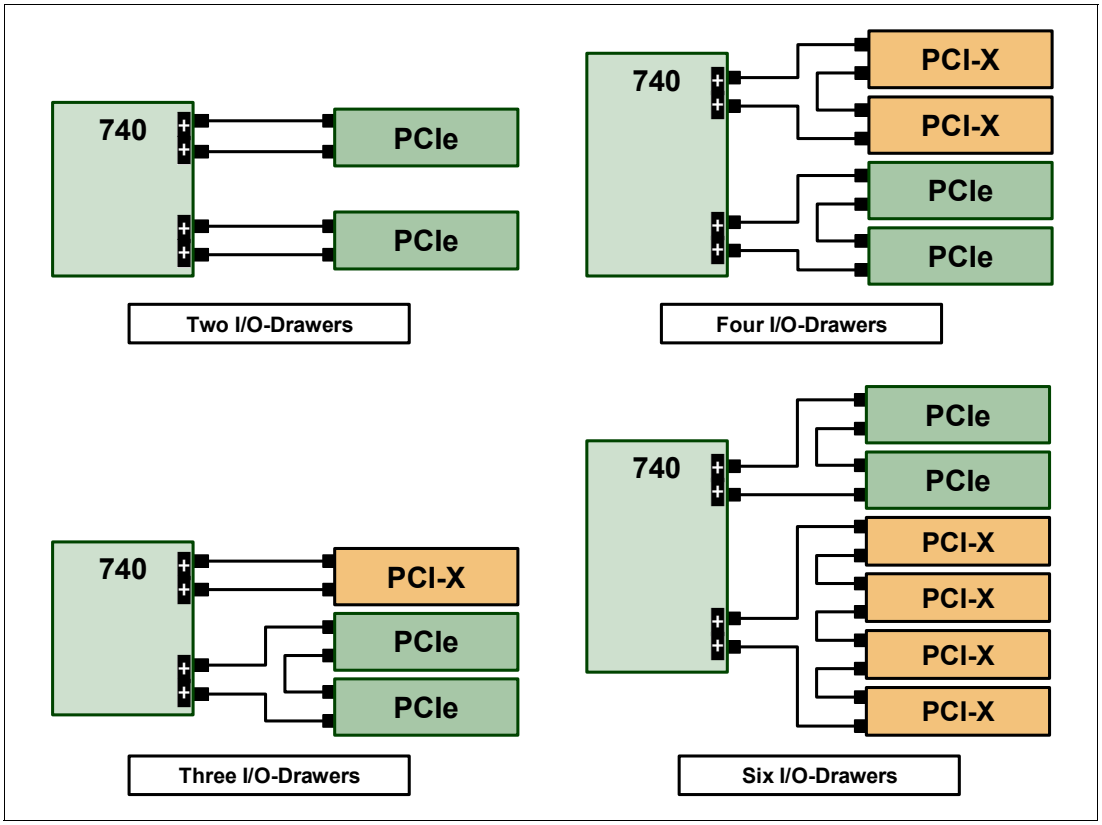


Figure 2-26 12X I/O Drawer configuration for a Power 740 with one GX++ slot

Supported 12X cable length for PCI-DDR 12X Expansion Drawer

Each #5796 drawer requires one Dual Port PCI DDR 12X Channel Adapter, either Short Run (#6446) or Long Run (#6457). The choice of adapters is dependent on the distance to the next 12X Channel connection in the loop, either to another I/O drawer or the system unit. Table 2-26 identifies the supported cable lengths for each 12X channel adapter. I/O drawers containing the Short Range adapter can be mixed in a single loop with I/O drawers containing the Long Range adapter. In Table 2-26, a “Yes” indicates that the 12X cable identified in that column can be used to connect the drawer configuration identified to the left. A “No” means that it cannot be used.

Table 2-26 Supported 12X cable length

Connection type	12X cable options			
	0.6 M	1.5 M	3.0 M	8.0 M
#5796 to #5796 with #6446 in both drawers	Yes	Yes	No	No
#5796 with #6446 adapter to #5796 with #6457 adapter	Yes	Yes	Yes	No
#5796 to #5796 with 6#457 adapter in both drawers	Yes	Yes	Yes	Yes
#5796 with #6446 adapter to system unit	No	Yes	Yes	No
#5796 with #6457 adapter to system unit	No	Yes	Yes	No

2.10.5 12X I/O Drawer PCIe and PCI-DDR 12X Expansion Drawer SPCN cabling

The System Power Control Network (SPCN) is used to control and monitor the status of power and cooling within the I/O drawer.

SPCN cables connect all ac powered expansion units (Figure 2-27):

1. Start at SPCN 0 (T1) of the CEC unit to J15 (T1) of the first expansion unit.
2. Cable all units from J16 (T2) of the previous unit to J15 (T1) of the next unit.
3. To complete the cabling loop, from J16 (T2) of the final expansion unit, connect to the CEC, SPCN 1 (T2).
4. Ensure that a complete loop exists from the CEC, through all attached expansions and back to the CEC drawer.

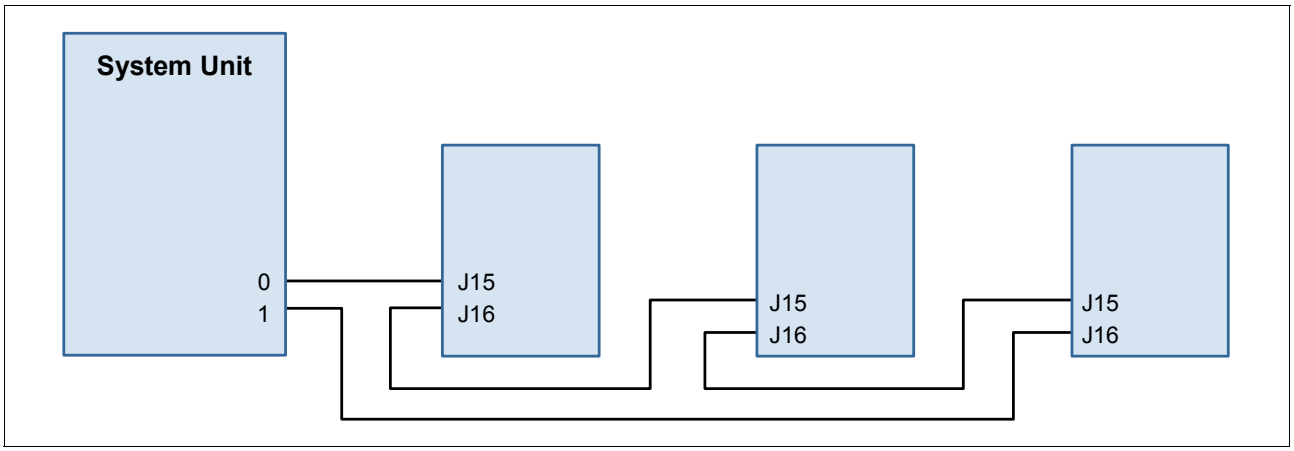


Figure 2-27 SPCN cabling examples

Table 2-27 shows SPCN cables to satisfy various length requirements.

Table 2-27 SPCN cables

Feature code	Description
#6006	SPCN cable drawer-to-drawer, 2 m
#6008[a]	SPCN cable rack-to-rack, 6 m
#6007	SPCN cable rack-to-rack, 15 m
#6029[a]	SPCN cable rack-to-rack, 30 m

a. Supported, but no longer orderable

2.11 External disk subsystems

This section describes the external disk subsystems that can be attached to the Power 720 and Power 740:

- EXP 12S SAS Expansion Drawer (#5886, supported, but no longer orderable)
- EXP24S SFF Gen2-bay Drawer for high-density storage (#5887)
- TotalStorage EXP24 Disk Drawer (#5786) and Tower (#5787)

- IBM 7031 TotalStorage EXP24 Ultra320 SCSI Expandable Storage Disk Enclosure (no longer orderable)
- IBM System Storage

The following sections describe the EXP 12S Expansion Drawer, the EXP24S SFF Gen2-bay Drawer, the 12X I/O Drawer PCIe with SFF Disks drawer, and IBM System Storage in more detail.

2.11.1 EXP 12S SAS Expansion Drawer

The EXP 12S SAS Expansion Drawer (#5886) is an expansion drawer that supports up to 12 hot-swap SAS Hard Disk Drives (HDD) or up to eight hot-swap Solid State Drives (SSD). The EXP 12S includes redundant ac power supplies and two power cords. Though the drawer is one set of 12 drives that is run by one SAS controller or one pair of SAS controllers, it has two SAS attachment ports and two Service Managers for redundancy. The EXP 12S takes up a 2-EIA space in a 19-inch rack. The SAS controller can be a SAS PCI-X or PCIe adapter or pair of adapters.

At the time of writing, a single disk-only EXP 12S SAS Expansion Drawer (#5886) is supported from the external SAS port of a Power 720 and Power 740 server. Additional EXP 12S Drawers can be supported using PCI-X or PCIe SAS adapters. A maximum of 28 EXP 12S Drawers are supported on a Power 720 and Power 740 server.

Note: The 4-core Power 720 server does not support the attachment of an EXP 12S Drawer.

The drawer can either be attached using the #EJ01 storage backplane, providing an external SAS port, or using one of the following adapters:

- PCIe LP 2-x4-port SAS 3 Gb adapter (#5278)
- PCIe 380 MB Cache Dual -x4 3 Gb SAS RAID adapter (#5805)
- PCI-X DDR Dual -x4 SAS adapter (#5900)
- PCIe Dual -x4 SAS adapter (#5901)
- PCIe 380 MB Cache Dual -x4 3 Gb SAS RAID adapter (#5903)
- PCI-X DDR 1.5 GB Cache SAS RAID adapter (#5904)
- PCI-X DDR Dual -x4 SAS adapter (#5912)
- PCIe2 1.8 GB Cache RAID SAS Adapter Tri-port 6 Gb (#5913)

Note: If you use a PCI-X SAS adapter within the Power 720 and 740, a PCI-X DDR 12X Expansion Drawer (#5796) or a 7314-G30 drawer is required.

With proper cabling and configuration, multiple wide ports are used to provide redundant paths to each dual-port SAS disk. The adapter manages SAS path redundancy and path switching in case a SAS drive failure occurs. The SAS Y cables attach to an EXP 12S Disk Drawer. Use SAS cable (YI) system to SAS enclosure, single controller/dual path 1.5 M (#3686, supported but not longer orderable) or SAS cable (YI) system to SAS enclosure, single controller/dual path 3 M (#3687) to attach SFF SAS drives in an EXP12S Drawer.

Figure 2-28 illustrates connecting a system external SAS port to a disk expansion drawer.

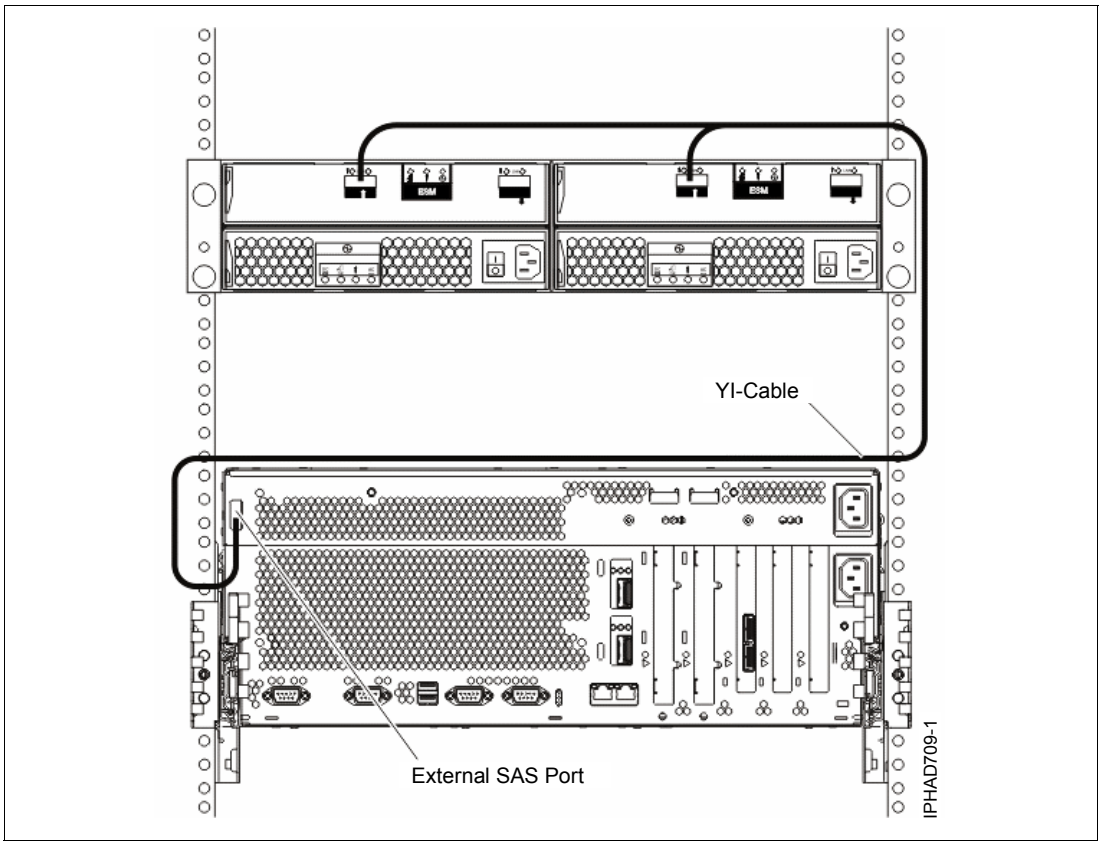


Figure 2-28 External SAS cabling

Use a SAS cable (YO) system to SAS enclosure, single controller/dual path 1.5 M (#3450) or SAS cable (YO) system to SAS enclosure, single controller/dual path 3 M (#3451) to attach SFF SAS drives in an EXP12S Drawer. In the EXP 12S Drawer, a high-availability I/O configuration can be created using a pair of #5278 adapters and SAS X cables to protect against the failure of a SAS adapter.

Various disk options are available to be installed in the EXP 12S drawer. Table 2-28 shows the available disk drive feature codes.

Table 2-28 Disk options for the EXP 12S drawer

Feature code	Description	OS support
#3586	69 GB 3.5" SAS Solid State Drive	AIX, Linux
#3647	146.8 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3648	300 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3649	450 GB 15 K RPM SAS Disk Drive	AIX, Linux
#3646	73.4 GB 15 K RPM SAS Disk Drive	AIX. Linux
#3587	69 GB 3.5” SAS Solid State Drive	IBM i
#3676	69.7 GB 15 K RPM SAS Disk Drive	IBM i
#3677	139.5 GB 15 K RPM SAS Disk Drive	IBM i

Feature code	Description	OS support
#3678	283.7 GB 15 K RPM SAS Disk Drive	IBM i
#3658	428.4 GB 15 K RPM SAS Disk Drive	IBM i

Note: A EXP 12S Drawer containing SSD drives cannot be attached to the CEC external SAS port on the Power 720 and Power 740 or through a PCIe LP 2-x4 port SAS adapter 3 Gb (#5278). If this configuration is required, use a high-profile PCIe SAS adapter or a PCI-X SAS adapter.

A second EXP 12S drawer can be attached to another drawer using two SAS EE cables, providing 24 SAS bays instead of 12 bays for the same SAS controller port. This is called cascading. In this configuration, all 24 SAS bays are controlled by a single controller or a single pair of controllers.

The EXP 12S Drawer can also be directly attached to the SAS port on the rear of the Power 720 and Power 740, providing a low-cost disk storage solution. The rear SAS port is provided by the storage backplane, eight SFF Bays/175MB RAID/Dual IOA (#EJ01).

A second unit cannot be cascaded to an EXP 12S Drawer attached in this way.

Note: If the internal disk bay of the Power 720 or Power 740 contains any SSD drives, an EXP 12S Drawer cannot be attached to the external SAS port on the Power 720 or Power 740 (this applies even if the I/O drawer only contains SAS disk drives).

For detailed information about the SAS cabling, see the Serial-attached SCSI cable planning documentation at:

http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7had/p7hadsascabling.htm

2.11.2 EXP24S SFF Gen2-bay Drawer

The EXP24S SFF Gen2-bay Drawer (#5887) is an expansion drawer supporting up to 24 hot-swap 2.5-inch SFF SAS hard disk drives (HDD) on POWER6 or POWER7 servers in 2U of 19-inch rack space.

The SFF bays of the EXP24S are different from the SFF bays of the POWER7 system units or 12X PCIe I/O Drawers (#5802 and #5803). The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the Gen-1 or SFF-1 bays of the POWER7 system unit or 12X PCIe I/O Drawers or vice versa.

The EXP 24S SAS ports are attached to SAS controllers, which can be a SAS PCI-X or PCIe adapter or pair of adapters. The EXP24S can also be attached to an imbedded SAS controller in a server with an imbedded SAS port. Attachment between the SAS controller and the EXP24S SAS ports is via the appropriate SAS Y or X cables.

The drawer can either be attached using the #EJ0F storage backplane, providing an external SAS port, or using the following adapters:

- PCI-X 1.5 GB Cache SAS RAID Adapter 3 Gb (#5908)
- PCIe Dual-x4 SAS Adapter 3 Gb (#5901 and #5278)
- PCIe2 1.8 GB Cache RAID SAS Adapter Tri-port 6 Gb (#5913)

Note: A single #5887 drawer can be cabled to the CEC external SAS port when a #5268 DASD backplane is part of the system. A 3Gb/s YI cable (#3686/#3687) is used to connect a #5887 to the CEC external SAS port.

A single #5887 will not be allowed to attach to the CEC external SAS port when a #EPC5 processor (4-core) is ordered/installed on a single socket Power 720 system.

The EXP24S can be ordered in one of three possible manufacturing-configured MODE settings (not customer set-up): 1, 2, or 4 sets of disk bays.

With IBM AIX, Linux, Virtual I/O Server, the EXP24S can be ordered with four sets of six bays (mode4), two sets of 12 bays (mode 2), or one set of 24 bays (mode 1). With IBM i the EXP24S can be ordered as one set of 24 bays (mode 1).

Note: Note the following information:

- The modes for the EXP24S SFF Gen2-bay Drawer are set by IBM Manufacturing. There is no option to reset after the drawer has been shipped.
- If you order multiple EXP24S, avoid mixing modes within that order. There is no externally visible indicator as to the drawer's mode.
- Several EXP24S cannot be cascaded on the external SAS connector. Only one #5887 is supported.
- The Power 720 or Power 740 support up to 14 EXP24S.

There are six SAS connectors on the rear of the EXP 24S drawer to which SAS adapters/controllers are attached. They are labeled T1, T2, and T3, and there are two T1, two T2, and two T3 (Figure 2-29):

- In mode 1, two or four of the six ports are used. Two T2 are used for a single SAS adapter, and two T2 and two T3 are used with a paired set of two adapters or dual adapters configuration.
- In mode 2 or mode 4, four ports will be used, two T2 and two T3, to access all SAS bays.

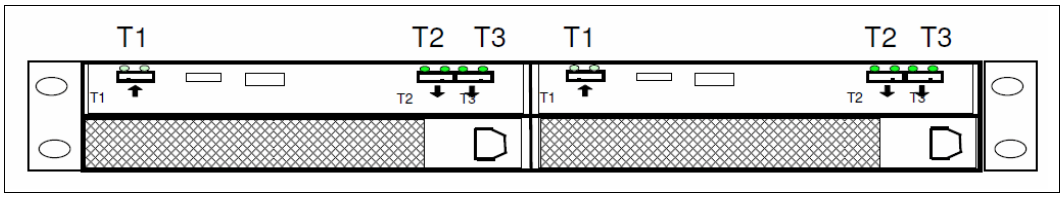


Figure 2-29 #5887 rear connectors

An EXP 24S in mode 4 can be attached to two or four SAS controllers and provide a lot of configuration flexibility. An EXP24S in mode 2 has similar flexibility. Up to 24 HDDs can be supported with any of the supported SAS adapters/controllers.

Include EXP24S no-charge specify codes with EXP24S orders to indicate to IBM Manufacturing the mode to which the drawer should be set and the adapter/controller/cable configuration that will be used. Table 2-29 lists the no-charge specify codes, the physical adapters/controllers/cables with their own chargeable feature numbers.

Table 2-29 EXP 24S Cabling

Feature code	Mode	Adapter/controller	Cable to drawer	Environment
#9359	1	One #5278	1 YO Cable	A, L, VIOS
#9360	1	Pair #5278	2 YO Cables	A, L, VIOS
#9361	2	Two #5278	2 YO Cables	A, L, VIOS
#9365	4	Four #5278	2 X Cable	A, L, VIOS
#9366	2	Two pair #5278	2 X Cables	A, L, VIOS
#9367	1	Pair #5805	2 YO Cables	A, i, L, VIOS
#9368	2	Two 5805	2 X Cables	A, L, VIOS
#9382	1	One #5904/06/08	1 YO Cable	A, i, L, VIOS
#9383	1	Pair #5904/06/08	2 YO Cables	A, i, L, VIOS
#9384	1	CEC SAS port	1 YI Cable	A, i, L, VIOS
#9385	1	Two #5913	2 YO Cables	A, i, L, VIOS
#9386	2	Four #5913	4 X Cables	A, L, VIOS

The following cabling options for the EXP 24S Drawer are available:

- X cables for #5278
  - 3 m (#3661)
  - 6 m (#3662)
  - 15 m (#3663)
- X cables for #5913 (all 6 Gb except for 15 m cable)
  - 3 m (#3454)
  - 6 m (#3455)
  - 10 m (#3456)
- YO cables for #5278
  - 1.5 m (#3691)
  - 3 m (#3692)
  - 6 m (#3693)
  - 15 m (#3694)
- YO cables for #5913 (all 6 Gb except for 15 m cable)
  - 1.5 m (#3450)
  - 3 m (#3451)
  - 6 m (#3452)
  - 10 m (#3453)
- YI cables for system unit SAS port (3 Gb)
  - 1.5 m (#3686)
  - 3 m (#3687)

Note: IBM plans to offer a 15-meter, 3 Gb bandwidth SAS cable for the #5913 PCIe2 1.8 GB Cache RAID SAS Adapter when attaching the EXP24S Drawer (#5887) for large configurations where the 10-meter cable is a distance limitation.

The EXP24S Drawer rails are fixed length and designed to fit Power Systems provided racks of 28 inches (711 mm) deep. EXP24S uses 2 EIA of space in a 19-inch-wide rack. Other racks might have different depths, and these rails will not adjust. No adjustable depth rails are orderable at this time.

For detailed information about the SAS cabling, see the serial-attached SCSI cable planning documentation at:

http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7had/p7hadsascabling.htm

2.11.3 TotalStorage EXP24 disk drawer and tower

The TotalStorage EXP24 is available as a 4 EIA unit drawer and mounts in a 19-inch rack (#5786) and a tower (#5787). The front of the IBM TotalStorage EXP24 Ultra320 SCSI Expandable Storage Disk Enclosure has bays for up to 12 disk drives organized in two SCSI groups of up to six drives. The rear also has bays for up to 12 disk drives organized in two additional SCSI groups of up to six drives plus slots for the four SCSI interface cards. Each SCSI drive group can be connected by either a Single Bus Ultra320 SCSI Repeater Card (#5741) or a Dual Bus Ultra320 SCSI Repeater Card (#5742). This allows the EXP24 to be configured as four sets of six bays, two sets of 12 bays, or two sets of six bays plus one set of 12 bays.

The EXP24 features #5786 and #5787 have three cooling fans and two power supplies to provide redundant power and cooling. The SCSI disk drives contained in the EXP24 are controlled by PCI-X SCSI adapters connected to the EXP24 SCSI repeater cards by SCSI cables. The PCI-X adapters are located in the Power 740 system unit or in an attached I/O drawer with PCI-X slots.

The 336 system maximum is achieved with a maximum of 24 disks in a maximum of 14 TotalStorage EXP24 disk drawers (#5786) or 14 TotalStorage EXP24 disk towers (#5787).

Note: The EXP24S SCSI disk drawer is an earlier technology drawer compared to the later SAS EXP12S drawer. It is used to house the older SCSI disk drives that are supported, but it is no longer orderable.

2.11.4 IBM TotalStorage EXP24

The IBM 7031 TotalStorage EXP24 Ultra320 SCSI Expandable Storage Disk Enclosure supports up to 24 Ultra320 SCSI Disk Drives arranged in four independent SCSI groups of up to six drives or in two groups of up to 12 drives. Each SCSI drive group can be connected by either a Single Bus Ultra320 SCSI Repeater Card or a Dual Bus Ultra320 SCSI Repeater Card, allowing a maximum of eight SCSI connections per TotalStorage EXP24.

The IBM 7031 Model D24 (7031-D24) is an Expandable Disk Storage Enclosure that is a horizontal 4 EIA by 19-inch rack drawer for mounting in equipment racks.

The IBM 7031 Model T24 (7031-T24) is an Expandable Disk Storage Enclosure that is a vertical tower for floor-standing applications.

Note: A new IBM 7031 TotalStorage EXP24 Ultra320 SCSI Expandable Storage Disk Enclosure cannot be ordered for the Power 720 and Power 740, and thus only existing 7031-D24 drawers or 7031-T24 towers can be moved to the Power 720 and 740 servers.

AIX and Linux partitions are supported along with the usage of a IBM 7031 TotalStorage EXP24 Ultra320 SCSI Expandable Storage Disk Enclosure.

2.11.5 IBM System Storage

The IBM System Storage Disk Systems products and offerings provide compelling storage solutions with superior value for all levels of business, from entry-level up to high-end storage systems.

IBM System Storage N series

The IBM System Storage N series is a Network Attached Storage (NAS) solution and provides the latest technology to customers to help them improve performance, virtualization manageability, and system efficiency at a reduced total cost of ownership. For more information about the IBM System Storage N series hardware and software, see:

http://www.ibm.com/systems/storage/network

IBM System Storage DS3000 family

The IBM System Storage DS3000 is an entry-level storage system designed to meet the availability and consolidation needs for a wide range of users. New features, including larger capacity 450 GB SAS drives, increased data protection features such as RAID 6, and more FlashCopies per volume, provide a reliable virtualization platform. For more information about the DS3000 family, see:

http://www.ibm.com/systems/storage/disk/ds3000/index.html

IBM System Storage DS5000

New DS5000 enhancements help reduce cost by introducing SSD drives. Also with the new EXP5060 expansion unit supporting 60 1 TB SATA drives in a 4U package, customers can see up to a one-third reduction in floor space over standard enclosures. With the addition of 1 Gbps iSCSI host attach, customers can reduce cost for their less demanding applications while continuing to provide high performance where necessary utilizing the 8 Gbps FC host ports. With the DS5000 family, you get consistent performance from a smarter design that simplifies your infrastructure, improves your TCO, and reduces your cost. For more information about the DS5000 family, see:

http://www.ibm.com/systems/storage/disk/ds5000/index.html

IBM Storwize V7000 Midrange Disk System

IBM® Storwize® V7000 is a virtualized storage system to complement virtualized server environments that provides unmatched performance, availability, advanced functions, and highly scalable capacity never seen before in midrange disk systems. Storwize V7000 is a powerful midrange disk system that has been designed to be easy to use and enable rapid deployment without additional resources. Storwize V7000 is virtual storage that offers greater efficiency and flexibility through built-in solid state drive (SSD) optimization and thin provisioning technologies. Storwize V7000 advanced functions also enable non-disruptive migration of data from existing storage, simplifying implementation and minimizing disruption to users. Storwize V7000 also enables you to virtualize and reuse existing disk systems,

supporting a greater potential return on investment (ROI). For more information about Storwize V7000, see:

http://www.ibm.com/systems/storage/disk/storwize_v7000/index.html

IBM XIV Storage System

IBM offers a mid–sized configuration of its self-optimizing, self-healing, resilient disk solution, the IBM XIV Storage System, storage reinvented for a new era. Now organizations with mid-size capacity requirements can take advantage of the latest IBM technology for their most demanding applications with as little as 27 TB usable capacity and incremental upgrades. For more information about XIV, see:

http://www.ibm.com/systems/storage/disk/xiv/index.html

IBM System Storage DS8000

The IBM System Storage DS8000 family is designed to offer high availability, multiplatform support, and simplified management tools. With its high capacity, scalability, broad server support, and virtualization features, the DS8000 family is well suited for simplifying the storage environment by consolidating data from multiple storage systems on a single system.

The high-end model DS8800 is the most advanced model in the IBM DS8000 family lineup and introduces new dual IBM POWER6-based controllers that usher in a new level of performance for the company's flagship enterprise disk platform. The DS8800 offers twice the maximum physical storage capacity than the previous model. For more information about the DS8000 family, see:

http://www.ibm.com/systems/storage/disk/ds8000/index.html

2.12 Hardware Management Console

The Hardware Management Console (HMC) is a dedicated workstation that provides a graphical user interface (GUI) for configuring, operating, and performing basic system tasks for the POWER7 processor-based systems (and the POWER5, POWER5+, POWER6, and POWER6+ processor-based systems) that function in either non-partitioned or clustered environments. In addition, the HMC is used to configure and manage partitions. One HMC is capable of controlling multiple POWER5, POWER5+, POWER6, POWER6+, and POWER7 processor-based systems.

Several HMC models are supported to manage POWER7 processor-based systems. Two models (7042-C08 and 7042-CR6) are available for ordering at the time of writing, but you can also use one of the withdrawn models listed in Table 2-30.

Table 2-30 HMC models supporting POWER7 processor technology based servers

Type-model	Availability	Description
7310-C05	Withdrawn	IBM 7310 Model C05 Desktop Hardware Management Console
7310-C06	Withdrawn	IBM 7310 Model C06 Deskside Hardware Management Console
7042-C06	Withdrawn	IBM 7042 Model C06 Deskside Hardware Management Console
7042-C07	Withdrawn	IBM 7042 Model C07 Deskside Hardware Management Console
7042-C08	Available	IBM 7042 Model C08 Deskside Hardware Management Console
7310-CR3	Withdrawn	IBM 7310 Model CR3 Rack-mounted Hardware Management Console

Type-model	Availability	Description
7042-CR4	Withdrawn	IBM 7042 Model CR4 Rack-mounted Hardware Management Console
7042-CR5	Withdrawn	IBM 7042 Model CR5 Rack-mounted Hardware Management Console
7042-CR6	Available	IBM 7042 Model CR6 Rack-mounted Hardware Management Console

At the time of writing, the HMC must be running V7R7.4.0. It can also support up to 48 POWER7 systems. Updates of the machine code, HMC functions, and hardware prerequisites can be found on Fix Central at this address:

http://www-933.ibm.com/support/fixcentral/

2.12.1 HMC functional overview

The HMC provides three groups of functions:

- Server
- Virtualization
- HMC management

Server management

The first group contains all functions related to the management of the physical servers under the control of the HMC:

- System password
- Status Bar
- Power On/Off
- Capacity on Demand
- Error management
  - System indicators
  - Error and event collection reporting
  - Dump collection reporting
  - Call Home
  - Customer notification
  - Hardware replacement (Guided Repair)
  - SNMP events
- Concurrent Add/Repair/Upgrade
- Redundant Service Processor
- Firmware Updates

Virtualization management

The second group contains all of the functions related to virtualization features such as a partitions configuration or the dynamic reconfiguration of resources:

- System Plans
- System Profiles
- Partitions (create, activate, shutdown)
- Profiles
- Partition Mobility
- DLPAR (processors, memory, I/O, and so on)
- Custom Groups

HMC Console management

The last group relates to the management of the HMC itself, its maintenance, security, and configuration, for example:

- Guided set-up wizard
- Electronic Service Agent set up wizard
- User Management
  - User IDs
  - Authorization levels
  - Customizable authorization
- Disconnect and reconnect
- Network Security
  - Remote operation enable and disable
  - User definable SSL certificates
- Console logging
- HMC Redundancy
- Scheduled Operations
- Back-up and Restore
- Updates, Upgrades
- Customizable Message of the day

The HMC provides both a graphical interface and a command-line interface (CLI) for all management tasks. Remote connection to the HMC using a web browser (as of HMC Version 7. Previous versions required a special client program, called WebSM) is possible. The CLI is also available by using the Secure Shell (SSH) connection to the HMC. It can be used by an external management system or a partition to remotely perform many HMC operations.

2.12.2 HMC connectivity to the POWER7 processor based systems

POWER5, POWER5+, POWER6, POWER6+, and POWER7 processor-technology based servers that are managed by an HMC require Ethernet connectivity between the HMC and the server's Service Processor. In addition, if dynamic LPAR, Live Partition Mobility, or PowerVM Active Memory Sharing operations are required on the managed partitions, Ethernet connectivity is needed between these partitions and the HMC. A minimum of two Ethernet ports are needed on the HMC to provide such connectivity. The rack-mounted 7042-CR5 HMC default configuration provides four Ethernet ports. The deskside 7042-C07 HMC standard configuration offers only one Ethernet port. Be sure to order an optional PCIe adapter to provide additional Ethernet ports.

For any logical partition in a server it is possible to use a Shared Ethernet Adapter that is configured via a Virtual I/O Server. Therefore, a partition does not require its own physical adapter to communicate with an HMC.

For the HMC to communicate properly with the managed server, eth0 of the HMC must be connected to either the HMC1 port or HMC2 port of the managed server, although other network configurations are possible. You can attach a second HMC to HMC Port 2 of the server for redundancy (or vice versa). These must be addressed by two separate subnets.

Figure 2-30 shows a simple network configuration to enable the connection from HMC to the server and to enable Dynamic LPAR operations. For more details about HMC and the possible network connections, see *Hardware Management Console V7 Handbook*, SG24-7491.

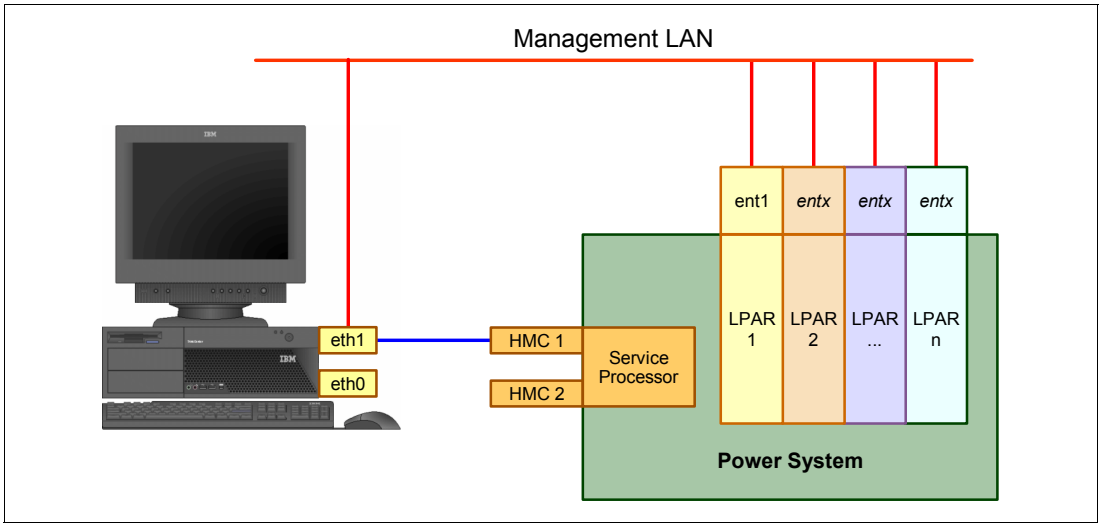


Figure 2-30 HMC to service processor and LPARs network connection

The default mechanism for allocation of the IP addresses for the service processor HMC ports is dynamic. The HMC can be configured as a DHCP server, providing the IP address at the time the managed server is powered on. In this case, the FSPs are allocated IP addresses from a set of address ranges predefined in the HMC software. These predefined ranges are identical for Version 710 of the HMC code and for previous versions.

If the service processor of the managed server does not receive a DHCP reply before time out, predefined IP addresses will be set up on both ports. Static IP address allocation is also an option. You can configure the IP address of the service processor ports with a static IP address by using the Advanced System Management Interface (ASMI) menus.

Note: The service processor is used to monitor and manage the system hardware resources and devices. The service processor offers two Ethernet 10/100 Mbps ports as connections. Note the following information:

- Both Ethernet ports are visible only to the service processor and can be used to attach the server to an HMC or to access the ASMI options from a client web browser using the HTTP server integrated into the service processor internal operating system.
- When not configured otherwise (DHCP or from a previous ASMI setting), both Ethernet ports of the first FSP have predefined IP addresses:
  - Service processor Eth0 or HMC1 port is configured as 169.254.2.147 with netmask 255.255.255.0.
  - Service processor Eth1 or HMC2 port is configured as 169.254.3.147 with netmask 255.255.255.0.

  For the second FSP of IBM Power 770 and 780, these default addresses are:
  - Service processor Eth0 or HMC1 port is configured as 169.254.2.146 with netmask 255.255.255.0.
  - Service processor Eth1 or HMC2 port is configured as 169.254.3.146 with netmask 255.255.255.0.

For more information about the service processor, see “Service processor” on page 144.

2.12.3 High availability using the HMC

The HMC is an important hardware component. When in operation, POWER7 processor-based servers and their hosted partitions can continue to operate when no HMC is available. However, in such conditions, certain operations cannot be performed, such as a DLPAR reconfiguration, a partition migration using PowerVM Live Partition Mobility, or the creation of a new partition. You might therefore decide to install two HMCs in a redundant configuration so that one HMC is always operational, even when performing maintenance of the other one for, example.

If redundant HMC function is desired, a server can be attached to two independent HMCs to address availability requirements. Both HMCs must have the same level of Hardware Management Console Licensed Machine Code Version 7 and installed fixes to manage POWER7 processor-based servers or an environment with a mixture of POWER5, POWER5+, POWER6, POWER6+, and POWER7 processor-based servers. The HMCs provide a locking mechanism so that only one HMC at a time has write access to the service processor. It is recommended that both HMCs are available on a public subnet to allow full syncronization of functionality. Depending on your environment, you have multiple options to configure the network.

Figure 2-31 shows one possible highly available HMC configuration managing two servers. These servers have only one CEC and therefore only one FSP. Each HMC is connected to one FSP port of all managed servers.

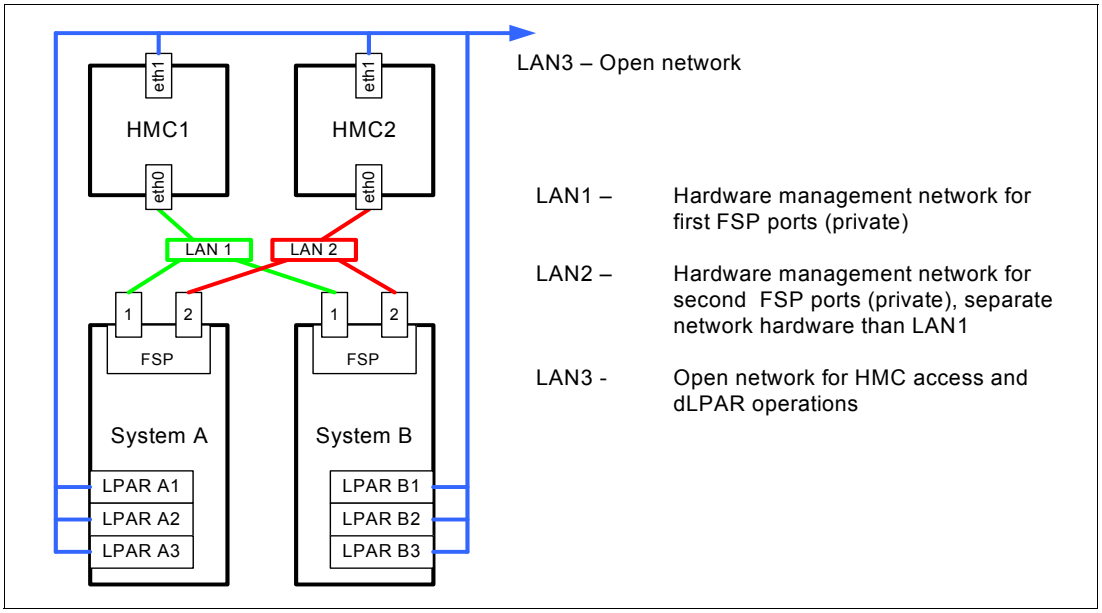


Figure 2-31 Highly available HMC and network architecture

Note that only hardware management networks (LAN1 and LAN2) are highly available (as shown in Figure 2-31) for simplicity. However, the management network (LAN3) can be made highly available by using a similar concept and adding more Ethernet adapters to LPARs and HMCs.

Both HMCs must be on a separate VLAN to protect from any network contention. Each HMC can be a DHCP server for its VLAN.

For more details about redundant HMCs, see the *Hardware Management Console V7 Handbook*, SG24-7491.

2.12.4 HMC code level

The HMC code must be at V7R7.4.0 to support the Power 720 and Power 740 systems.

In a dual-HMC configuration, both must be at the same version and release of the HMC.

Tips: Note these tips:

- When upgrading the code of a dual-HMC configuration, a good practice is to disconnect one HMC to avoid having both HMCs connected to the same server but running different levels of code. If no profiles or partition changes take place during the upgrade, both HMCs can stay connected. If the HMCs are at different levels and a profile change is made from the HMC at level V7R7.4.0, for example, the format of the data stored in the server could be changed, causing the HMC at a previous level (for example, 3.50) to possibly go into a recovery state, because it does not understand the new data format.
- Compatibility rules exist between the various software that is executing within a POWER7 processor-based server environment:
  - HMC
  - Virtual I/O Server
  - System firmware
  - Partition operating systems

  To check which combinations are supported, and to identify required upgrades, you can use the Fix Level Recommendation Tool web page:

  http://www14.software.ibm.com/webapp/set2/flrt/home

If you want to migrate a LPAR from a POWER6 processor-based server onto a POWER7 processor-based server using PowerVM Live Partition Mobility, consider this information: If the source server is managed by one HMC and the destination server is managed by a different HMC, ensure that the HMC managing the POWER6 processor-based server is at V7R7.3.5 or later and the HMC managing the POWER7 processor-based server is at V7R7.4.0 or later.

2.13 IBM Systems Director Management Console

The newly released IBM Systems Director Management Console (SDMC) is intended to be used in the same manner as the HMC. It provides the same functionality, including hardware, service, and virtualization management, for Power Systems server and Power Systems blades. Because SDMC uses IBM Systems Director Express® Edition, it also provides all Systems Director Express capabilities, such as monitoring of operating systems and creating event action plans.

No configuration changes are required when a client moves from HMC management to SDMC management.

Much of the SDMC function is equivalent to the HMC. This includes:

- Server (host) management
- Virtualization management
- Redundancy and high availability

  The SDMC offers console redundancy similar to the HMC.

The scalability and performance of the SDMC matches that of a current HMC. This includes both the number of systems (hosts) and the number of partitions (virtual servers) that can be managed. Currently, 48 small-tier entry servers or 32 large-tier servers can be managed by the SDMC with up to 1,024 partitions (virtual servers) configured across those managed systems (hosts).

The SDMC can be obtained as a hardware appliance in the same manner as an HMC. Hardware appliances support managing all Power Systems servers. The SDMC can optionally be obtained in a virtual appliance format, capable of running on VMware (ESX/i 4, or later) and KVM (Red Hat Enterprise Linux (RHEL) 5.5). The virtual appliance is only supported for managing small-tier Power servers and Power Systems blades.

Note: At the time of writing, the SDMC is not supported for the Power 720 (8202-E4C) and Power 740 (8205-E6C) models.

IBM intends to enhance the IBM Systems Director Management Console (SDMC) to support the Power 720 (8202-E4C) and Power 740 (8205-E6C). IBM also intends for the current HMC 7042-CR6 to be upgradable to an IBM SDMC that supports the Power 710 (8202-E4C) and Power 740 (8205-E6C).

Table 2-31 details whether the SDMC software appliance, hardware appliance, or both are supported for each model.

Table 2-31 Type of SDMC appliance support for POWER7-based server

POWER7 models	Type of SDMC appliance supported
7891-73X (IBM BladeCenter PS703)	Hardware or software appliance
7891-74X (IBM BladeCenter PS704)	Hardware or software appliance
8202-E4B (IBM Power 720 Express)	Hardware or software appliance
8205-E6B (IBM Power 740 Express)	Hardware or software appliance
8406-70Y (IBM BladeCenter PS700)	Hardware or software appliance
8406-71Y (IBM BladeCenter PS701 and PS702)	Hardware or software appliance
8231-E2B (IBM Power 710 and IBM Power 730 Express)	Hardware or software appliance
8233-E8B (IBM Power 750 Express)	Hardware or software appliance
8236-E8C (IBM Power 755)	Hardware or software appliance
9117-MMB (IBM Power 770)	Hardware appliance only
9179-MHB (IBM Power 780)	Hardware appliance only
9119-FHB (IBM Power 795)	Hardware appliance only

The IBM SDMC Hardware Appliance requires an IBM 7042-CR6 rack-mounted Hardware Management Console with the IBM SDMC indicator (#0963).

Note: When ordering #0963, the features #0031(No Modem), #1946 (additional 4 GB memory), and #1998 (additional 500 GB SATA HDD) are being configured automatically. Feature #0963 replaces the HMC software with IBM Systems Director Management Console Hardware Appliance V6.7.3 (5765-MCH).

Neither an external modem (#0032) nor an internal modem (#0033) can be selected with the IBM SDMC indicator (#0963).

To run HMC LMC (#0962), you cannot order the additional storage (#1998). However, you can order the additional memory (#1946) if wanted.

The IBM SDMC Virtual Appliance requires IBM Systems Director Management Console V6.7.3 (5765-MCV).

Note: If you want to use the software appliance, you have to provide the hardware and virtualization environment.

At a minimum, the following resources must be available to the virtual machine:

- 2.53 GHz Intel Xeon E5630, Quad Core processor
- 500 GB storage
- 8 GB memory

The following hypervisors are supported:

- VMware (ESXi 4.0 or later)
- KVM (RHEL 5.5)

The SDMC on POWER6 processor-based servers and blades requires eFirmware level 3.5.7. A SDMC on Power Systems POWER7 processor-based servers and blades requires eFirmware level 7.3.0.

For more detailed information about the SDMC, see *IBM Systems Director Management Console: Introduction and Overview*, SG24-7860.

2.14 Operating system support

The IBM POWER7 processor-based systems support three families of operating systems:

- AIX
- IBM i
- Linux

In addition, the Virtual I/O server can be installed in special partitions that provide support to the other operating systems for using features such as virtualized I/O devices, PowerVM Live Partition Mobility, or PowerVM Active Memory Sharing.

Note: For details about the software available on IBM Power Systems, visit the Power Systems Software site:

http://www.ibm.com/systems/power/software/index.html

2.14.1 Virtual I/O Server

The minimum required level of Virtual I/O server for both the Power 720 and Power 740 is VIOS 2.2.1.0.

IBM regularly updates the Virtual I/O server code. To find information about the latest updates, visit the Fix Central website:

http://www-933.ibm.com/support/fixcentral/

2.14.2 IBM AIX operating system

The following sections discuss the various levels of AIX operating system support.

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX operating system. Information about these packages and downloading and obtaining the CD-ROM is on the Fix Central website:

http://www-933.ibm.com/support/fixcentral/

The Fix Central website also provides information about how to obtain the fixes shipping on CD-ROM.

The Service Update Management Assistant, which can help you automate the task of checking and downloading operating system downloads, is part of the base operating system. For more information about the **suma** command, go to the following website:

http://www14.software.ibm.com/webapp/set2/sas/f/genunix/suma.html

IBM AIX Version 5.3

The minimum level of AIX Version 5.3 to support the Power 720 and Power 740 is AIX 5.3 with the 5300-12 Technology Level and Service Pack 5 or later.

A partition using AIX Version 5.3 will be executing in POWER6 or POWER6+ compatibility mode. This means that although the POWER7 processor has the ability to run four hardware threads per core simultaneously, using AIX 5.3 limits the number of hardware threads per core to two.

IBM AIX Version 6.1

The minimum level of AIX Version 6.1 to support the Power 720 and Power 740 is:

- AIX 6.1 with the 6100-07 Technology Level or later
- AIX 6.1 with the 6100-06 Technology Level and Service Pack 6 or later
- AIX 6.1 with the 6100-05 Technology Level and Service Pack 7 or later

A partition using AIX 6.1 with TL6 can run in POWER6, POWER6+, or POWER7 mode. It is best to run the partition in POWER7 mode to allow exploitation of new hardware capabilities such as SMT4 and Active Memory Expansion (AME).

IBM AIX Version 7.1

The minimum level of AIX Version 7.1 to support the Power 720 and Power 740 is as follows:

- AIX 7.1 with the 7100-01 Technology Level or later
- AIX 7.1 with the 7100-00 Technology Level and Service Pack 4 or 1 later

A partition using AIX 7.1 can run in POWER6, POWER6+, or POWER7 mode. It is best to run the partition in POWER7 mode to allow exploitation of new hardware capabilities such as SMT4 and Active Memory Expansion (AME).

2.14.3 IBM i operating system

The IBM i operating system is supported on the Power 720 and Power 740 with the minimum required levels:

- IBM i Version 6.1 with i 6.1.1 machine code or later
- IBM i Version 7.1 or later

IBM periodically releases maintenance packages (service packs or technology levels) for the IBM i operating system. Information about these packages and downloading and obtaining the CD-ROM is on the Fix Central website:

http://www-933.ibm.com/support/fixcentral/

2.14.4 Linux operating system

Linux is an open source operating system that runs on numerous platforms from embedded systems to mainframe computers. It provides a UNIX-like implementation across many computer architectures.

The supported versions of Linux on POWER7 processor-based servers are:

- SUSE Linux Enterprise Server 11 Service Pack 1, or later, with one current maintenance update available from SUSE to enable all planned functionality
- Red Hat Enterprise Linux AP 5 Update 7 for POWER, or later
- Red Hat Enterprise Linux 6.1 for POWER, or later

If you want to configure Linux partitions in virtualized Power Systems you have to be aware of these conditions:

- Not all devices and features that are supported by the AIX operating system are supported in logical partitions running the Linux operating system.
- Linux operating system licenses are ordered separately from the hardware. You can acquire Linux operating system licenses from IBM, to be included with the POWER7 processor-based servers, or from other Linux distributors.

For information about the features and external devices supported by Linux, go to:

http://www.ibm.com/systems/p/os/linux/index.html

For information about SUSE Linux Enterprise Server 10, refer to:

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

For information about Red Hat Enterprise Linux Advanced Server, see:

http://www.redhat.com/rhel/features

2.14.5 Java supported versions

There are unique considerations when running Java 1.4.2 on POWER7 servers. For best exploitation of the outstanding performance capabilities and most recent improvements of POWER7 technology, IBM recommends upgrading Java-based applications to Java 7, Java 6, or Java 5 whenever possible. For more information, visit:

http://www.ibm.com/developerworks/java/jdk/aix/service.html

2.14.6 Boost performance and productivity with IBM compilers

IBM XL C, XL C/C++, and XL Fortran compilers for AIX and for Linux exploit the latest POWER7 processor architecture. Release after release, these compilers continue to help improve application performance and capability, exploiting architectural enhancements made available through the advancement of the POWER technology.

IBM compilers are designed to optimize and tune your applications for execution on IBM POWER platforms to help you unleash the full power of your IT investment, to create and maintain critical business and scientific applications, to maximize application performance, and to improve developer productivity.

The performance gain from years of compiler optimization experience is seen in the continuous release-to-release compiler improvements that support the POWER4™

processors, through to the POWER4+™, POWER5™, POWER5+™, and POWER6 processors, and now including the new POWER7 processors. With the support of the latest POWER7 processor chip, IBM advances a more than 20-year investment in the XL compilers for POWER series and PowerPC® series architectures.

XL C, XL C/C++, and XL Fortran features introduced to exploit the latest POWER7 processor include vector unit and vector scalar extension (VSX) instruction set to efficiently manipulate vector operations in your application, vector functions within the Mathematical Acceleration Subsystem (MASS) libraries for improved application performance, built-in functions or intrinsics and directives for direct control of POWER instructions at the application level, and architecture and tune compiler options to optimize and tune your applications.

COBOL for AIX enables you to selectively target code generation of your programs to either exploit POWER7 systems architecture or to be balanced among all supported Power Systems. The performance of COBOL for AIX applications is improved by means of an enhanced back-end optimizer. The back-end optimizer, a component common also to the IBM XL compilers, lets your applications leverage the latest industry-leading optimization technology.

The performance of PL/I for AIX applications has been improved through both front-end changes and back-end optimizer enhancements. The back-end optimizer, a component common also to the IBM XL compilers, lets your applications leverage the latest industry-leading optimization technology. It will produce for PL/I, code that is intended to perform well across all hardware levels, including POWER7, of AIX.

IBM Rational® Development Studio for IBM i 7.1 provides programming languages for creating modern business applications. This includes the ILE RPG, ILE COBOL, C, and C++ compilers as well as the heritage RPG and COBOL compilers. The latest release includes performance improvements and XML processing enhancements for ILE RPG and ILE COBOL, improved COBOL portability with a new COMP-5 data type, and easier Unicode migration with relaxed USC2 rules in ILE RPG. Rational has also released a product called Rational Open Access: RPG Edition. This product opens up the ILE RPG file I/O processing, enabling partners, tool providers, and users to write custom I/O handlers that can access other devices like databases, services, and web user interfaces.

IBM Rational Developer for Power Systems Software™ provides a rich set of integrated development tools that support the XL C/C++ for AIX compiler, the XL C for AIX compiler, and the COBOL for AIX compiler. Rational Developer for Power Systems Software offers capabilities of file management, searching, editing, analysis, build, and debug, all integrated into an Eclipse workbench. XL C/C++, XL C, and COBOL for AIX developers can boost productivity by moving from older, text-based, command-line development tools to a rich set of integrated development tools.

The IBM Rational Power Appliance solution provides a workload optimized system and integrated development environment for AIX development on IBM Power Systems. IBM Rational Power Appliance includes a Power Express server preinstalled with a comprehensive set of Rational development software along with the AIX operating system. The Rational development software includes support for Collaborative Application Lifecycle Management (C/ALM) through Rational Team Concert, a set of software development tools from Rational Developer for Power Systems Software, and a choice between the XL C/C++ for AIX or COBOL for AIX compilers.

2.15 Energy management

The Power 720 and Power 740 servers are designed with features to help clients become more energy efficient. The IBM Systems Director Active Energy Manager exploits EnergyScale technology, enabling advanced energy management features to dramatically and dynamically conserve power and further improve energy efficiency. Intelligent Energy optimization capabilities enable the POWER7 processor to operate at a higher frequency for increased performance and performance per watt or dramatically reduce frequency to save energy.

2.15.1 IBM EnergyScale technology

IBM EnergyScale technology provides functions to help the user understand and dynamically optimize the processor performance versus processor energy consumption, and system workload, to control IBM Power Systems power and cooling usage.

On POWER7 processor-based systems, the thermal power management device (TPMD) card is responsible for collecting the data from all system components, changing operational parameters in components, and to interact with the IBM Systems Director Active Energy Manager (an IBM Systems Directors plug-in) for energy management and control.

IBM EnergyScale makes use of power and thermal information collected from the system to implement policies that can lead to better performance or better energy utilization. IBM EnergyScale features include these:

- Power trending

  EnergyScale provides continuous collection of real-time server energy consumption. This enables administrators to predict power consumption across their infrastructure and to react to business and processing needs. For example, administrators can use such information to predict datacenter energy consumption at various times of the day, week, or month.

- Thermal reporting

  IBM Director Active Energy Manager can display measured ambient temperature and calculated exhaust heat index temperature. This information can help identify data center hot spots that need attention.

- Power Saver Mode

  Power Saver Mode lowers the processor frequency and voltage on a fixed amount, reducing the energy consumption of the system while still delivering predictable performance. This percentage is predetermined to be within a safe operating limit and is not user configurable. The server is designed for a fixed frequency drop of up to 30% down from nominal frequency (the actual value depends on the server type and configuration). Power Saver Mode is not supported during boot or re-boot, although it is a persistent condition that will be sustained after the boot when the system starts executing instructions.

- Dynamic Power Saver Mode

  Dynamic Power Saver Mode varies processor frequency and voltage based on the utilization of the POWER7 processors. Processor frequency and utilization are inversely proportional for most workloads, implying that as the frequency of a processor increases, its utilization decreases, given a constant workload. Dynamic Power Saver Mode takes advantage of this relationship to detect opportunities to save power, based on measured real-time system utilization.

When a system is idle, the system firmware will lower the frequency and voltage to Power Energy Saver Mode values. When fully utilized, the maximum frequency will vary, depending on whether the user favors power savings or system performance. If an administrator prefers energy savings and a system is fully utilized, the system is designed to reduce the maximum frequency to 95% of nominal values. If performance is favored over energy consumption, the maximum frequency can be increased to up to 109% of nominal frequency for extra performance.

Dynamic Power Saver Mode is mutually exclusive with Power Saver Mode. Only one of these modes can be enabled at a given time.

- Power Capping

  Power Capping enforces a user-specified limit on power usage. Power Capping is not a power-saving mechanism. It enforces power caps by throttling the processors in the system, degrading performance significantly. The idea of a power cap is to set a limit that must never be reached but that frees up extra power never used in the data center. The *margined* power is this amount of extra power that is allocated to a server during its installation in a datacenter. It is based on the server environmental specifications that usually are never reached because server specifications are always based on maximum configurations and worst-case scenarios. The user must set and enable an energy cap from the IBM Director Active Energy Manager user interface.

- Soft Power Capping

  There are two power ranges into which the power cap can be set. One is Power Capping, as described previously, and the other is Soft Power Capping. Soft Power Capping extends the allowed energy capping range further, beyond a region that can be guaranteed in all configurations and conditions. If the energy management goal is to meet a particular consumption limit, then Soft Power Capping is the mechanism to use.

- Processor Core Nap Mode

  The IBM POWER7 processor uses a low-power mode called Nap that stops processor execution when there is no work to do on that processor core. The latency of exiting Nap is very small, typically not generating any impact on applications running. Because of that, the POWER Hypervisor can use Nap mode as a general-purpose idle state. When the operating system detects that a processor thread is idle, it yields control of a hardware thread to the POWER Hypervisor. The POWER Hypervisor immediately puts the thread into Nap mode. Nap mode allows the hardware to turn the clock off on most of the circuits inside the processor core. Reducing active energy consumption by turning off the clocks allows the temperature to fall, which further reduces leakage (static) power of the circuits, causing a cumulative effect. Nap mode saves from 10 - 15% of power consumption in the processor core.

- Processor core Sleep Mode

  To be able to save even more energy, the POWER7 processor has an even lower power mode called Sleep. Before a core and its associated L2 and L3 caches enter Sleep mode, caches are flushed and transition lookaside buffers (TLB) are invalidated, and hardware clock is turned off in the core and in the caches. Voltage is reduced to minimize leakage current. Processor cores inactive in the system (such as CoD processor cores) are kept in Sleep mode. Sleep mode saves about 35% of power consumption in the processor core and associated L2 and L3 caches.

- Fan control and altitude input

  System firmware will dynamically adjust fan speed based on energy consumption, altitude, ambient temperature, and energy savings modes. Power Systems are designed to operate in worst-case environments, in hot ambient temperatures, at high altitudes, and with high-power components. In a typical case, one or more of these constraints are not valid. When no power savings setting is enabled, fan speed is based on ambient

temperature and assumes a high-altitude environment. When a power savings setting is enforced (either Power Energy Saver Mode or Dynamic Power Saver Mode), fan speed varies based on power consumption, ambient temperature, and altitude available. System altitude can be set in IBM Director Active Energy Manager. If no altitude is set, the system assumes a default value of 350 meters above sea level.

- Processor Folding

  Processor Folding is a consolidation technique that dynamically adjusts, over the short-term, the number of processors available for dispatch to match the number of processors demanded by the workload. As the workload increases, the number of processors made available increases. As the workload decreases, the number of processors made available decreases. Processor Folding increases energy savings during periods of low to moderate workload because unavailable processors remain in low-power idle states (Nap or Sleep) longer.

- EnergyScale for I/O

  IBM POWER7 processor-based systems automatically power off hot-pluggable PCI adapter slots that are empty or not being used. System firmware automatically scans all pluggable PCI slots at regular intervals, looking for those that meet the criteria for being not in use and powering them off. This support is available for all POWER7 processor-based servers and the expansion units that they support.

- Server Power Down

  If overall data center processor utilization is low, workloads can be consolidated on fewer numbers of servers so that some servers can be turned off completely. It makes sense to do this when there will be long periods of low utilization, such as weekends. AEM provides information, such as the power that will be saved and the time that it will take to bring a server back online, that can be used to help make the decision to consolidate and power off. As with many of the features available in IBM Systems Director and Active Energy Manager, this function is scriptable and can be automated.

- Partition Power Management

  Available with Active Energy Manager 4.3.1 and later, and POWER7 systems with 730 firmware release and later, is the capability to set a power savings mode for partitions or the system processor pool. As in the system-level power savings modes, the per-partition power savings modes can be used to achieve a balance between the power consumption and the performance of a partition. Only partitions that have dedicated processing units can have a unique power savings setting. Partitions that run in shared processing mode will have a common power savings setting, which is that of the system processor pool. This is because processing unit fractions cannot be power-managed.

  As in the case of system-level power savings, two Dynamic Power Saver options are offered:

  – Favor partition performance
  – Favor partition power savings

  The user must configure this setting from Active Energy Manager. When Dynamic Power Saver is enabled in either mode, system firmware continuously monitors the performance and utilization of each of the computer's POWER7 processor cores that belong to the partition. Based on this utilization and performance data, the firmware dynamically adjusts the processor frequency and voltage, reacting within milliseconds to adjust workload performance and also deliver power savings when the partition is under-utilized.

  In addition to the two Dynamic Power Saver options, the customer can select to have no power savings on a given partition. This option will leave the processor cores assigned to the partition running at their nominal frequencies and voltages.

A new power savings mode called *Inherit Host Setting* is available and is only applicable to partitions. When configured to use this setting, a partition adopts the power savings mode of its hosting server. By default, all partitions with dedicated processing units, and the system processor pool, are set to the Inherit Host Setting.

On POWER7 processor-based systems, several EnergyScales are imbedded in the hardware and do not require an operating system or external management component. More advanced functionality requires Active Energy Manager (AEM) and IBM Systems Director.

Table 2-32 provides a list of all features supported, underlining all cases where AEM is not required. Table 2-32 also notes the features that can be activated by traditional user interfaces (that is, ASMI and HMC).

Table 2-32 AEM support

Feature	Active Energy Manager (AEM) required	ASMI	HMC
Power Trending	Y	N	N
Thermal Reporting	Y	N	N
Static Power Saver	N	Y	Y
Dynamic Power Saver	Y	N	N
Power Capping	Y	N	N
Energy-optimized Fans	N	-	-
Processor Core Nap	N	-	-
Processor Core Sleep	N	-	-
Processor Folding	N	-	-
EnergyScale for I/O	N	-	-
Server Power Down	Y	-	-
Partition Power Management	Y	-	-

The Power 720 and Power 740 systems implement all the EnergyScale capabilities listed in 2.15.1, “IBM EnergyScale technology” on page 94.

2.15.2 Thermal power management device card

The thermal power management device (TPMD) card is a separate micro-controller installed on certain POWER6 processor-based systems and on all POWER7 processor-based systems. It runs real-time firmware whose sole purpose is to manage system energy.

The TPMD card monitors the processor modules, memory, environmental temperature, and fan speed, and based on this information, it can act upon the system to maintain optimal power and energy conditions (for example, it can increase the fan speed to react to a temperature change). It also interacts with the IBM Systems Director Active Energy Manager to report power and thermal information and to receive input from AEM on policies to be set. The TPMD is part of the EnergyScale infrastructure.

3

Virtualization

As you look for ways to maximize the return on your IT infrastructure investments, consolidating workloads becomes an attractive proposition.

IBM Power Systems combined with PowerVM technology is designed to help you consolidate and simplify your IT environment, with the following key capabilities:

- Improve server utilization and sharing I/O resources to reduce total cost of ownership and make better use of IT assets.
- Improve business responsiveness and operational speed by dynamically re-allocating resources to applications as needed, to better match changing business needs or handle unexpected changes in demand.
- Simplify IT infrastructure management by making workloads independent of hardware resources, thereby enabling you to make business-driven policies to deliver resources based on time, cost, and service-level requirements.

This chapter discusses the virtualization technologies and features on IBM Power Systems:

- POWER Hypervisor™
- POWER Modes
- Partitioning
- Active Memory Expansion
- PowerVM
- System Planning Tool

3.1 POWER Hypervisor

Combined with features designed into the POWER7 processors, the POWER Hypervisor delivers functions that enable other system technologies, including logical partitioning technology, virtualized processors, IEEE VLAN compatible virtual switch, virtual SCSI adapters, virtual Fibre Channel adapters, and virtual consoles. The POWER Hypervisor is a basic component of the system's firmware and offers the following functions:

- Provides an abstraction between the physical hardware resources and the logical partitions that use them
- Enforces partition integrity by providing a security layer between logical partitions
- Controls the dispatch of virtual processors to physical processors (see "Processing mode" on page 111)
- Saves and restores all processor state information during a logical processor context switch
- Controls hardware I/O interrupt management facilities for logical partitions
- Provides virtual LAN channels between logical partitions that help to reduce the need for physical Ethernet adapters for inter-partition communication
- Monitors the Service Processor and will perform a reset or reload if it detects the loss of the Service Processor, notifying the operating system if the problem is not corrected

The POWER Hypervisor is always active, regardless of the system configuration and also when not connected to the managed console. It requires memory to support the resource assignment to the logical partitions on the server. The amount of memory required by the POWER Hypervisor firmware varies according to several factors. Factors influencing the POWER Hypervisor memory requirements include:

- Number of logical partitions
- Number of physical and virtual I/O devices used by the logical partitions
- Maximum memory values specified in the logical partition profiles

The minimum amount of physical memory required to create a partition will be the size of the system's Logical Memory Block (LMB). The default LMB size varies according to the amount of memory configured in the CEC (Table 3-1).

Table 3-1 Configured CEC memory-to-default Logical Memory Block size

Configurable CEC memory	Default Logical Memory Block
Greater than 8 GB up to 16 GB	64 MB
Greater than 16 GB up to 32 GB	128 MB
Greater than 32 GB	256 MB

In most cases, however, the actual minimum requirements and recommendations of the supported operating systems are above 256 MB. Physical memory is assigned to partitions in increments of LMB.

The POWER Hypervisor provides the following types of virtual I/O adapters:

- Virtual SCSI
- Virtual Ethernet
- Virtual Fibre Channel
- Virtual (TTY) console

Virtual SCSI

The POWER Hypervisor provides a virtual SCSI mechanism for virtualization of storage devices. The storage virtualization is accomplished using two paired adapters:

- A virtual SCSI server adapter
- A virtual SCSI client adapter

A Virtual I/O Server partition or an IBM i partition can define virtual SCSI server adapters. Other partitions are *client* partitions. The Virtual I/O Server partition is a special logical partition, as described in 3.4.4, “Virtual I/O Server” on page 117. The Virtual I/O Server software is included on all PowerVM Editions and when using the PowerVM Standard Edition and PowerVM Enterprise Edition, dual Virtual I/O Servers can be deployed to provide maximum availability for client partitions when performing Virtual I/O Server maintenance.

Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to use a fast and secure communication without any need for physical interconnection. The virtual Ethernet allows a transmission speed in the range of 1 - 3 Gbps, depending on the maximum transmission unit (MTU) size and CPU entitlement. Virtual Ethernet support began with IBM AIX Version 5.3, or an appropriate level of Linux supporting virtual Ethernet devices (see 3.4.9, “Operating system support for PowerVM” on page 128). The virtual Ethernet is part of the base system configuration.

Virtual Ethernet has these major features:

- The virtual Ethernet adapters can be used for both IPv4 and IPv6 communication and can transmit packets with a size up to 65 408 bytes. Therefore, the maximum MTU for the corresponding interface can be up to 65 394 (65 390 if VLAN tagging is used).
- The POWER Hypervisor presents itself to partitions as a virtual 802.1Q-compliant switch. The maximum number of VLANs is 4 096. Virtual Ethernet adapters can be configured as either untagged or tagged (following the IEEE 802.1Q VLAN standard).
- A partition can support 256 virtual Ethernet adapters. Besides a default port VLAN ID, the number of additional VLAN ID values that can be assigned per virtual Ethernet adapter is 20, which implies that each virtual Ethernet adapter can be used to access 21 virtual networks.
- 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.

Any virtual Ethernet can also have connectivity outside of the server if a layer-2 bridge to a physical Ethernet adapter is set in one Virtual I/O Server partition. See 3.4.4, “Virtual I/O Server” on page 117, for more details about shared Ethernet), also known as Shared Ethernet Adapter.

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 Fibre Channel

A virtual Fibre Channel adapter is a virtual adapter that provides client logical partitions with a Fibre Channel connection to a storage area network through the Virtual I/O Server logical partition. The Virtual I/O Server logical partition provides the connection between the virtual Fibre Channel adapters on the Virtual I/O Server logical partition and the physical Fibre Channel adapters on the managed system. Figure 3-1 depicts the connections between the client partition virtual Fibre Channel adapters and the external storage. For additional information, see 3.4.8, “N_Port ID virtualization” on page 127.

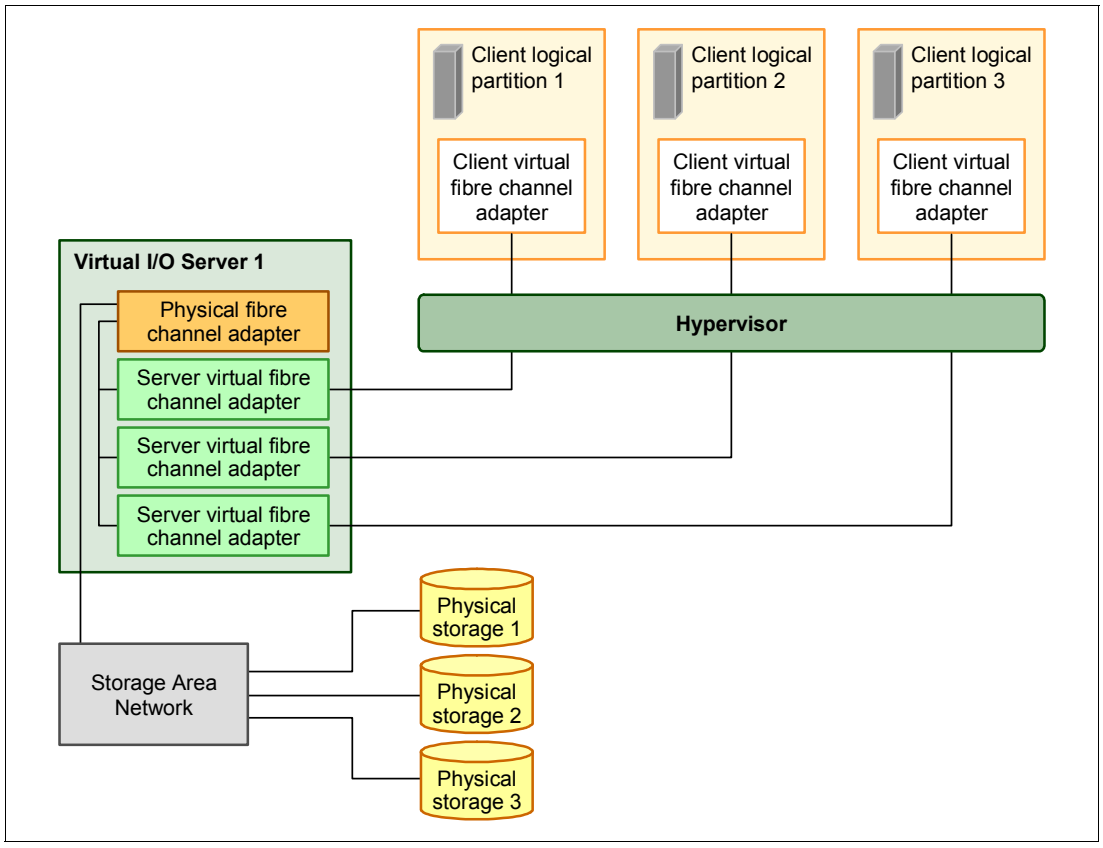


Figure 3-1 Connectivity between virtual Fibre Channels adapters and external SAN devices

Virtual (TTY) console

Each partition must have access to a system console. Tasks such as operating system installation, network setup, and various problem analysis activities require a dedicated system console. The POWER Hypervisor provides the virtual console by using a virtual TTY or serial adapter and a set of Hypervisor calls to operate on them. Virtual TTY does not require the purchase of any additional features or software, such as the PowerVM Edition features.

Depending on the system configuration, the operating system console can be provided by the Hardware Management Console virtual TTY, IVM virtual TTY, or from a terminal emulator that is connected to a system port.

3.2 POWER processor modes

Although, strictly speaking, not a virtualization feature, the POWER modes are described here because they affect various virtualization features.

On Power System servers, partitions can be configured to run in several modes, including:

- POWER6 compatibility mode

  This execution mode is compatible with Version 2.05 of the Power Instruction Set Architecture (ISA). For more information, visit the following address:

  http://www.power.org/resources/reading/PowerISA_V2.05.pdf

- POWER6+ compatibility mode

  This mode is similar to POWER6, with eight additional Storage Protection Keys.

- POWER7 mode

  This is the native mode for POWER7 processors, implementing the v2.06 of the Power Instruction Set Architecture. For more information, visit the following address:

  http://www.power.org/resources/downloads/PowerISA_V2.06_PUBLIC.pdf

The selection of the mode is made on a per-partition basis, from the HMC, by editing the partition profile (Figure 3-2).

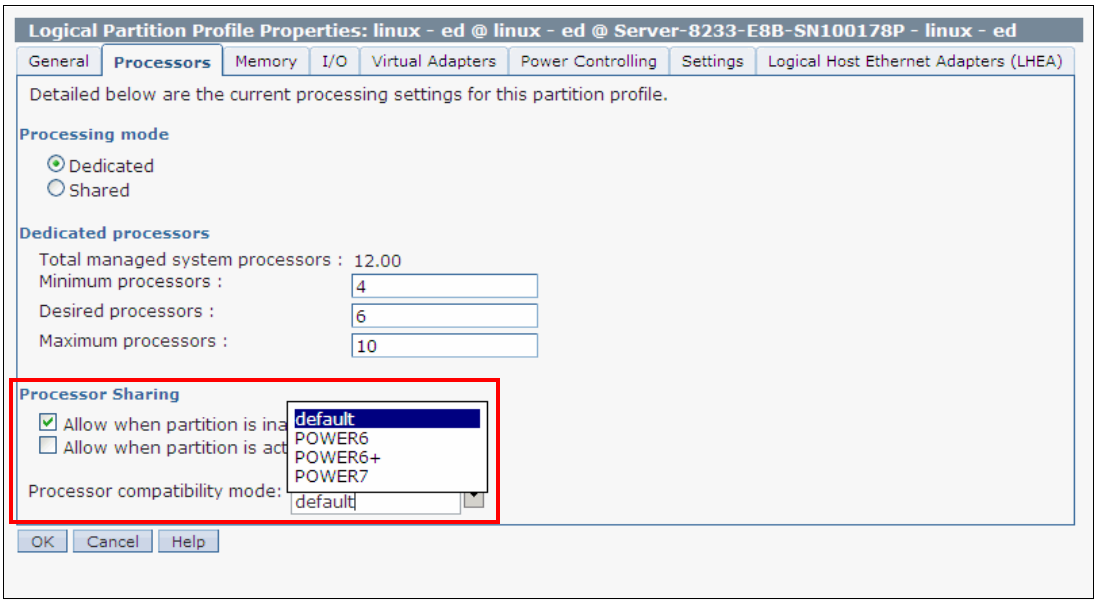


Figure 3-2 Configuring partition profile compatibility mode from the HMC

Table 3-2 lists the differences between these modes.

Table 3-2 Differences between POWER6 mode and POWER7 mode

POWER6 and POWER6+ mode	POWER7 mode	Customer value
2-thread SMT	4-thread SMT	Throughput performance, processor core utilization
VMX (Vector Multimedia Extension/AltiVec)	Vector Scalar Extension (VSX)	High-performance computing
Affinity OFF by default	3-tier memory, Micropartition Affinity	Improved system performance for system images spanning sockets and nodes
► Barrier Synchronization ► Fixed 128-byte Array, Kernel Extension Access	► Enhanced Barrier Synchronization ► Variable Sized Array, User Shared Memory Access	High-performance computing parallel programming synchronization facility
► 64-core and 128-thread scaling	► 32-core and 128-thread scaling ► 64-core and 256-thread scaling ► 256-core and 1024-thread scaling	Performance and Scalability for Large Scale-Up Single System Image Workloads (such as OLTP, ERP scale-up, WPAR consolidation)
EnergyScale CPU Idle	EnergyScale CPU Idle and Folding with NAP and SLEEP	Improved Energy Efficiency

3.3 Active Memory Expansion

Power Active Memory Expansion Enablement is an optional feature of POWER7 processor-based servers that must be specified when creating the configuration in the e-Config tool, as follows:

IBM Power 720 #4793
IBM Power 740 #4794

This feature enables memory expansion on the system. Using compression/decompression of memory content can effectively expand the maximum memory capacity, providing additional server workload capacity and performance.

Active Memory Expansion is an innovative POWER7 technology that allows the effective maximum memory capacity to be much larger than the true physical memory maximum. Compression/decompression of memory content can allow memory expansion up to 100%, which in turn enables a partition to perform significantly more work or support more users with the same physical amount of memory. Similarly, it can allow a server to run more partitions and do more work for the same physical amount of memory.

Active Memory Expansion is available for partitions running AIX 6.1, Technology Level 4 with SP2, or later.

Active Memory Expansion uses CPU resource of a partition to compress/decompress the memory contents of this same partition. The trade-off of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies based on how compressible the memory content is, and it also depends on having adequate spare CPU

capacity available for this compression/decompression. Tests in IBM laboratories, using sample work loads, showed excellent results for many workloads in terms of memory expansion per additional CPU utilized. Other test workloads had more modest results.

Clients have much control over Active Memory Expansion usage. Each individual AIX partition can turn on or turn off Active Memory Expansion. Control parameters set the amount of expansion desired in each partition to help control the amount of CPU used by the Active Memory Expansion function. An initial program load (IPL) is required for the specific partition that is turning memory expansion on or off. After it is turned on, monitoring capabilities are available in standard AIX performance tools, such as **lparstat**, **vmstat**, **topas**, and **svmon**.

Figure 3-3 represents the percentage of CPU that is used to compress memory for two partitions with separate profiles. The green curve corresponds to a partition that has spare processing power capacity. The blue curve corresponds to a partition constrained in processing power.

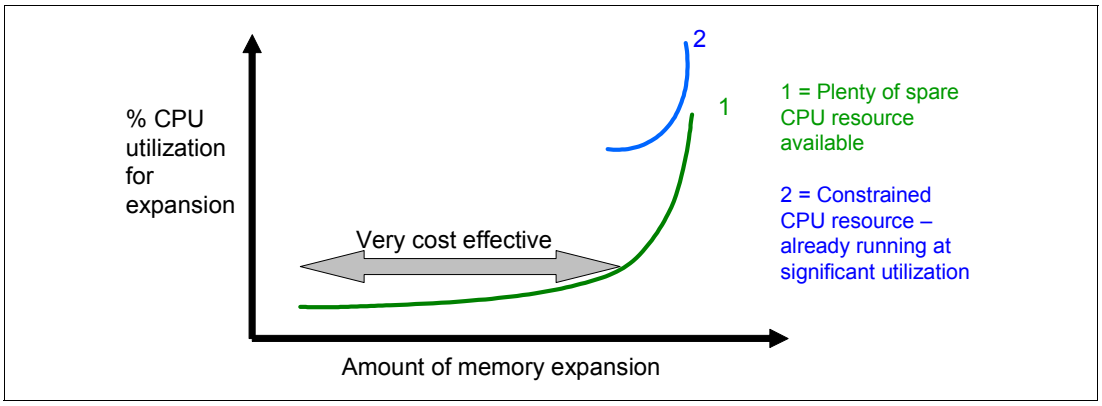


Figure 3-3 CPU usage versus memory expansion effectiveness

Both cases show that there is a knee-of-curve relationship for CPU resource required for memory expansion:

- Busy processor cores do not have resources to spare for expansion.
- The more memory expansion done, the more CPU resource required.

The knee varies depending on how compressible the memory contents are. This example demonstrates the need for a case-by-case study of whether memory expansion can provide a positive return on investment.

To help you perform this study, a planning tool is included with AIX 6.1 Technology Level 4, allowing you to sample actual workloads and estimate how expandable the partition's memory is and how much CPU resource is needed. Any Power System server can run the planning tool. Figure 3-4 shows an example of the output returned by this planning tool. The tool outputs various real memory and CPU resource combinations to achieve the desired effective memory. It also recommends one particular combination. In this example, the tool recommends that you allocate 58% of a processor, to benefit from 45% extra memory capacity.

```
Active Memory Expansion Modeled Statistics:
-----------------------
Modeled Expanded Memory Size :   8.00 GB

Expansion      True Memory       Modeled Memory       CPU Usage
 Factor        Modeled Size      Gain                 Estimate
---------    --------------    -----------------    -----------
  1.21          6.75 GB         1.25 GB [ 19%]          0.00
  1.31          6.25 GB         1.75 GB [ 28%]          0.20
  1.41          5.75 GB         2.25 GB [ 39%]          0.35
  1.51          5.50 GB         2.50 GB [ 45%]          0.58
  1.61          5.00 GB         3.00 GB [ 60%]          1.46

Active Memory Expansion Recommendation:
---------------------
The recommended AME configuration for this workload is to configure
the LPAR with a memory size of 5.50 GB and to configure a memory
expansion factor of 1.51.  This will result in a memory expansion of
45% from the LPAR's current memory size.  With this configuration,
the estimated CPU usage due to Active Memory Expansion is
approximately 0.58 physical processors, and the estimated overall
peak CPU resource required for the LPAR is 3.72 physical processors.
```

Figure 3-4 Output from Active Memory Expansion planning tool

After you select the value of the memory expansion factor that you want to achieve, you can use this value to configure the partition from the HMC (Figure 3-5).

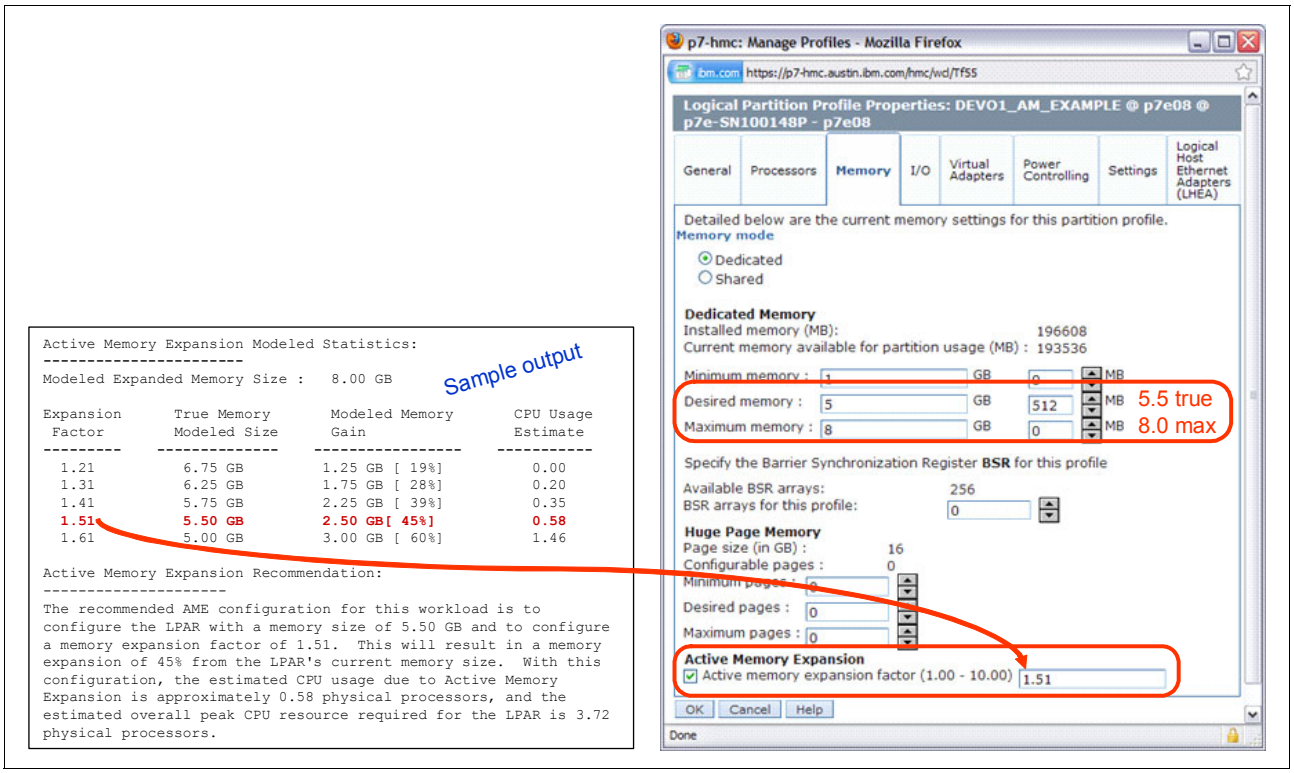


Figure 3-5 Using the planning tool result to configure the partition

On the HMC menu describing the partition, check the **Active Memory Expansion** box and enter true and maximum memory, and the memory expansion factor. To turn off expansion, clear the check box. In both cases, a reboot of the partition is needed to activate the change.

In addition, a one-time, 60-day trial of Active Memory Expansion is available to provide more exact memory expansion and CPU measurements. The trial can be requested using the Capacity on Demand web page.

http://www.ibm.com/systems/power/hardware/cod/

Active Memory Expansion can be ordered with the initial order of the server or as an MES order. A software key is provided when the enablement feature is ordered that is applied to the server. Rebooting is not required to enable the physical server. The key is specific to an individual server and is permanent. It cannot be moved to a separate server. This feature is ordered per server, independently of the number of partitions using memory expansion.

From the HMC, you can see whether the Active Memory Expansion feature has been activated (Figure 3-6).

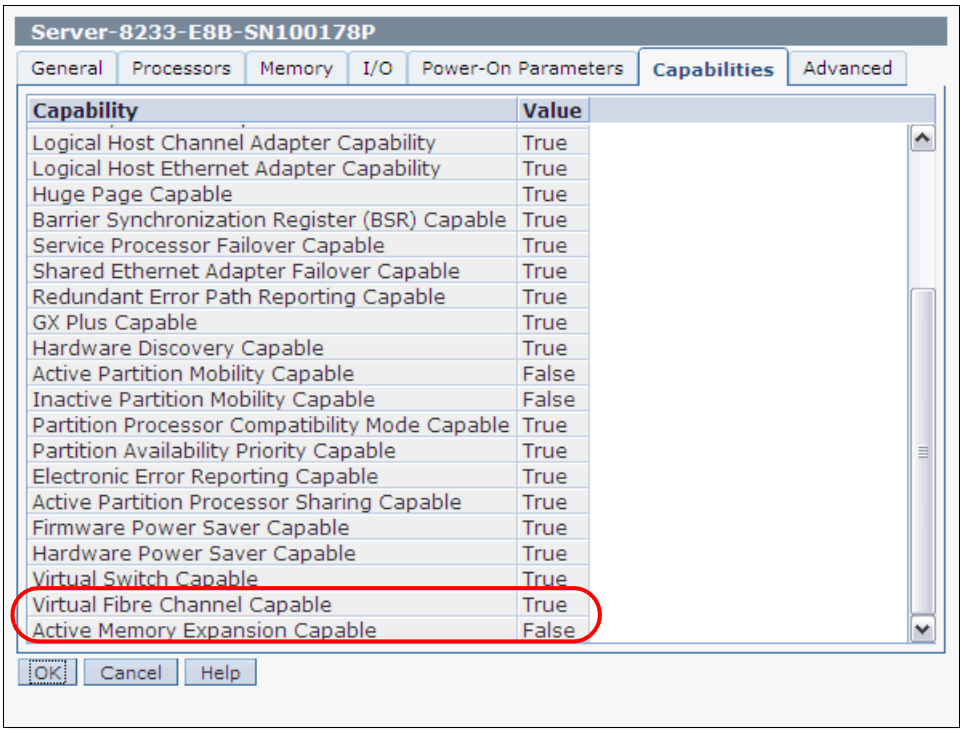


Figure 3-6 Server capabilities listed from the HMC

Note: If you want to move an LPAR using Active Memory Expansion to another system using Live Partition Mobility, the target system must support AME (the target system must have AME activated with the software key). If the target system does not have AME activated, the mobility operation fails during the pre-mobility check phase, and an appropriate error message displays to the user.

For detailed information regarding Active Memory Expansion, you can download the document *Active Memory Expansion: Overview and Usage Guide* from this location:

http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&appname=STGE_PO_PO_USEN&htmlfid=POW03037USEN

3.4 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that deliver industry-leading virtualization on the IBM Power Systems. It is the new umbrella branding term for Power Systems Virtualization (Logical Partitioning, Micro-Partitioning®, Power Hypervisor, Virtual I/O Server, Live Partition Mobility, Workload Partitions, and more). As with Advanced Power Virtualization in the past, PowerVM is a combination of hardware enablement and value-added software. Section 3.4.1, “PowerVM editions” on page 109, discusses the licensed features of each of the three separate editions of PowerVM.

3.4.1 PowerVM editions

This section provides information about the virtualization capabilities of the PowerVM. The three editions of PowerVM are suited for various purposes:

- PowerVM Express Edition

  PowerVM Express Edition is designed for customers looking for an introduction to more advanced virtualization features at a highly affordable price, generally in single-server projects.

- PowerVM Standard Edition

  This edition provides advanced virtualization functions and is intended for production deployments and server consolidation.

- PowerVM Enterprise Edition

  This edition is suitable for large server deployments such as multi-server deployments and cloud infrastructure. It includes unique features like Active Memory Sharing and Live Partition Mobility.

Table 3-3 lists the versions of PowerVM that are available on Power 720 and Power 740.

Table 3-3 Availability of PowerVM per POWER7 processor technology based server model

PowerVM editions	Express	Standard	Enterprise
IBM Power 720	#5225	#5227	#5228
IBM Power 740	#5225	#5227	#5228

For more information about the features included on each version of PowerVM, see *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940.

Note: At the time of writing, the IBM Power 720 (8202-E4C) and Power 740 (8205-E6C) have to be managed by the Hardware Management Console or by the Integrated Virtualization Manager.

3.4.2 Logical partitions (LPARs)

LPARs and virtualization increase utilization of system resources and add a new level of configuration possibilities. This section provides details and configuration specifications about this topic.

Dynamic logical partitioning

Logical partitioning was introduced with the POWER4 processor-based product line and the AIX Version 5.1 operating system. This technology offered the capability to divide a pSeries® system into separate logical systems, allowing each LPAR to run an operating environment on dedicated attached devices, such as processors, memory, and I/O components.

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

Micro-Partitioning

Micro-Partitioning technology allows you to allocate fractions of processors to a logical partition. This technology was introduced with POWER5 processor-based systems. A logical partition using fractions of processors is also known as a Shared Processor Partition or micro-partition. Micro-partitions run over a set of processors called Shared Processor Pool. Virtual processors are used to let the operating system manage the fractions of processing power assigned to the logical partition. From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system has been enhanced to be made aware of the difference. Physical processors are abstracted into virtual processors that are available to partitions. The meaning of the term *physical processor* in this section is a *processor core*. For example, a 2-core server has two physical processors.

When defining a shared processor partition, several options have to be defined:

- The minimum, desired, and maximum processing units

  Processing units are defined as processing power, or the fraction of time that the partition is dispatched on physical processors. Processing units define the capacity entitlement of the partition.

- The Shared Processor Pool

  Pick one from the list with the names of each configured Shared Processor Pool. This list also displays the pool ID of each configured Shared Processor Pool in parentheses. If the name of the desired Shared Processor Pool is not available here, you must first configure the desired Shared Processor Pool using the Shared Processor Pool Management window. Shared processor partitions use the default Shared Processor Pool called DefaultPool by default. See 3.4.3, “Multiple Shared Processor Pools” on page 112, for details about multiple Shared Processor Pools.

- Whether the partition will be able to access extra processing power to “fill up” its virtual processors above its capacity entitlement (selecting either to cap or uncap your partition)

  If there is spare processing power available in the Shared Processor Pool or other partitions are not using their entitlement, an uncapped partition can use additional processing units if its entitlement is not enough to satisfy its application processing demand.

- The weight (preference) in the case of an uncapped partition

- The minimum, desired, and maximum number of virtual processors

The POWER Hypervisor calculates partition’s processing power based on minimum, desired, and maximum values, processing mode, and is also based on requirements of other active partitions. The actual entitlement is never smaller than the processing unit’s desired value, but can exceed that value in the case of an uncapped partition and up to the number of virtual processors allocated.

A partition can be defined with a processor capacity as small as 0.10 processing units. This represents 0.10 of a physical processor. Each physical processor can be shared by up to 10 shared processor partitions and the partition’s entitlement can be incremented fractionally by as little as 0.01 of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under control of the POWER Hypervisor. The shared processor partitions are created and managed by the managed console or Integrated Virtualization Management.

The IBM Power 720 supports up to eight cores, and has these maximums:

- Up to eight dedicated partitions
- Up to 80 micro-partitions (10 micro-partitions per physical active core)

The Power 740 allows up to 16 cores in a single system, supporting the following maximums:

- Up to 16 dedicated partitions
- Up to 160 micro-partitions (10 micro-partitions per physical active core)

An important point is that the maximums stated are supported by the hardware, but the practical limits depend on the application workload demands

Additional information about virtual processors includes:

- A virtual processor can be running (dispatched) either on a physical processor or as standby waiting for a physical processor to become available.
- Virtual processors do not introduce any additional abstraction level. They are only a dispatch entity. When running on a physical processor, virtual processors run at the same speed as the physical processor.
- Each partition's profile defines CPU entitlement that determines how much processing power any given partition should receive. The total sum of CPU entitlement of all partitions cannot exceed the number of available physical processors in a Shared Processor Pool.
- The number of virtual processors can be changed dynamically through a dynamic LPAR operation.

Processing mode

When you create a logical partition you can assign entire processors for dedicated use, or you can assign partial processing units from a Shared Processor Pool. This setting defines the processing mode of the logical partition. Figure 3-7 shows a diagram of the concepts discussed in this section.

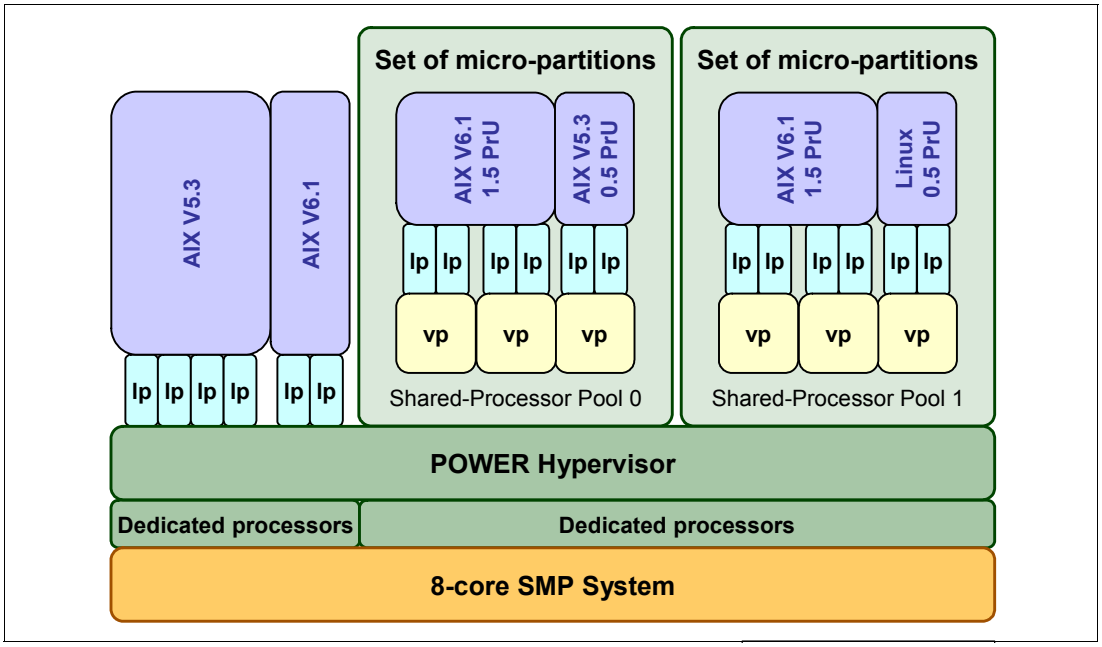


Figure 3-7 Logical partitioning concepts

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to partitions. The simultaneous multithreading feature in the POWER7 processor core allows the core to execute instructions from two or four independent software threads simultaneously. To support this feature we use the concept of *logical processors*. The operating system (AIX, IBM i, or Linux) sees one physical processor as two or four logical processors if the simultaneous multithreading feature is on. It can be turned off and on dynamically while the operating system is executing (for AIX, use the **smtctl** command). If simultaneous multithreading is off, each physical processor is presented as one logical processor, and thus only one thread.

Shared dedicated mode

On POWER7 processor technology based servers, you can configure dedicated partitions to become processor donors for idle processors that they own, allowing for the donation of spare CPU cycles from dedicated processor partitions to a Shared Processor Pool. The dedicated partition maintains absolute priority for dedicated CPU cycles. Enabling this feature might help to increase system utilization, without compromising the computing power for critical workloads in a dedicated processor.

Shared mode

In shared mode, logical partitions use virtual processors to access fractions of physical processors. Shared partitions can define any number of virtual processors (the maximum number is 10 times the number of processing units assigned to the partition). From the POWER Hypervisor point of view, virtual processors represent dispatching objects. The POWER Hypervisor dispatches virtual processors to physical processors according to the partition's processing units entitlement. One processing unit represents one physical processor's processing capacity. At the end of the POWER Hypervisor's dispatch cycle (10 ms), all partitions should receive total CPU time equal to their processing unit's entitlement. The logical processors are defined on top of virtual processors. So, even with a virtual processor, the concept of a logical processor exists and the number of logical processors depends on whether the simultaneous multithreading is turned on or off.

3.4.3 Multiple Shared Processor Pools

Multiple Shared Processor Pools (MSPPs) is a capability supported on POWER7 processor and POWER6 processor based servers. This capability allows a system administrator to create a set of micro-partitions with the purpose of controlling the processor capacity that can be consumed from the physical Shared Processor Pool.

To implement MSPPs, there is a set of underlying techniques and technologies. Figure 3-8 shows an overview of the architecture of multiple Shared Processor Pools.

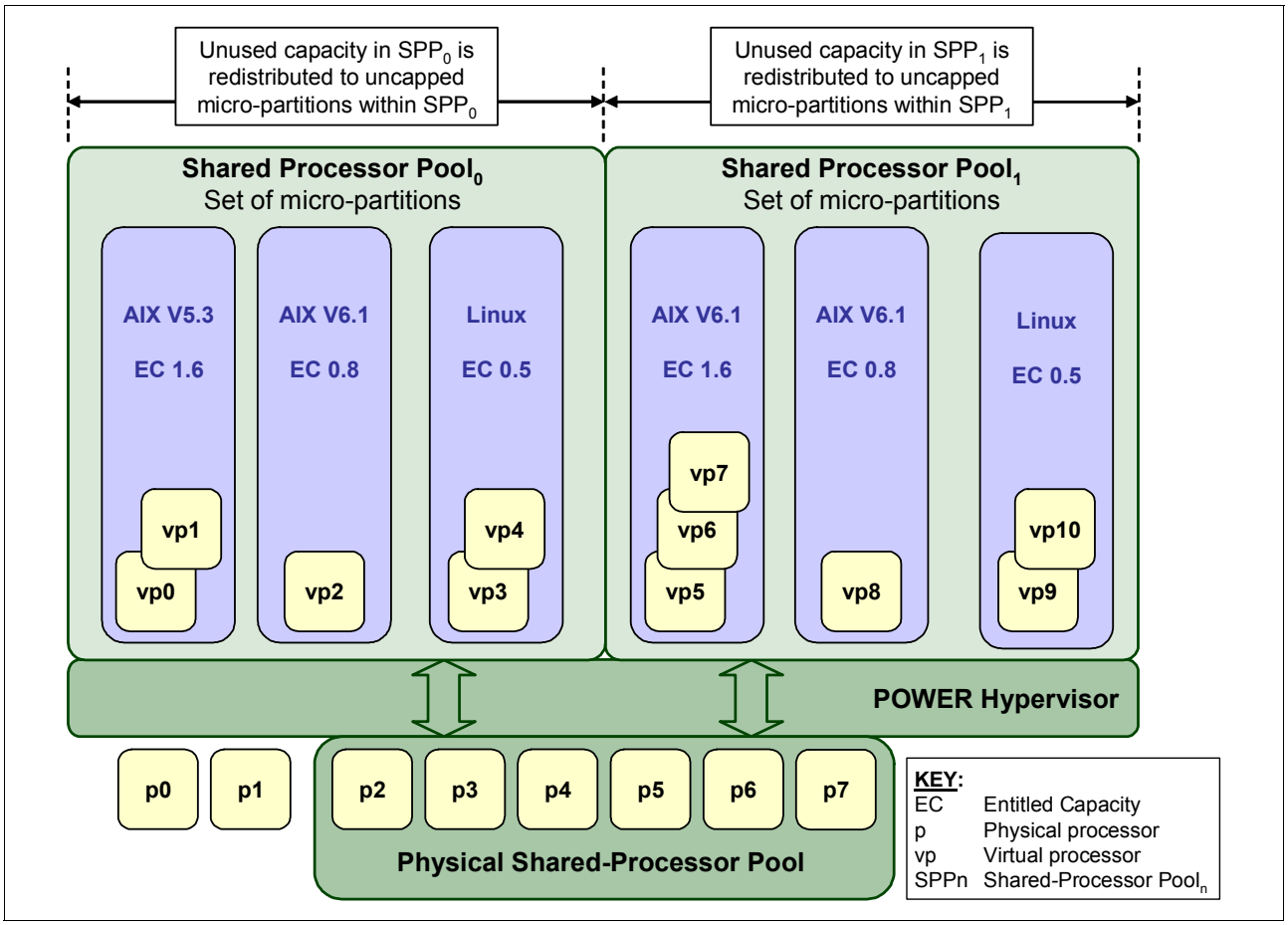


Figure 3-8 Overview of the architecture of multiple Shared Processor Pools

Micro-partitions are created and then identified as members of either the default Shared Processor Pool$_0$ or a user-defined Shared Processor Pool$_n$. The virtual processors that exist within the set of micro-partitions are monitored by the POWER Hypervisor, and processor capacity is managed according to user-defined attributes.

If the Power Systems server is under heavy load, each micro-partition within a Shared Processor Pool is guaranteed its processor entitlement plus any capacity that it might be allocated from the reserved pool capacity if the micro-partition is uncapped.

If some micro-partitions in a Shared Processor Pool do not use their capacity entitlement, the unused capacity is ceded and other uncapped micro-partitions within the same Shared Processor Pool are allocated the additional capacity according to their uncapped weighting. In this way, the entitled pool capacity of a Shared Processor Pool is distributed to the set of micro-partitions within that Shared Processor Pool.

All Power Systems servers that support the multiple Shared Processor Pools capability will have a minimum of one (the default) Shared Processor Pool and up to a maximum of 64 Shared Processor Pools.

Default Shared Processor Pool (SPP_0)

On any Power Systems server supporting multiple Shared Processor Pools, a default Shared Processor Pool is always automatically defined. The default Shared Processor Pool has a pool identifier of zero (SPP-ID = 0) and can also be referred to as SPP_0. The default Shared Processor Pool has the same attributes as a user-defined Shared Processor Pool except that these attributes are not directly under the control of the system administrator. They have fixed values (Table 3-4).

Table 3-4 Attribute values for the default Shared Processor Pool (SPP_0)

SPP_0 attribute	**Value**
Shared Processor Pool ID	0.
Maximum pool capacity	The value is equal to the capacity in the physical Shared Processor Pool.
Reserved pool capacity	0.
Entitled pool capacity	Sum (total) of the entitled capacities of the micro-partitions in the default Shared Processor Pool.

Creating multiple Shared Processor Pools

The default Shared Processor Pool (SPP_0) is automatically activated by the system and is always present.

All other Shared Processor Pools exist, but by default they are inactive. By changing the maximum pool capacity of a Shared Processor Pool to a value greater than zero, it becomes active and can accept micro-partitions (either transferred from SPP_0 or newly created).

Levels of processor capacity resolution

The levels of processor capacity resolution implemented by the POWER Hypervisor and multiple Shared Processor Pools are:

- $Level_0$

  The first level, $Level_0$, is the resolution of capacity within the same Shared Processor Pool. Unused processor cycles from within a Shared Processor Pool are harvested and then redistributed to any eligible micro-partition within the same Shared Processor Pool.

- $Level_1$

  This is the second level of processor capacity resolution. When all $Level_0$ capacity has been resolved within the multiple Shared Processor Pools, the POWER Hypervisor harvests unused processor cycles and redistributes them to eligible micro-partitions regardless of the multiple Shared Processor Pools structure.

Figure 3-9 shows the levels of unused capacity redistribution implemented by the POWER Hypervisor.

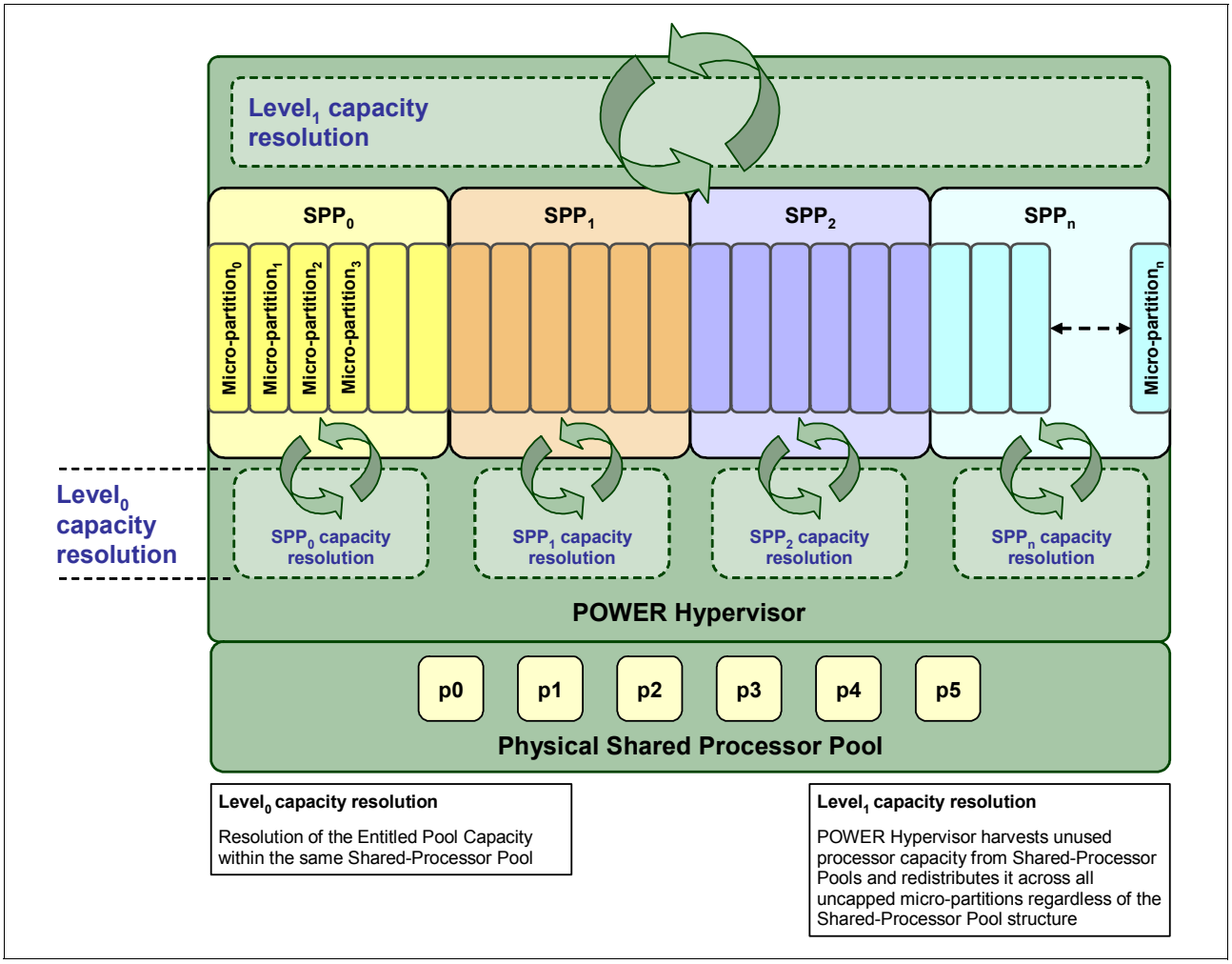


Figure 3-9 The levels of unused capacity redistribution

Capacity allocation above the entitled pool capacity ($Level_1$)

The POWER Hypervisor initially manages the entitled pool capacity at the Shared Processor Pool level. This is where unused processor capacity within a Shared Processor Pool is harvested and then redistributed to uncapped micro-partitions within the same Shared Processor Pool. This level of processor capacity management is sometimes referred to as $Level_0$ capacity resolution.

At a higher level, the POWER Hypervisor harvests unused processor capacity from the multiple Shared Processor Pools that do not consume all of their entitled pool capacity. If a particular Shared Processor Pool is heavily loaded and several of the uncapped micro-partitions within it require additional processor capacity (above the entitled pool capacity), then the POWER Hypervisor redistributes some of the extra capacity to the uncapped micro-partitions. This level of processor capacity management is sometimes referred to as $Level_1$ capacity resolution.

To redistribute unused processor capacity to uncapped micro-partitions in multiple Shared Processor Pools above the entitled pool capacity, the POWER Hypervisor uses a higher level of redistribution, $Level_1$.

Important: $Level_1$ capacity resolution: When allocating additional processor capacity in excess of the entitled pool capacity of the Shared Processor Pool, the POWER Hypervisor takes the uncapped weights of *all micro-partitions in the system* into account, *regardless of the multiple Shared Processor Pool structure*.

Where there is unused processor capacity in under-utilized Shared Processor Pools, the micro-partitions within the Shared Processor Pools cede the capacity to the POWER Hypervisor.

In busy Shared Processor Pools, where the micro-partitions have used all of the entitled pool capacity, the POWER Hypervisor allocates additional cycles to micro-partitions, in which *all* of these statements are true:

- The maximum pool capacity of the Shared Processor Pool hosting the micro-partition has not been met.
- The micro-partition is uncapped.
- The micro-partition has enough virtual-processors to take advantage of the additional capacity.

Under these circumstances, the POWER Hypervisor allocates additional processor capacity to micro-partitions on the basis of their uncapped weights independent of the Shared Processor Pool hosting the micro-partitions. This can be referred to as $Level_1$ capacity resolution. Consequently, when allocating additional processor capacity in excess of the entitled pool capacity of the Shared Processor Pools, the POWER Hypervisor takes the uncapped weights of all micro-partitions in the system into account, regardless of the Multiple Shared Processor Pools structure.

Dynamic adjustment of maximum pool capacity

The maximum pool capacity of a Shared Processor Pool, other than the default Shared Processor $Pool_0$, can be adjusted dynamically from the managed console, using either the graphical or command-line interface (CLI).

Dynamic adjustment of Reserve Pool Capacity

The reserved pool capacity of a Shared Processor Pool, other than the default Shared Processor $Pool_0$, can be adjusted dynamically from the managed console, using either the graphical or CLI interface.

Dynamic movement between Shared Processor Pools

A micro-partition can be moved dynamically from one Shared Processor Pool to another using the managed console using either the graphical or CLI interface. Because the entitled pool capacity is partly made up of the sum of the entitled capacities of the micro-partitions, removing a micro-partition from a Shared Processor Pool reduces the entitled pool capacity for that Shared Processor Pool. Similarly, the entitled pool capacity of the Shared Processor Pool that the micro-partition joins will increase.

Deleting a Shared Processor Pool

Shared Processor Pools cannot be deleted from the system. However, they are deactivated by setting the maximum pool capacity and the reserved pool capacity to zero. The Shared Processor Pool will still exist but will not be active. Use the managed console interface to deactivate a Shared Processor Pool. A Shared Processor Pool cannot be deactivated unless all micro-partitions hosted by the Shared Processor Pool have been removed.

Live Partition Mobility and Multiple Shared Processor Pools

A micro-partition might leave a Shared Processor Pool because of PowerVM Live Partition Mobility. Similarly, a micro-partition might join a Shared Processor Pool in the same way. When performing PowerVM Live Partition Mobility, you are given the opportunity to designate a destination Shared Processor Pool on the target server to receive and host the migrating micro-partition.

Because several simultaneous micro-partition migrations are supported by PowerVM Live Partition Mobility, it is conceivable to migrate the entire Shared Processor Pool from one server to another.

3.4.4 Virtual I/O Server

The Virtual I/O Server is part of all PowerVM Editions. It is a special-purpose partition that allows the sharing of physical resources between logical partitions to allow more efficient utilization (for example, consolidation). In this case, the Virtual I/O Server owns the physical resources (SCSI, Fibre Channel, network adapters, and optical devices) and allows client partitions to share access to them, thus minimizing the number of physical adapters in the system. The Virtual I/O Server eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The Virtual I/O Server supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses. Figure 3-10 shows an overview of a Virtual I/O Server configuration.

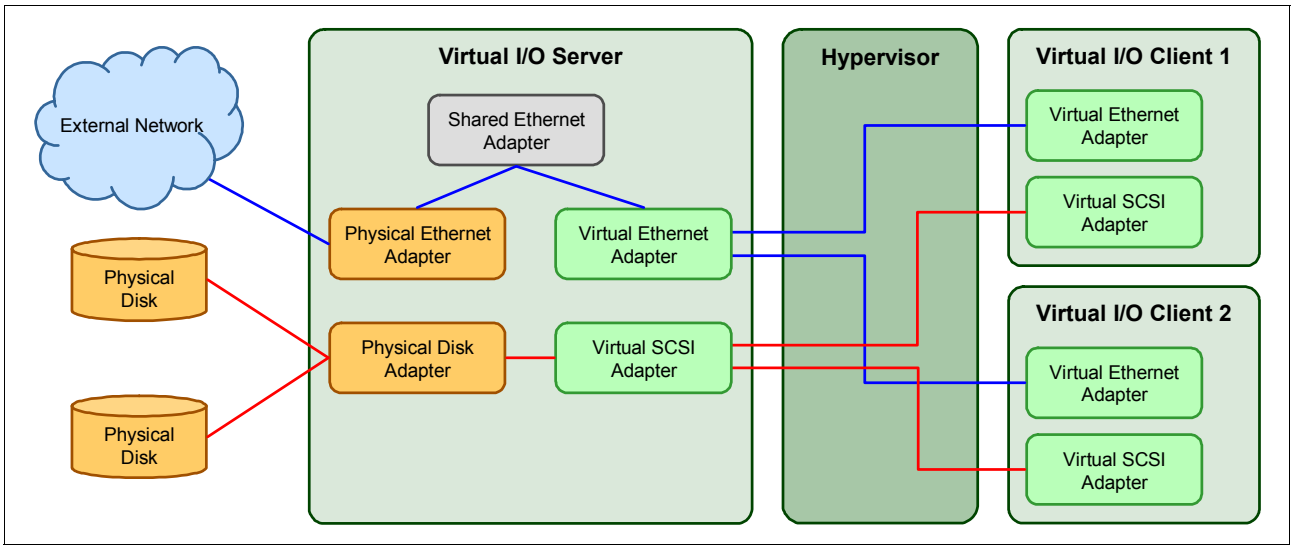


Figure 3-10 Architectural view of the Virtual I/O Server

Because the Virtual I/O server is an operating system-based appliance server, 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 system backup DVD that is provided to clients who order any PowerVM edition. This dedicated software is only for the Virtual I/O Server (and IVM in case it is used) and is only supported in special Virtual I/O Server partitions. Three major virtual devices are supported by the Virtual I/O Server:

- Shared Ethernet Adapter
- Virtual SCSI
- Virtual Fibre Channel adapter

The Virtual Fibre Channel adapter is used with the NPIV feature, described in 3.4.8, “N_Port ID virtualization” on page 127.

Shared Ethernet Adapter

A Shared Ethernet Adapter (SEA) can be used to connect a physical Ethernet network to a virtual Ethernet network. The Shared Ethernet Adapter provides this access by connecting the internal Hypervisor VLANs with the VLANs on the external switches. Because the Shared Ethernet Adapter processes packets at layer 2, the original MAC address and VLAN tags of the packet are visible to other systems on the physical network. IEEE 802.1 VLAN tagging is supported.

The Shared Ethernet Adapter also provides the ability for several client partitions to share one physical adapter. With an SEA, you can connect internal and external VLANs using a physical adapter. The Shared Ethernet Adapter service can only be hosted in the Virtual I/O Server, not in a general-purpose AIX or Linux partition, and acts as a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (EtherChannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the POWER Hypervisor on the Virtual I/O Server

Tip: A Linux partition can provide a bridging function also by using the `brctl` command.

Figure 3-11 shows a configuration example of an SEA with one physical and two virtual Ethernet adapters. An SEA can include up to 16 virtual Ethernet adapters on the Virtual I/O Server that share the same physical access.

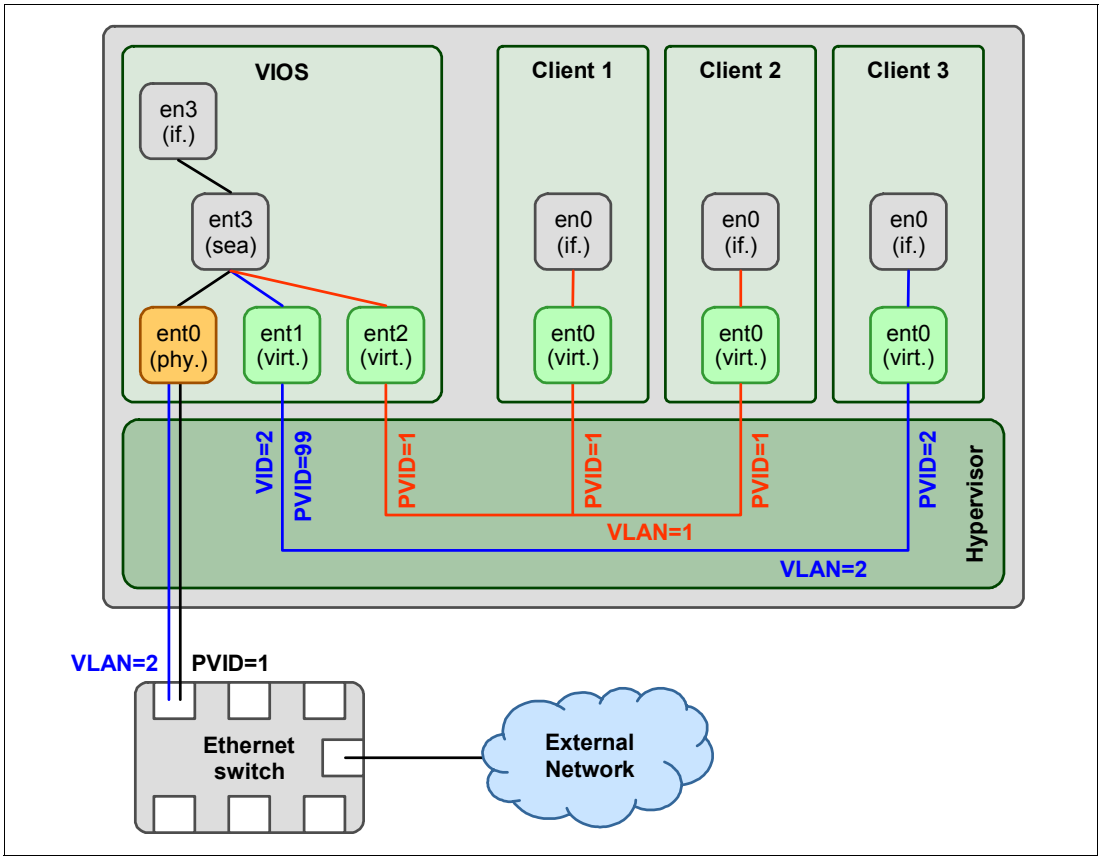


Figure 3-11 Architectural view of a Shared Ethernet Adapter

A single SEA setup can have up to 16 Virtual Ethernet trunk adapters, and each virtual Ethernet trunk adapter can support up to 20 VLAN networks. Therefore, a possibility is for a single physical Ethernet to be shared between 320 internal VLAN networks. The number of shared Ethernet adapters that can be set up in a Virtual I/O Server partition is limited only by the resource availability, because there are no configuration limits.

Unicast, broadcast, and multicast are supported, so protocols that rely on broadcast or multicast, such as Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), Boot Protocol (BOOTP), and Neighbor Discovery Protocol (NDP) can work on an SEA.

Note: A Shared Ethernet Adapter does not need to have an IP address configured to be able to perform the Ethernet bridging functionality. Configuring IP on the Virtual I/O Server is convenient because the Virtual I/O Server can then be reached by TCP/IP, for example, to perform dynamic LPAR operations or to enable remote login. This task can be done either by configuring an IP address directly on the SEA device or on an additional virtual Ethernet adapter in the Virtual I/O Server. This leaves the SEA without the IP address, allowing for maintenance on the SEA without losing IP connectivity in case SEA failover is configured.

For a more detailed discussion about virtual networking, see:

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

Virtual SCSI

Virtual SCSI is used to refer to a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a client/server relationship. The Virtual I/O Server logical partition owns the physical resources and acts as a server or, in SCSI terms, a target device. The client logical partitions access the virtual SCSI backing storage devices provided by the Virtual I/O Server as clients.

The virtual I/O adapters (virtual SCSI server adapter and a virtual SCSI client adapter) are configured using a managed console or through the Integrated Virtualization Manager on smaller systems. The virtual SCSI server (target) adapter is responsible for executing any SCSI commands that it receives. It is owned by the Virtual I/O Server partition. The virtual SCSI client adapter allows a client partition to access physical SCSI and SAN-attached devices and LUNs that are assigned to the client partition. The provisioning of virtual disk resources is provided by the Virtual I/O Server.

Physical disks presented to the Virtual/O Server can be exported and assigned to a client partition in a number of ways:

- The entire disk is presented to the client partition.
- The disk is divided into several logical volumes, which can be presented to a single client or multiple clients.
- As of Virtual I/O Server 1.5, files can be created on these disks, and file-backed storage devices can be created.

The logical volumes or files can be assigned to separate partitions. Therefore, virtual SCSI enables sharing of adapters and disk devices.

Figure 3-12 shows an example where one physical disk is divided into two logical volumes by the Virtual I/O Server. Each of the two client partitions is assigned one logical volume, which is then accessed through a virtual I/O adapter (VSCSI Client Adapter). Inside the partition, the disk is seen as a normal hdisk.

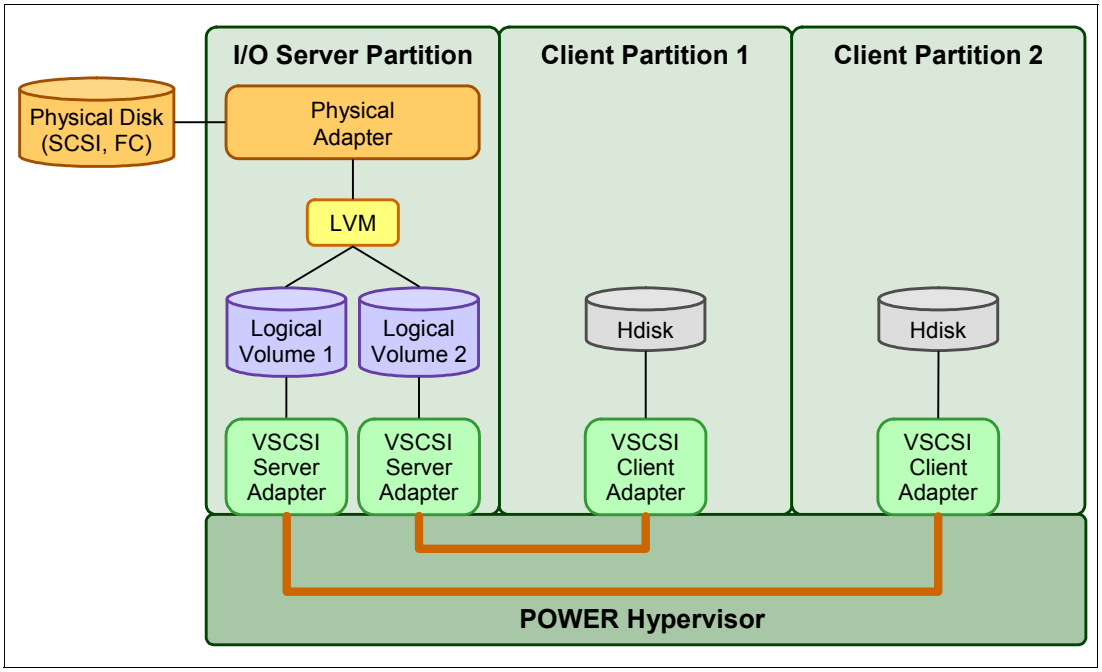


Figure 3-12 Architectural view of virtual SCSI

At the time of writing, virtual SCSI supports Fibre Channel, parallel SCSI, iSCSI, SAS, SCSI RAID devices and optical devices, including DVD-RAM and DVD-ROM. Other protocols such as SSA and tape devices are not supported.

For more information about the specific storage devices supported for Virtual I/O Server, see:

http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html

Virtual I/O Server functions

The Virtual I/O Server has a number of features, including monitoring solutions:

- Support for Live Partition Mobility starting on POWER6 processor-based systems with the PowerVM Enterprise Edition. For more information about Live Partition Mobility, see 3.4.5, “PowerVM Live Partition Mobility” on page 121.
- Support for virtual SCSI devices backed by a file, which are then accessed as standard SCSI-compliant LUNs.
- Support for virtual Fibre Channel devices that are used with the NPIV feature.
- Virtual I/O Server Expansion Pack with additional security functions such as Kerberos (Network Authentication Service for users and Client and Server Applications), Simple Network Management Protocol (SNMP) v3, and Lightweight Directory Access Protocol (LDAP) client functionality.
- System Planning Tool (SPT) and Workload Estimator, which are designed to ease the deployment of a virtualized infrastructure. For more information about System Planning Tool, see 3.5, “System Planning Tool” on page 130.

- Includes IBM Systems Director agent and a number of pre-installed Tivoli agents, such as these:
  - Tivoli Identity Manager (TIM), to allow easy integration into an existing Tivoli Systems Management infrastructure
  - Tivoli Application Dependency Discovery Manager (ADDM), which creates and automatically maintains application infrastructure maps including dependencies, change-histories, and deep configuration values
- vSCSI eRAS.
- Additional CLI statistics in **`svmon`**, **`vmstat`**, **`fcstat`**, and **`topas`**.
- Monitoring solutions to help manage and monitor the Virtual I/O Server and shared resources. New commands and views provide additional metrics for memory, paging, processes, Fibre Channel HBA statistics, and virtualization.

For more information about the Virtual I/O Server and its implementation, see *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940.

3.4.5 PowerVM Live Partition Mobility

PowerVM Live Partition Mobility allows you to move a running logical partition, including its operating system and running applications, from one system to another without any shutdown and without disrupting the operation of that logical partition. Inactive partition mobility allows you to move a powered-off logical partition from one system to another.

Partition mobility provides systems management flexibility and improves system availability, as follows:

- Avoid planned outages for hardware or firmware maintenance by moving logical partitions to another server and then performing the maintenance. Live Partition Mobility can help lead to zero downtime maintenance because you can use it to work around scheduled maintenance activities.
- Avoid downtime for a server upgrade by moving logical partitions to another server and then performing the upgrade. This approach allows your users to continue their work without disruption.
- Avoid unplanned downtime. With preventive failure management, if a server indicates a potential failure, you can move its logical partitions to another server before the failure occurs. Partition mobility can help avoid unplanned downtime.
- Take advantage of server optimization:
  - Consolidation: You can consolidate workloads running on several small, under-used servers onto a single large server.
  - Deconsolidation: You can move workloads from server to server to optimize resource use and workload performance within your computing environment. With active partition mobility, you can manage workloads with minimal downtime.

Mobile partition's operating system requirements

The operating system running in the mobile partition has to be AIX or Linux. The Virtual I/O Server partition itself cannot be migrated. All versions of AIX and Linux supported on the IBM POWER7 processor-based servers also support partition mobility.

Source and destination system requirements

The source partition must be one that has only virtual devices. If there are any physical devices in its allocation, they must be removed before the validation or migration is initiated. An N_Port ID virtualization (NPIV) device is considered virtual and is compatible with partition migration.

The hypervisor must support the Partition Mobility functionality, also called migration process, available on POWER 6 and POWER 7 processor-based hypervisors. Firmware must be at firmware level eFW3.2 or later. All POWER7 processor-based hypervisors support Live Partition Mobility. Source and destination systems can have separate firmware levels, but they must be compatible with each other.

A option is to migrate partitions back and forth between POWER6 and POWER7 processor-based servers. Partition Mobility leverages the POWER6 Compatibility Modes that are provided by POWER7 processor-based servers. On the POWER7 processor-based server, the migrated partition is then executing in POWER6 or POWER6+ Compatibility Mode.

If you want to move an active logical partition from a POWER6 processor-based server to a POWER7 processor-based server so that the logical partition can take advantage of the additional capabilities available with the POWER7 processor, you can perform these steps:

1. Set the partition-preferred processor compatibility mode to the default mode. When you activate the logical partition on the POWER6 processor-based server, it runs in the POWER6 mode.
2. Move the logical partition to the POWER7 processor-based server. Both the current and preferred modes remain unchanged for the logical partition until you restart the logical partition.
3. Restart the logical partition on the POWER7 processor-based server. The hypervisor evaluates the configuration. Because the preferred mode is set to default and the logical partition now runs on a POWER7 processor-based server, the highest mode available is the POWER7 mode. The hypervisor determines that the most fully featured mode that is supported by the operating environment installed in the logical partition is the POWER7 mode and changes the current mode of the logical partition to the POWER7 mode.

Now the current processor compatibility mode of the logical partition is the POWER7 mode and the logical partition runs on the POWER7 processor-based server.

Tip: The “Migration combinations of processor compatibility modes for active Partition Mobility” web page offers presentations about the supported migrations:

`http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm`

The Virtual I/O Server on the source system provides the access to the client resources and must be identified as a mover service partition (MSP). The Virtual Asynchronous Services Interface (VASI) device allows the MSP to communicate with the hypervisor. It is created and managed automatically by the managed console and will be configured on both the source and destination Virtual I/O Servers, which are designated as the mover service partitions for the mobile partition, to participate in active mobility. Other requirements include a similar time-of-day on each server, systems must not be running on battery power, and shared storage (external hdisk with `reserve_policy=no_reserve`). In addition, all logical partitions must be on the same open network with RMC established to the managed console.

The managed console is used to configure, validate, and orchestrate. Use the managed console to configure the Virtual I/O Server as an MSP and to configure the VASI device. A managed console wizard validates your configuration and identifies issues that can cause the migration to fail. During the migration, the managed console controls all phases of the process.

Improved Live Partition Mobility benefits

The possibility to move partitions between POWER6 and POWER7 processor-based servers greatly facilitates the deployment of POWER7 processor-based servers, as follows:

- Installation of the new server can be performed while the application is executing on a POWER6 server. After the POWER7 processor-based server is ready, the application can be migrated to its new hosting server without application downtime.
- When adding POWER7 processor-based servers to a POWER6 environment, you get the additional flexibility to perform workload balancing across the entire set of POWER6 and POWER7 processor-based servers.
- When performing server maintenance, you get the additional flexibility to use POWER6 servers for hosting applications usually hosted on POWER7 processor-based servers, and vice-versa, allowing you to perform this maintenance with no application planned downtime.

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

3.4.6 Active Memory Sharing

Active Memory Sharing is an IBM PowerVM advanced memory virtualization technology that provides system memory virtualization capabilities to IBM Power Systems, allowing multiple partitions to share a common pool of physical memory.

Active Memory Sharing is only available with the Enterprise version of PowerVM.

The physical memory of an IBM Power System can be assigned to multiple partitions either in a dedicated or in a shared mode. The system administrator has the capability to assign some physical memory to a partition and some physical memory to a pool that is shared by other partitions. A single partition can have either dedicated or shared memory:

- With a pure dedicated memory model, the system administrator's task is to optimize available memory distribution among partitions. When a partition suffers degradation because of memory constraints, and other partitions have unused memory, the administrator can manually issue a dynamic memory reconfiguration.
- With a shared memory model, the system automatically decides the optimal distribution of the physical memory to partitions and adjusts the memory assignment based on partition load. The administrator reserves physical memory for the shared memory pool, assigns partitions to the pool, and provides access limits to the pool.

Active Memory Sharing can be exploited to increase memory utilization on the system either by decreasing the global memory requirement or by allowing the creation of additional partitions on an existing system. Active Memory Sharing can be used in parallel with Active Memory Expansion on a system running a mixed workload of several operating systems. For example, AIX partitions can take advantage of Active Memory Expansion. Other operating systems take advantage of Active Memory Sharing.

For additional information regarding Active Memory Sharing, see *IBM PowerVM Virtualization Active Memory Sharing*, REDP-4470.

3.4.7 Active Memory Deduplication

In a virtualized environment, the systems might have a considerable amount of duplicated information stored on its RAM after each partition has its own operating system, and some of them might even share the same kind of applications. On heavily loaded systems, this might lead to a shortage of the available memory resources, forcing paging by the AMS partition operating systems, the AMD pool, or both, which might decrease overall system performance.

Figure 3-13 shows the standard behavior of a system without Active Memory Deduplication (AMD) enabled on its AMS shared memory pool. Identical pages within the same or different LPARs each require their own unique physical memory page, consuming space with repeated information.

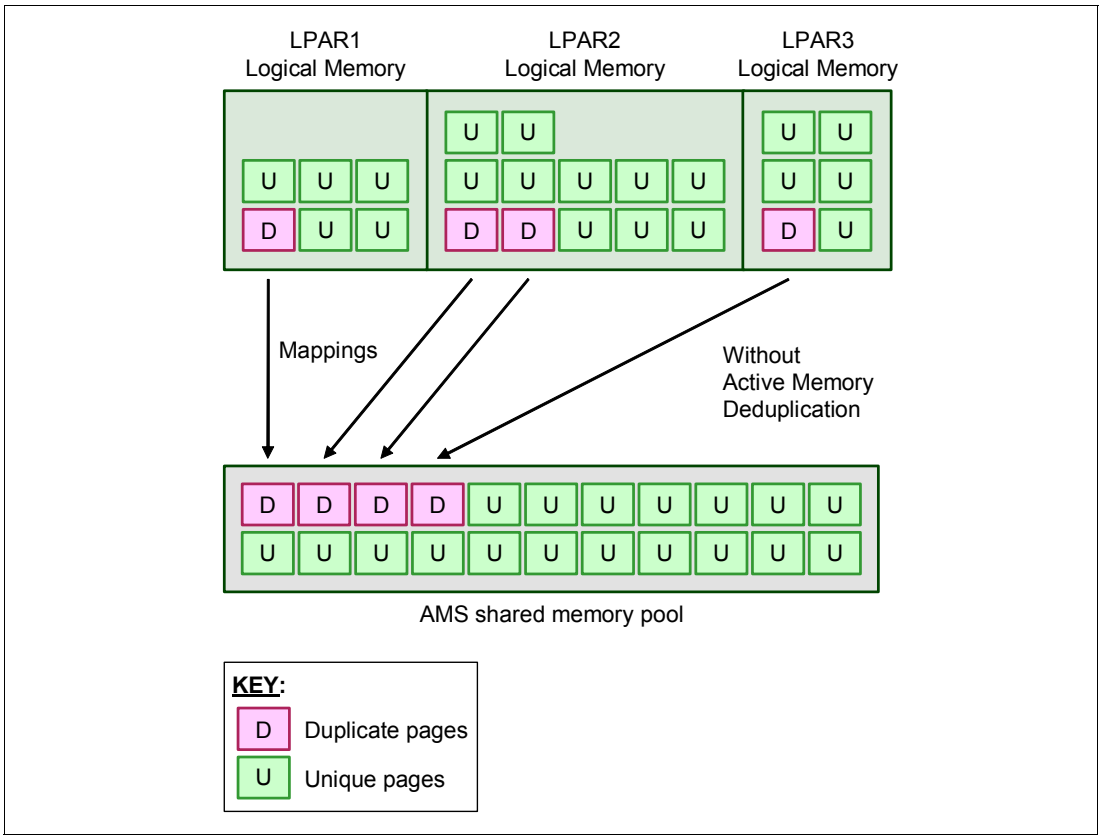


Figure 3-13 AMS shared memory pool without AMD enabled

Active Memory Deduplication allows the Hypervisor to dynamically map identical partition memory pages to a single physical memory page within a shared memory pool. This enables a better utilization of the AMS shared memory pool, increasing the system's overall performance by avoiding paging. Deduplication can cause the hardware to incur fewer cache misses, which also leads to improved performance.

Figure 3-14 shows the behavior of a system with Active Memory Deduplication enabled on its AMS shared memory pool. Duplicated pages from different LPARs are stored just once, providing the AMS pool with more free memory.

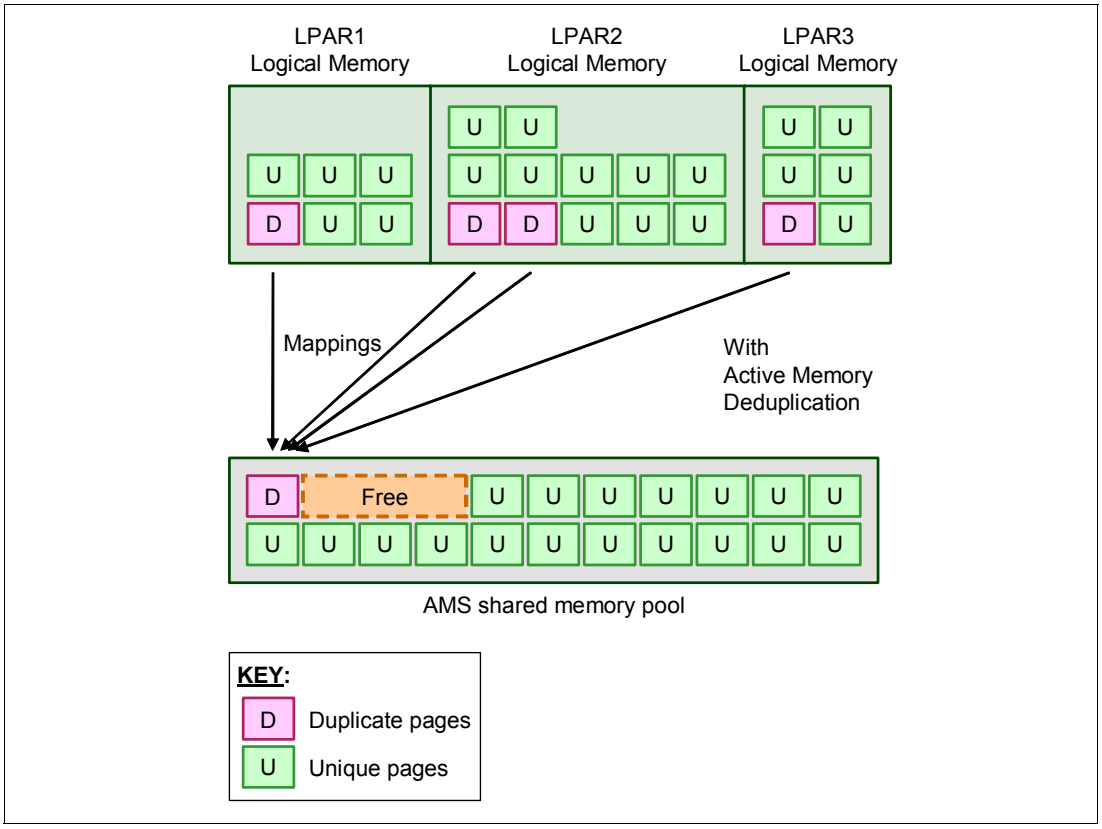


Figure 3-14 Identical memory pages mapped to a single physical memory page with Active Memory Duplication enabled

AMD depends on the Active Memory Sharing (AMS) feature to be available, and it consumes CPU cycles donated by the AMS pool's VIOS partitions to identify deduplicated pages. The operating systems running on the AMS partitions can hint to the PowerVM Hypervisor that certain pages (such as frequently referenced read-only code pages) are particularly good for deduplication.

To perform deduplication, the Hypervisor cannot compare every memory page in the AMS pool with every other page. Instead, it computes a small signature for each page that it visits, and stores the signatures in an internal table. Each time that a page is inspected, its signature is looked up against the known signatures in the table. If a match is found, the memory pages are compared to be sure that the pages are really duplicates. When an actual duplicate is found, the Hypervisor remaps the partition memory to the existing memory page and returns the duplicate page to the AMS pool.

Figure 3-15 shows two pages being written in the AMS memory pool and having its signatures matched on the deduplication table.

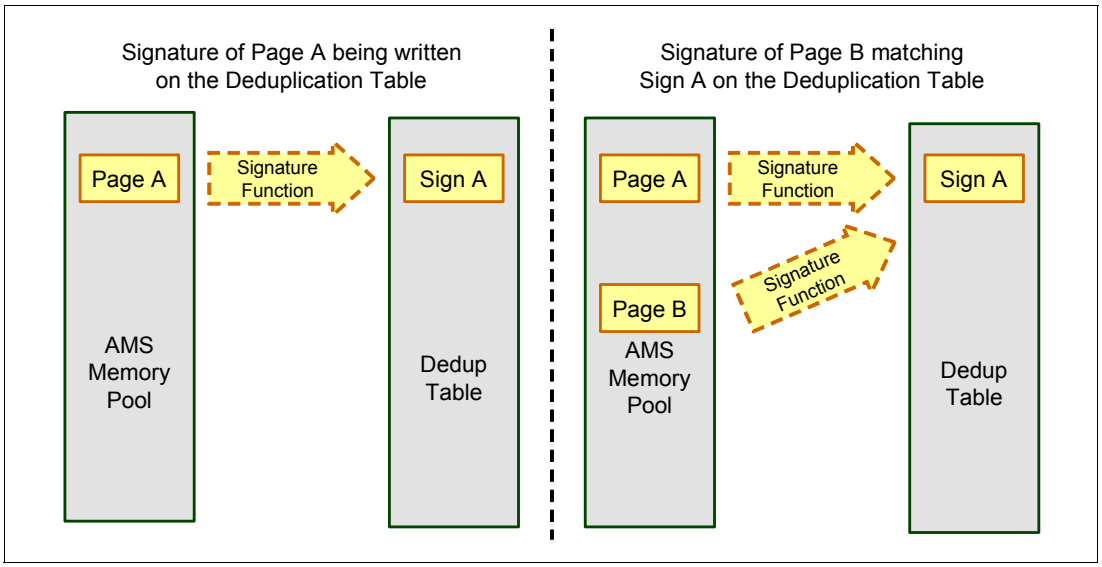


Figure 3-15 Memory pages having their signatures matched by Active Memory Deduplication

From the LPAR point of view, the AMD feature is completely transparent. If an LPAR attempts to modify a deduplicated page, the Hypervisor takes a free page from the AMS pool, copies the duplicate page content into the new page, and maps the LPAR's reference to the new page so that the LPAR can modify its own unique page.

System administrators can dynamically configure the size of the deduplication table, ranging from 1/8192 up to 1/256 of the configured maximum AMS memory pool size. Having this table too small might lead to missed deduplication opportunities. Conversely, having a table that is too large might waste a small amount of overhead space.

The management of the Active Memory Deduplication feature is done via a managed console, allowing administrators to:

- Enable and disable Active Memory Deduplication at an AMS Pool level.
- Display deduplication metrics.
- Display and modify the deduplication table size.

Figure 3-16 shows the Active Memory Deduplication function being enabled to a shared memory pool.

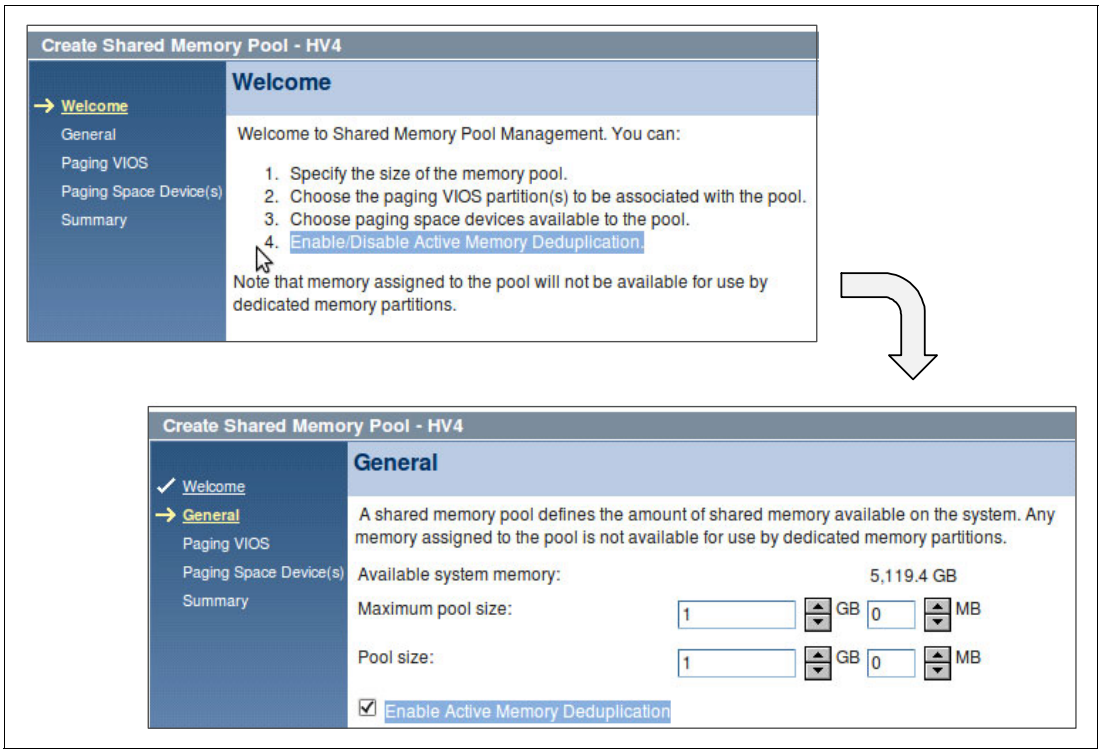


Figure 3-16 Enabling the Active Memory Deduplication for a shared memory pool

The Active Memory Deduplication feature requires these minimum components:

- PowerVM Enterprise edition
- System firmware level 740
- AIX Version 6: AIX 6.1 TL7 or later
- AIX Version 7: AIX 7.1 TL1 SP1 or later
- IBM i: 7.14 or 7.2 or later
- SLES 11 SP2 or later
- RHEL 6.2 or later

3.4.8 N_Port ID virtualization

N_Port ID virtualization (NPIV) is a technology that allows multiple logical partitions to access independent physical storage through the same physical Fibre Channel adapter. This adapter is attached to a Virtual I/O Server partition that acts only as a pass-through, managing the data transfer through the POWER Hypervisor.

Each partition using NPIV is identified by a pair of unique worldwide port names, enabling you to connect each partition to independent physical storage on a SAN. Unlike virtual SCSI, only the client partitions see the disk.

For additional information and requirements for NPIV, see:

- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

NPIV is supported in PowerVM Express, Standard, and Enterprise Editions, on the IBM Power 720 and Power 740 servers.

3.4.9 Operating system support for PowerVM

Table 3-5 summarizes the PowerVM features supported by the operating systems compatible with the POWER7 processor-based servers.

Table 3-5 PowerVM features supported by AIX, IBM i and Linux

Feature	AIX V5.3	AIX V6.1	AIX V7.1	IBM i 6.1.1	IBM i 7.1	RHEL V5.7	RHEL V6.1	SLES V10 SP4	SLES V11 SP1
Virtual SCSI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Ethernet	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shared Ethernet Adapter	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Fibre Channel	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual Tape	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Logical Partitioning	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O adapter add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O processor add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O memory add	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR I/O memory remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Micro-Partitioning	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shared Dedicated Capacity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multiple Shared Processor Pools	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Virtual I/O Server	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Suspend/Resume	No	Yes	Yes	No	No	No	No	No	No
Shared Storage Pools	Yes	Yes	Yes	Yes	Yes[a]	No	No	No	No
Thin Provisioning	Yes	Yes	Yes	Yes[b]	Yes[b]	No	No	No	No
Active Memory Sharing and Active Memory Deduplication	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes

Feature	AIX V5.3	AIX V6.1	AIX V7.1	IBM i 6.1.1	IBM i 7.1	RHEL V5.7	RHEL V6.1	SLES V10 SP4	SLES V11 SP1
Live Partition Mobility	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Simultaneous Multi-Threading (SMT)	Yes[c]	Yes[d]	Yes	Yes[e]	Yes	Yes[c]	Yes[c]	Yes[c]	Yes
Active Memory Expansion	No	Yes[f]	Yes	No	No	No	No	No	No

a. Requires IBM i 7.1 TR1.
b. Will become a fully provisioned device when used by IBM i.
c. Only supports two threads.
d. AIX 6.1 up to TL4 SP2 only supports two threads, and supports four threads as of TL4 SP3.
e. IBM i 6.1.1 and later support SMT4.
f. On AIX 6.1 with TL4 SP2 and later.

3.4.10 POWER7 Linux programming support

IBM Linux Technology Center (LTC) contributes to the development of Linux by providing support for IBM hardware in Linux distributions. In particular, the LTC makes tools and code available to the Linux communities to take advantage of the POWER7 technology and develop POWER7 optimized software.

Table 3-6 lists the support of specific programming features for various versions of Linux.

Table 3-6 Linux support for POWER7 features

Features	Linux releases				Comments
	SLES 10 SP4	SLES 11	RHEL 5.7	RHEL 6.1	
POWER6 compatibility mode	Yes	Yes	Yes	Yes	-
POWER7 mode	No	Yes	No	Yes	-
Strong Access Ordering	No	Yes	No	Yes	Can improve Lx86 performance
Scale to 256 cores/ 1024 threads	No	Yes	No	Yes	Base OS support available
4-way SMT	No	Yes	No	Yes	-
VSX Support	No	Yes	No	Yes	Full exploitation requires Advance Toolchain
Distro toolchain mcpu/mtune=p7	No	Yes	No	Yes	SLES11/GA toolchain has minimal P7 enablement necessary to support kernel build

Features	Linux releases				Comments
	SLES 10 SP4	SLES 11	RHEL 5.7	RHEL 6.1	
Advance Toolchain Support	Yes, execution restricted to Power6 instructions	Yes	Yes, execution restricted to Power6 instructions	Yes	Alternative IBM GNU Toolchain
64k base page size	No	Yes	Yes	Yes	-
Tickless idle	No	Yes	No	Yes	Improved energy utilization and virtualization of partially to fully idle partitions.

For information regarding Advance Toolchain, go to the following address:

http://www.ibm.com/developerworks/wikis/display/hpccentral/How+to+use+Advance+Toolchain+for+Linux+on+POWER

You can also visit the University of Illinois Linux on Power Open Source Repository:

- http://ppclinux.ncsa.illinois.edu
- ftp://linuxpatch.ncsa.uiuc.edu/toolchain/at/at05/suse/SLES_11/release_notes.at05-2.1-0.html
- ftp://linuxpatch.ncsa.uiuc.edu/toolchain/at/at05/redhat/RHEL5/release_notes.at05-2.1-0.html

3.5 System Planning Tool

The IBM System Planning Tool (SPT) helps you design systems to be partitioned with logical partitions. You can also plan for and design non-partitioned systems by using the SPT. The resulting output of your design is called a *system plan*, which is stored in a `.sysplan` file. This file can contain plans for a single system or multiple systems. The `.sysplan` file can be used for these reasons:

- To create reports
- As input to the IBM configuration tool (e-Config)
- To create and deploy partitions on your system (or systems) automatically

System plans that are generated by the SPT can be deployed on the system by the Hardware Management Console (HMC), Systems Director Management Console (SDMC), or Integrated Virtualization Manager (IVM).

Note: Ask your IBM representative or Business Partner to use the Customer Specified Placement manufacturing option if you want to automatically deploy your partitioning environment on a new machine. SPT looks for the resource's allocation to be the same as that specified in your `.sysplan` file.

You can create an entirely new system configuration, or you can create a system configuration based on any of the following items:

- Performance data from an existing system that the new system is to replace
- Performance estimates that anticipate future workloads that you must support
- Sample systems that you can customize to fit your needs

Integration between the SPT and both the Workload Estimator (WLE) and IBM Performance Management (PM) allows you to create a system that is based on performance and capacity data from an existing system or that is based on new workloads that you specify.

You can use the SPT before you order a system to determine what you must order to support your workload. You can also use the SPT to determine how you can partition a system that you already have.

The System Planning tool is an effective way of documenting and backing up key system settings and partition definitions. It allows the user to create records of systems and export them to their personal workstation or backup system of choice. These same backups can then be imported back onto the same managed console when needed. This can be useful when cloning systems enabling the user to import the system plan to any managed console multiple times.

The SPT and its supporting documentation can be found on the IBM System Planning Tool site:

http://www.ibm.com/systems/support/tools/systemplanningtool/

4

Continuous availability and manageability

This chapter provides information about IBM reliability, availability, and serviceability (RAS) design and features. This set of technologies, implemented on IBM Power Systems servers, provides the possibility to improve your architecture's total cost of ownership (TCO) by reducing unplanned down time.

The elements of RAS can be described as follows:

- Reliability: Indicates how infrequently a defect or fault in a server manifests itself.
- Availability: Indicates how infrequently the functionality of a system or application is impacted by a fault or defect.
- Serviceability: Indicates how well faults and their effects are communicated to users and services and how efficiently and nondisruptively the faults are repaired.

Each successive generation of IBM servers is designed to be more reliable than the previous server family. POWER7 processor-based servers have new features to support new levels of virtualization, help ease administrative burden, and increase system utilization.

Reliability starts with components, devices, and subsystems designed to be fault-tolerant. POWER7 uses lower voltage technology, improving reliability with stacked latches to reduce soft error susceptibility. 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 to help ensure high product quality levels.

The processor and memory subsystem contain a number of features designed to avoid or correct environmentally induced, single-bit, intermittent failures as well as handle solid faults in components, including selective redundancy to tolerate certain faults without requiring an outage or parts replacement.

4.1 Reliability

Highly reliable systems are built with highly reliable components. On IBM POWER processor-based systems, this basic principle is expanded upon with a clear design for reliability architecture and methodology. A concentrated, systematic, architecture-based approach is designed to improve overall system reliability with each successive generation of system offerings.

4.1.1 Designed for reliability

Systems designed with fewer components and interconnects have fewer opportunities to fail. Simple design choices such as integrating processor cores on a single POWER chip can dramatically reduce the opportunity for system failures. In this case, an 8-core server can include one-fourth as many processor chips (and chip socket interfaces) as with a double CPU-per-processor design. Not only does this case reduce the total number of system components, it reduces the total amount of heat generated in the design, resulting in an additional reduction in required power and cooling components. POWER7 processor-based servers also integrate L3 cache into the processor chip for a higher integration of parts.

Parts selection also plays a critical role in overall system reliability. IBM uses three grades of components, grade 3 defined as industry standard (off-the-shelf). As shown in Figure 4-1, using stringent design criteria and an extensive testing program, the IBM manufacturing team can produce grade 1 components that are expected to be 10 times more reliable than industry standard. Engineers select grade 1 parts for the most critical system components. Newly introduced organic packaging technologies, rated grade 5, achieve the same reliability as grade 1 parts.

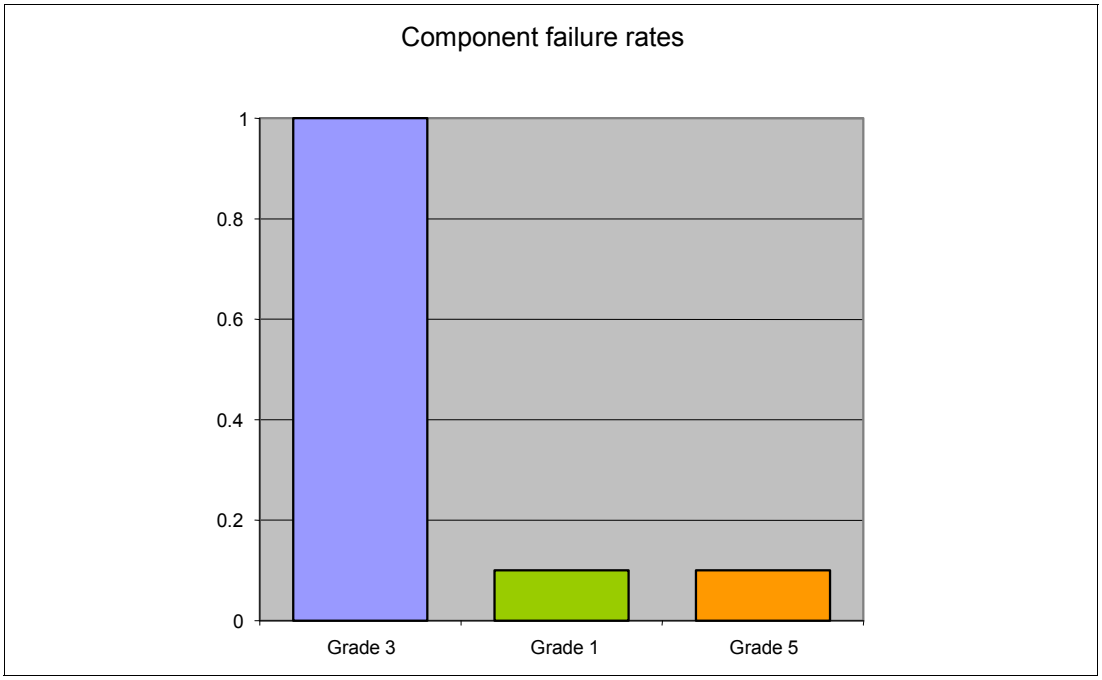


Figure 4-1 Component failure rates

4.1.2 Placement of components

Packaging is designed to deliver both high performance and high reliability. For example, the reliability of electronic components is directly related to their thermal environment. That is, large decreases in component reliability are directly correlated with relatively small increases in temperature. POWER processor-based systems are carefully packaged to ensure adequate cooling. Critical system components such as the POWER7 processor chips are positioned on printed circuit cards so that they receive fresh air during operation. In addition, POWER processor-based systems are built with redundant, variable-speed fans that can automatically increase output to compensate for increased heat in the central electronic complex.

4.1.3 Redundant components and concurrent repair

High opportunity components—those that most affect system availability—are protected with redundancy and the ability to be repaired concurrently.

The use of redundant components allows the system to remain operational:

- POWER7 cores, which include redundant bits in L1 instruction and data caches, L2 caches, and L2 and L3 directories
- Power 720 and Power 740 main memory DIMMs, which use an innovative ECC algorithm from IBM research that improves bit error correction and memory failures
- Redundant and hot-swap cooling
- Redundant and hot-swap power supplies
- Redundant 12X loops to I/O subsystem

For maximum availability, be sure to connect power cords from the same system to two separate Power Distribution Units (PDUs) in the rack, and to connect each PDU to independent power sources. Tower form factor power cords must be plugged into two independent power sources to achieve maximum availability.

Tip: Check your configuration for optional redundant components before ordering your system.

4.2 Availability

IBM hardware and microcode capability to continuously monitor execution of hardware functions is generally described as the process of First Failure Data Capture (FFDC). This process includes the strategy of predictive failure analysis, which refers to the ability to track intermittent correctable errors and to vary components off-line before they reach the point of hard failure, causing a system outage, and without the need to recreate the problem.

The POWER7 family of systems continues to offer and introduce significant enhancements designed to increase system availability to drive towards a high-availability objective with hardware components that can perform the following automatic functions:

- Self-diagnose and self-correct during run time.
- Automatically reconfigure to mitigate potential problems from suspect hardware.
- Self-heal or automatically substitute good components for failing components.

Note: POWER7 processor-based servers are independent of the operating system for error detection and fault isolation within the central electronics complex.

Throughout this chapter we describe IBM POWER7 processor-based systems technologies focused on keeping a system up and running. For a specific set of functions focused on detecting errors before they become serious enough to stop computing work, see 4.3.1, “Detecting errors” on page 183.

4.2.1 Partition availability priority

Also available is the ability to assign availability priorities to partitions. If an alternate processor recovery event requires spare processor resources and there are no other means of obtaining the spare resources, the system determines which partition has the lowest priority and attempts to claim the needed resource. On a properly configured POWER processor-based server, this approach allows that capacity to first be obtained from a low-priority partition instead of a high-priority partition.

This capability is relevant to the total system availability because it gives the system an additional stage before an unplanned outage. In the event that insufficient resources exist to maintain full system availability, these servers attempt to maintain partition availability by user-defined priority.

Partition availability priority is assign to partitions using a *weight value* or integer rating. The lowest priority partition is rated at 0 (zero) and the highest priority partition is valued at 255. The default value is set at 127 for standard partitions and 192 for Virtual I/O Server (VIOS) partitions. You can vary the priority of individual partitions.

Partition availability priorities can be set for both dedicated and shared processor partitions. The POWER Hypervisor uses the relative partition weight value among active partitions to favor higher priority partitions for processor sharing, adding and removing processor capacity, and favoring higher priority partitions for normal operation.

Note that the partition specifications for *minimum*, *desired*, and *maximum* capacity are also taken into account for capacity-on-demand options, and if total system-wide processor capacity becomes disabled because of deconfigured failed processor cores. For example, if total system-wide processor capacity is sufficient to run all partitions, at least with the minimum capacity, the partitions are allowed to start or continue running. If processor capacity is insufficient to run a partition at its minimum value, then starting that partition results in an error condition that must be resolved.

4.2.2 General detection and deallocation of failing components

Runtime correctable or recoverable errors are monitored to determine whether there is a pattern of errors. If these components reach a predefined error limit, the service processor initiates an action to deconfigure the faulty hardware, helping to avoid a potential system outage and to enhance system availability.

Persistent deallocation

To enhance system availability, a component that is identified for deallocation or deconfiguration on a POWER processor-based system is flagged for persistent deallocation. Component removal can occur either dynamically (while the system is running) or at boot time (IPL), depending both on the type of fault and when the fault is detected.

In addition, runtime unrecoverable hardware faults can be deconfigured from the system after the first occurrence. The system can be rebooted immediately after failure and resume operation on the remaining stable hardware. This prevents the same faulty hardware from

affecting system operation again. The repair action is deferred to a more convenient, less critical time.

The following persistent deallocation functions are included:

- Processor.
- L2/L3 cache lines. (Cache lines are dynamically deleted.)
- Memory.
- Deconfigure or bypass failing I/O adapters.

Processor instruction retry

As in POWER6, the POWER7 processor has the ability to do processor instruction retry and alternate processor recovery for a number of core related faults. Doing this significantly reduces exposure to both permanent and intermittent errors in the processor core.

Intermittent errors, often due to cosmic rays or other sources of radiation, are generally not repeatable.

With the instruction retry function, when an error is encountered in the core, in caches and certain logic functions, the POWER7 processor first automatically retries the instruction. If the source of the error was truly transient, the instruction succeeds and the system can continue as before.

Note that on IBM systems prior to POWER6, such an error typically caused a checkstop.

Alternate processor retry

Hard failures are more difficult, being permanent errors that will be replicated each time that the instruction is repeated. Retrying the instruction does not help in this situation because the instruction will continue to fail.

As in POWER6, POWER7 processors have the ability to extract the failing instruction from the faulty core and retry it elsewhere in the system for a number of faults, after which the failing core is dynamically deconfigured and scheduled for replacement.

Dynamic processor deallocation

Dynamic processor deallocation enables automatic deconfiguration of processor cores when patterns of recoverable core-related faults are detected. Dynamic processor deallocation prevents a recoverable error from escalating to an unrecoverable system error, which might otherwise result in an unscheduled server outage. Dynamic processor deallocation relies on the service processor's ability to use FFDC-generated recoverable error information to notify the POWER Hypervisor when a processor core reaches its predefined error limit. Then the POWER Hypervisor dynamically deconfigures the failing core and is called out for replacement. The entire process is transparent to the partition owning the failing instruction.

Single processor checkstop

As in the POWER6 processor, the POWER7 processor provides single core check stopping for certain processor logic, command, or control errors that cannot be handled by the availability enhancements in the preceding section.

This significantly reduces the probability of any one processor affecting total system availability by containing most processor checkstops to the partition that was using the processor at the time that full checkstop went into effect.

Even with all these availability enhancements to prevent processor errors from affecting system-wide availability into play, there will be errors that can result in a system-wide outage.

4.2.3 Memory protection

A memory protection architecture that provides good error resilience for a relatively small L1 cache might be inadequate for protecting the much larger system main store. Therefore, a variety of protection methods are used in POWER processor-based systems to avoid uncorrectable errors in memory.

Memory protection plans must take into account many factors, including these:

- Size
- Desired performance
- Memory array manufacturing characteristics

POWER7 processor-based systems have a number of protection schemes designed to prevent, protect, or limit the effect of errors in main memory:

- Chipkill

  Chipkill is an enhancement that enables a system to sustain the failure of an entire DRAM chip. An ECC word uses 18 DRAM chips from two DIMM pairs, and a failure on any of the DRAM chips can be fully recovered by the ECC algorithm. The system can continue indefinitely in this state with no performance degradation until the failed DIMM can be replaced.

- 72-byte ECC

  In POWER7, an ECC word consists of 72 bytes of data. Of these, 64 are used to hold application data. The remaining eight bytes are used to hold check bits and additional information about the ECC word.

  This innovative ECC algorithm from IBM research works on DIMM pairs on a rank basis (a rank is a group of nine DRAM chips on the Power 710 and 730). With this ECC code, the system can dynamically recover from an entire DRAM failure (Chipkill), but it can also correct an error even if another *symbol* (a byte, accessed by a 2-bit line pair) experiences a fault (an improvement from the Double Error Detection/Single Error Correction ECC implementation found on the POWER6 processor-based systems).

- Hardware scrubbing

  Hardware scrubbing is a method used to deal with intermittent errors. IBM POWER processor-based systems periodically address all memory locations. Any memory locations with a correctable error are rewritten with the correct data.

- CRC

  The bus that is transferring data between the processor and the memory uses CRC error detection with a failed operation-retry mechanism and the ability to dynamically retune bus parameters when a fault occurs. In addition, the memory bus has spare capacity to substitute a data bit-line whenever it is determined to be faulty.

POWER7 memory subsystem

The POWER7 processor chip contains two memory controllers with four channels per Memory Controller. Each channel connects to a single DIMM, but as the channels work in pairs, a processor chip can address four DIMM pairs, two pairs per memory controller.

The bus transferring data between the processor and the memory uses CRC error detection with a failed operation retry mechanism and the ability to dynamically retune bus parameters when a fault occurs. In addition, the memory bus has spare capacity to substitute a spare data bit-line for one that is determined to be faulty.

Figure 4-2 shows a POWER7 processor chip, with its memory interface, consisting of two controllers and four DIMMs per controller. Advanced memory buffer chips are exclusive to IBM and help to increase performance, acting as read/write buffers. Power 720 and Power 740 uses one memory controller. Advanced memory buffer chips are on the system planar and support two DIMMs each.

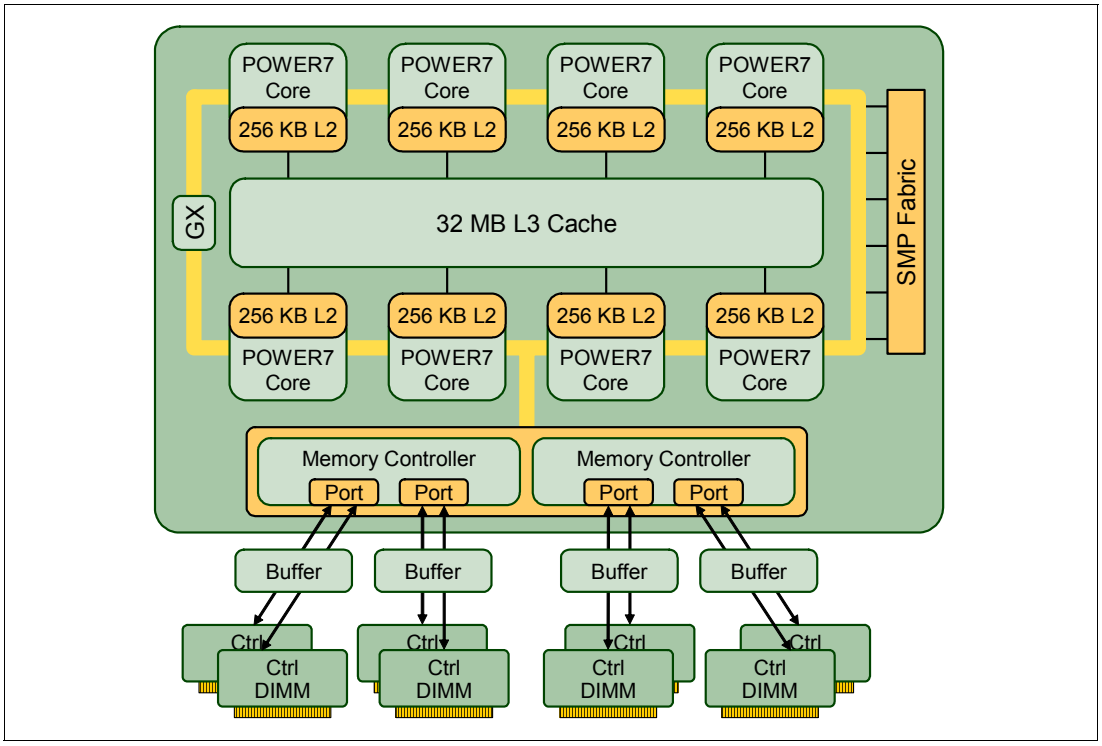


Figure 4-2 POWER7 memory subsystem

Memory page deallocation

While coincident cell errors in separate memory chips are statistically rare, IBM POWER processor-based systems can contain these errors using a memory page deallocation scheme for partitions running IBM AIX and IBM i operating systems as well as for memory pages owned by the POWER Hypervisor. If a memory address experiences an uncorrectable or repeated correctable single cell error, the Service Processor sends the memory page address to the POWER Hypervisor to be marked for deallocation.

Pages used by the POWER Hypervisor are deallocated as soon as the page is released.

In other cases, the POWER Hypervisor notifies the owning partition that the page must be deallocated. Where possible, the operating system moves any data currently contained in that memory area to another memory area and removes the pages associated with this error from its memory map, no longer addressing these pages. The operating system performs memory page deallocation without any user intervention and is transparent to users and applications.

The POWER Hypervisor maintains a list of pages marked for deallocation during the current platform Initial Program Load (IPL). During a partition IPL, the partition receives a list of all the bad pages in its address space. In addition, if memory is dynamically added to a partition (through a dynamic LPAR operation), the POWER Hypervisor warns the operating system when memory pages are included that need to be deallocated.

Finally, if an uncorrectable error in memory is discovered, the logical memory block associated with the address with the uncorrectable error is marked for deallocation by the POWER Hypervisor. This deallocation will take effect on a partition reboot if the logical memory block is assigned to an active partition at the time of the fault.

In addition, the system will deallocate the entire memory group associated with the error on all subsequent system reboots until the memory is repaired. This precaution is intended to guard against future uncorrectable errors while waiting for parts replacement.

Memory persistent deallocation

Defective memory discovered at boot time is automatically switched off. If the Service Processor detects a memory fault at boot time, it marks the affected memory as bad so it is not to be used on subsequent reboots.

If the Service Processor identifies faulty memory in a server that includes CoD memory, the POWER Hypervisor attempts to replace the faulty memory with available CoD memory. Faulty resources are marked as deallocated and working resources are included in the active memory space. Because these activities reduce the amount of CoD memory available for future use, repair of the faulty memory must be scheduled as soon as is convenient.

Upon reboot, if not enough memory is available to meet minimum partition requirements, the POWER Hypervisor will reduce the capacity of one or more partitions.

Depending on the configuration of the system, the HMC Service Focal Point™, OS Service Focal Point, or Service Processor will receive a notification of the failed component, and will trigger a service call.

4.2.4 Cache protection

POWER7 processor-based systems are designed with cache protection mechanisms, including cache line delete in both L2 and L3 arrays, Processor Instruction Retry and Alternate Processor Recovery protection on L1-I and L1-D, and redundant “Repair” bits in L1-I, L1-D, and L2 caches, as well as L2 and L3 directories.

L1 instruction and data array protection

The POWER7 processor instruction and data caches are protected against intermittent errors using Processor Instruction Retry and against permanent errors by Alternate Processor Recovery, both mentioned earlier. L1 cache is divided into sets. POWER7 processor can deallocate all but one before doing a Processor Instruction Retry.

In addition, faults in the Segment Lookaside Buffer (SLB) array are recoverable by the POWER Hypervisor. The SLB is used in the core to perform address translation calculations.

L2 and L3 array protection

The L2 and L3 caches in the POWER7 processor are protected with double-bit detect single-bit correct error detection code (ECC). Single-bit errors are corrected before forwarding to the processor, and subsequently written back to L2/L3.

In addition, the caches maintain a cache line delete capability. A threshold of correctable errors detected on a cache line can result in the data in the cache line being purged and the cache line removed from further operation without requiring a reboot. An ECC uncorrectable error detected in the cache can also trigger a purge and delete of the cache line. This results in no loss of operation because an unmodified copy of the data can be held on system

memory to reload the cache line from main memory. Modified data is handled through Special Uncorrectable Error handling.

L2 and L3 deleted cache lines are marked for persistent deconfiguration on subsequent system reboots until they can be replaced.

4.2.5 Special Uncorrectable Error handling

While it is rare, an uncorrectable data error can occur in memory or a cache. IBM POWER processor-based systems attempt to limit the impact of an uncorrectable error to the least possible disruption, using a well-defined strategy that first considers the data source. Sometimes, an uncorrectable error is temporary in nature and occurs in data that can be recovered from another repository, for example:

- Data in the instruction L1 cache is never modified within the cache itself. Therefore, an uncorrectable error discovered in the cache is treated like an ordinary cache miss, and correct data is loaded from the L2 cache.
- The L2 and L3 cache of the POWER7 processor-based systems can hold an unmodified copy of data in a portion of main memory. In this case, an uncorrectable error simply triggers a reload of a cache line from main memory.

In cases where the data cannot be recovered from another source, a technique called Special Uncorrectable Error (SUE) handling is used to prevent an uncorrectable error in memory or cache from immediately causing the system to terminate. Rather, the system tags the data and determines whether it will ever be used again:

- If the error is irrelevant, it will not force a checkstop.
- If the data is used, termination can be limited to the program/kernel or hypervisor owning the data, or a freeze of the I/O adapters controlled by an I/O hub controller if data is going to be transferred to an I/O device.

When an uncorrectable error is detected, the system modifies the associated ECC word, thereby signaling to the rest of the system that the “standard” ECC is no longer valid. The Service Processor is then notified and takes appropriate actions. When running AIX V5.2 or later, or Linux, and a process attempts to use the data, the operating system is informed of the error and might terminate, or only terminate a specific process associated with the corrupt data, depending on the operating system and firmware level and whether the data was associated with a kernel or non-kernel process.

It is only in the case where the corrupt data is used by the POWER Hypervisor that the entire system must be rebooted, thereby preserving overall system integrity.

Depending on system configuration and source of the data, errors encountered during I/O operations might not result in a machine check. Instead, the incorrect data is handled by the processor host bridge (PHB) chip. When the PHB chip detects a problem, it rejects the data, preventing data from being written to the I/O device.

The PHB then enters a freeze mode, halting normal operations. Depending on the model and type of I/O being used, the freeze might include the entire PHB chip or simply a single bridge, resulting in the loss of all I/O operations that use the frozen hardware until a power-on reset of the PHB is done. The impact to partitions depends on how the I/O is configured for redundancy. In a server configured for failover availability, redundant adapters spanning multiple PHB chips can enable the system to recover transparently, without partition loss.

4.2.6 PCI extended error handling

IBM estimates that PCI adapters can account for a significant portion of the hardware-based errors on a large server. Whereas servers that rely on boot-time diagnostics can identify failing components to be replaced by hot-swap and reconfiguration, runtime errors pose a more significant problem.

PCI adapters are generally complex designs involving extensive on-board instruction processing, often on embedded micro controllers. They tend to use industry-standard-grade components with an emphasis on product cost relative to high reliability. In certain cases, they might be more likely to encounter internal microcode errors, or many of the hardware errors described for the rest of the server.

The traditional means of handling these problems is through adapter internal error reporting and recovery techniques in combination with operating system device driver management and diagnostics. In certain cases, an error in the adapter might cause transmission of bad data on the PCI bus itself, resulting in a hardware-detected parity error and causing a global machine check interrupt, eventually requiring a system reboot to continue.

PCI extended error handling (EEH) enabled adapters respond to a special data packet generated from the affected PCI slot hardware by calling system firmware, which examines the affected bus, allows the device driver to reset it, and continues without a system reboot. For Linux, EEH support extends to the majority of frequently used devices, although various third-party PCI devices might not provide native EEH support.

To detect and correct PCIe bus errors, POWER7 processor-based systems use CRC detection and instruction retry correction, while for PCI-X they use ECC.

Figure 4-3 shows the location and various mechanisms used throughout the I/O subsystem for PCI extended error handling.

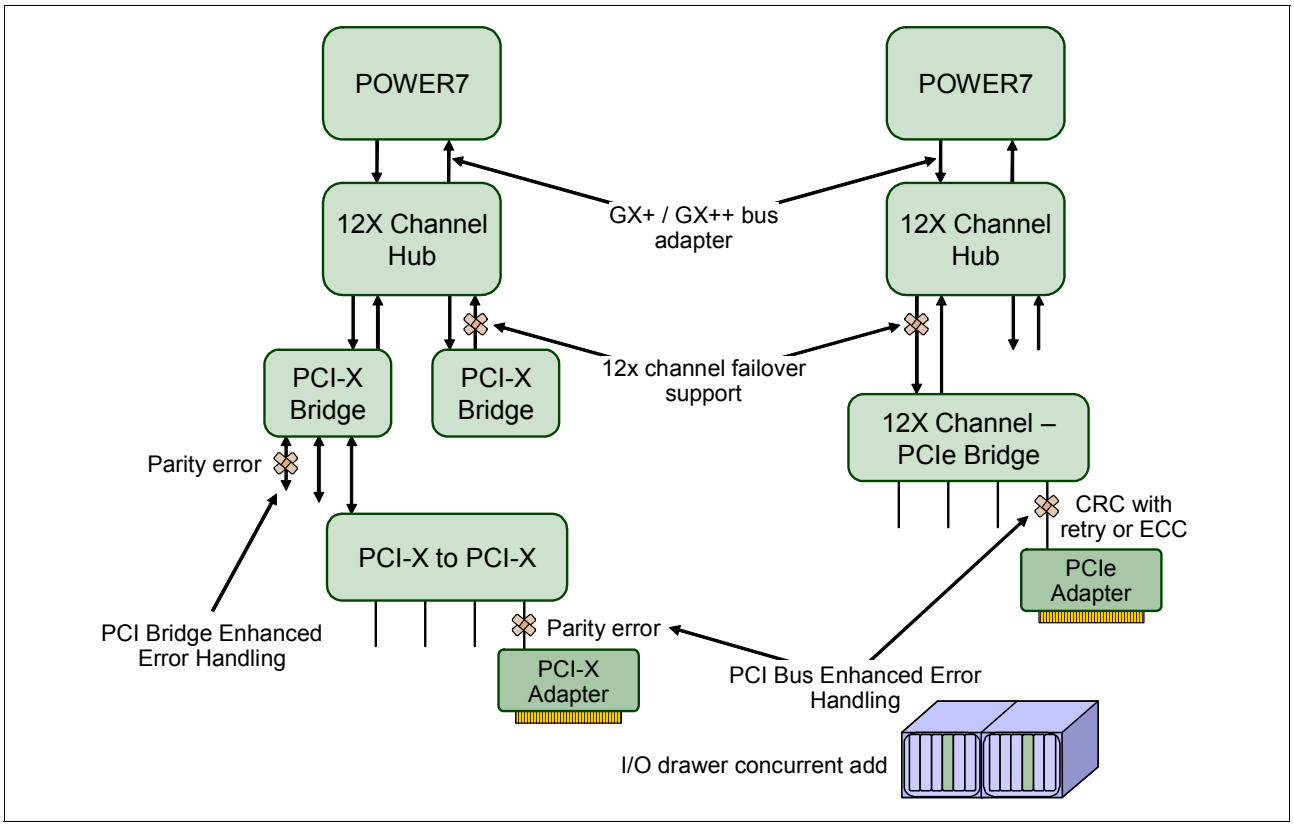


Figure 4-3 PCI extended error handling

4.3 Serviceability

IBM Power Systems design considers both IBM and client needs. The IBM Serviceability Team has enhanced the base service capabilities and continues to implement a strategy that incorporates best-of-breed service characteristics from diverse IBM systems offerings.

Serviceability includes system installation, system upgrades and downgrades (MES), and system maintenance and repair.

The goal of the IBM Serviceability Team is to design and provide the most efficient system service environment that includes:

- Easy access to service components, design for Customer Set Up (CSU), Customer Installed Features (CIF), and Customer Replaceable Units (CRU)
- On demand service education
- Error detection and fault isolation (ED/FI)
- First-failure data capture (FFDC)
- An automated guided repair strategy that uses common service interfaces for a converged service approach across multiple IBM server platforms

By delivering on these goals, IBM Power Systems servers enable faster and more accurate repair, and reduce the possibility of human error.

Client control of the service environment extends to firmware maintenance on all of the POWER processor-based systems. This strategy contributes to higher systems availability with reduced maintenance costs.

This section provides an overview of the progressive steps of error detection, analysis, reporting, notifying, and repairing that are found in all POWER processor-based systems.

4.3.1 Detecting

The first and most crucial component of a solid serviceability strategy is the ability to accurately and effectively detect errors when they occur. Although not all errors are a guaranteed threat to system availability, those that go undetected can cause problems because the system does not have the opportunity to evaluate and act if necessary. POWER processor-based systems employ System z® server-inspired error detection mechanisms that extend from processor cores and memory to power supplies and hard drives.

Service processor

The service processor is a microprocessor that is powered separately from the main instruction processing complex. The service processor provides the capabilities for:

- POWER Hypervisor (system firmware) and Hardware Management Console connection surveillance
- Several remote power control options
- Reset and boot features
- Environmental monitoring

  The service processor monitors the server's built-in temperature sensors, sending instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range. Using an architected operating system interface, the service processor notifies the operating system of potential environmentally related problems so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached.

  The service processor can also post a warning and initiate an orderly system shutdown in these cases:

  – The operating temperature exceeds the critical level (for example, failure of air conditioning or air circulation around the system).
  – The system fan speed is out of operational specification (for example, because of multiple fan failures).
  – The server input voltages are out of operational specification.

  The service processor can immediately shut down a system when the following cases occur:

  – Temperature exceeds the critical level or remains above the warning level for too long.
  – Internal component temperatures reach critical levels.
  – Non-redundant fan failures occur.
- Placing calls

  On systems without a Hardware Management Console, the service processor can place calls to report surveillance failures with the POWER Hypervisor, critical environmental faults, and critical processing faults even when the main processing unit is inoperable.

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

- Availability

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

  Note: The auto-restart (reboot) option has to be enabled from the Advanced System Manager Interface or from the Control (Operator) Panel. Figure 4-4 shows this option using the ASMI.

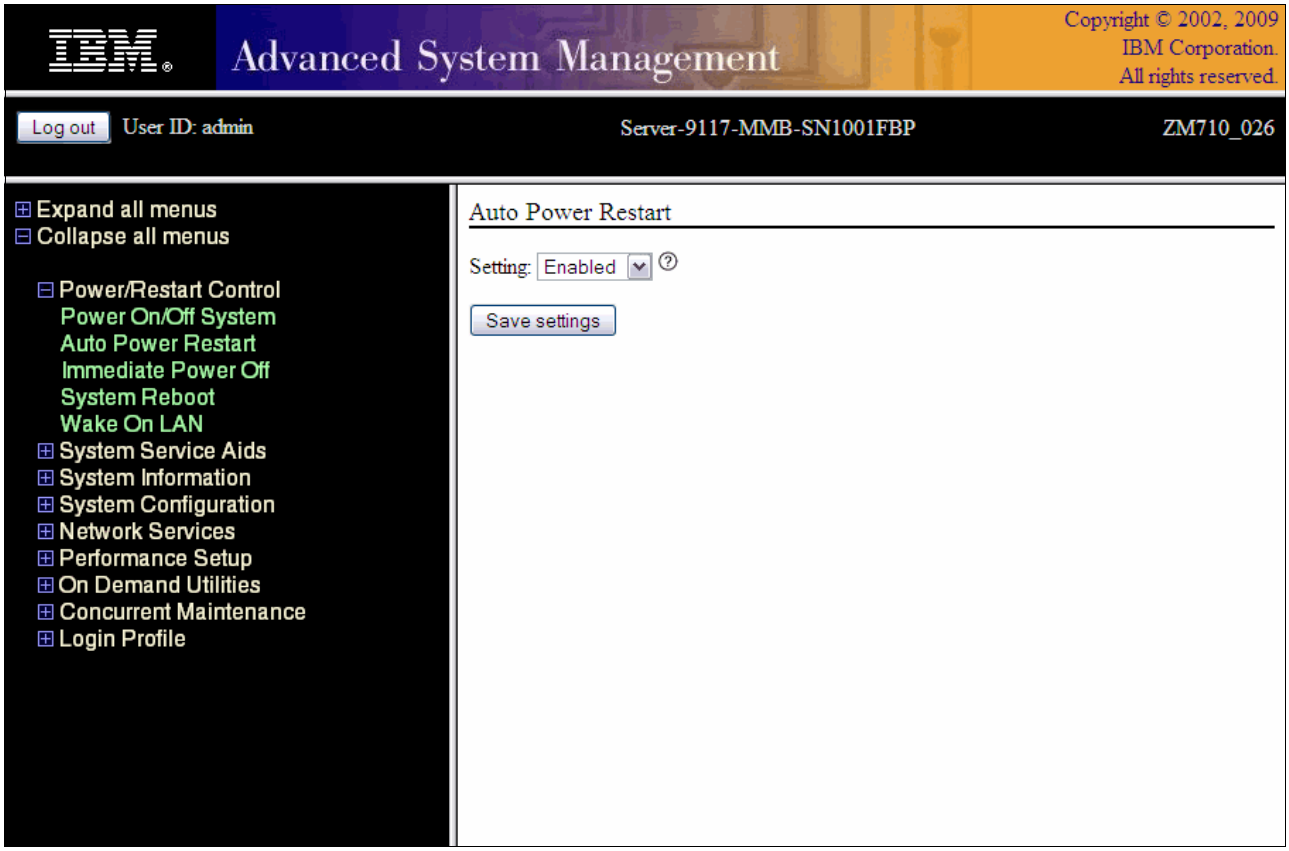


Figure 4-4 ASMI Auto Power Restart setting window interface

- Fault monitoring

  Built-in self-test (BIST) checks processor, cache, memory, and associated hardware that is required for proper booting of the operating system when the system is powered on at the initial installation or after a hardware configuration change (for example, an upgrade). If a non-critical error is detected or if the error occurs in a resource 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 random access memory (NVRAM). When

the operating system completes booting, the information is passed from the NVRAM to the system error log where it is analyzed by error log analysis (ELA) routines. Appropriate actions are taken to report the boot-time error for subsequent service, if required.

- Concurrent access to the service processors menus of the Advanced System Management Interface (ASMI)

  This access allows nondisruptive abilities to change system default parameters, interrogate service processor progress and error logs, set and reset server indicators (Guiding Light for midrange and high-end servers, Light Path for low-end servers), accessing all service processor functions without having to power down the system to the standby state. This way allows the administrator or service representative to dynamically access the menus from any web-browser-enabled console that is attached to the Ethernet service network, concurrently with normal system operation.

- Managing the interfaces for connecting uninterruptible power source systems to the POWER processor-based systems, performing Timed Power-On (TPO) sequences, and interfacing with the power and cooling subsystem

Error checkers

IBM POWER processor-based systems contain specialized hardware detection circuitry that is used to detect erroneous hardware operations. Error checking hardware ranges from parity error detection coupled with processor instruction retry and bus retry, to ECC correction on caches and system buses. All IBM hardware error checkers have distinct attributes:

- Continuous monitoring of system operations to detect potential calculation errors.
- Attempts to isolate physical faults based on runtime detection of each unique failure.
- Ability to initiate a wide variety of recovery mechanisms designed to correct the problem. The POWER processor-based systems include extensive hardware and firmware recovery logic.

Fault isolation registers

Error checker signals are captured and stored in hardware fault isolation registers (FIRs). The associated logic circuitry is used to limit the domain of an error to the first checker that encounters the error. In this way, runtime error diagnostics can be deterministic so that for every check station, the unique error domain for that checker is defined and documented. Ultimately, the error domain becomes the field-replaceable unit (FRU) call, and manual interpretation of the data is not normally required.

First-failure data capture

FFDC is an error isolation technique, which ensures that when a fault is detected in a system through error checkers or other types of detection methods, the root cause of the fault will be captured without the need to recreate the problem or run an extended tracing or diagnostics program.

For the vast majority of faults, a good FFDC design means that the root cause is detected automatically without intervention by a service representative. Pertinent error data related to the fault is captured and saved for analysis. In hardware, FFDC data is collected from the fault isolation registers and from the associated logic. In firmware, this data consists of return codes, function calls, and so forth.

FFDC *check stations* are carefully positioned within the server logic and data paths to ensure that potential errors can be quickly identified and accurately tracked to a field-replaceable unit (FRU).

This proactive diagnostic strategy is a significant improvement over the classic, less accurate *reboot and diagnose* service approaches.

Figure 4-5 shows a schematic of a fault isolation register implementation.

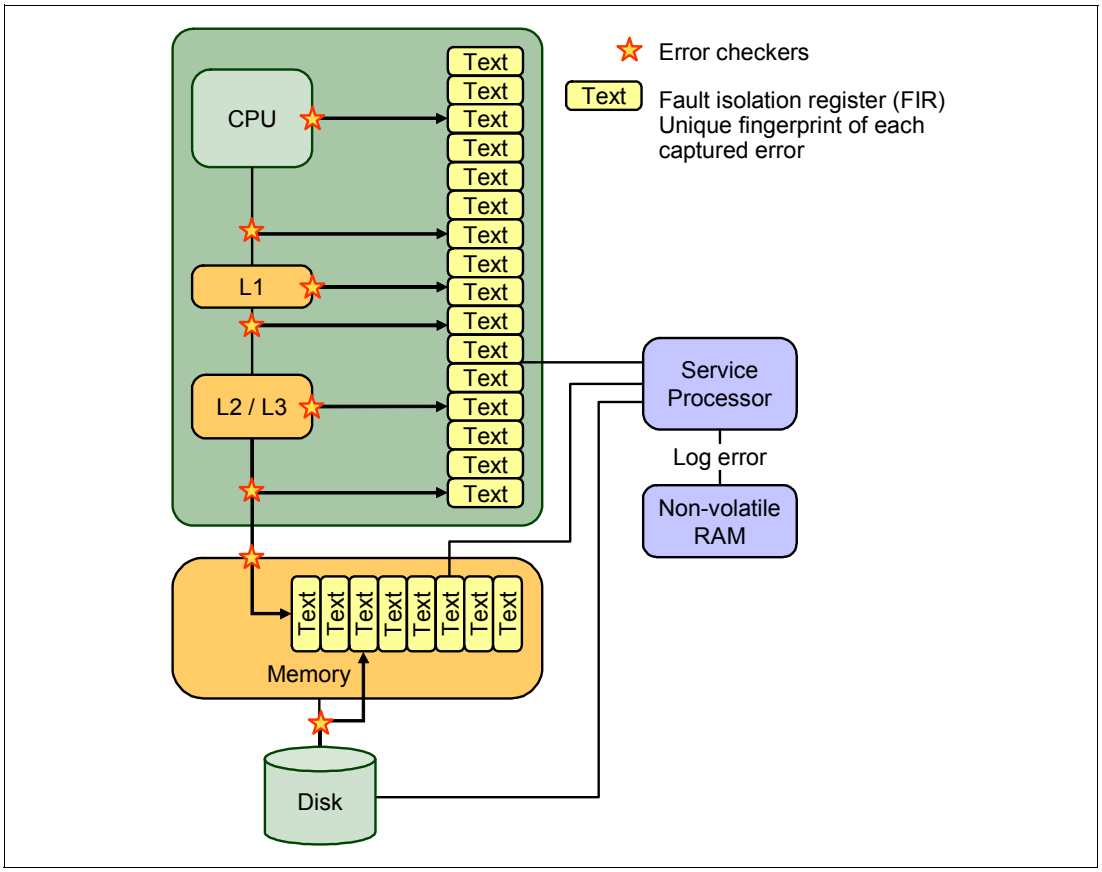


Figure 4-5 Schematic of FIR implementation

Fault isolation

The service processor interprets error data that is captured by the FFDC checkers (saved in the FIRs or other firmware-related data capture methods) to determine the root cause of the error event.

Root cause analysis might indicate that the event is recoverable, meaning that a service action point or need for repair has not been reached. Alternatively, it could indicate that a service action point has been reached, where the event exceeded a predetermined threshold or was unrecoverable. Based on the isolation analysis, recoverable error threshold counts might be incremented. No specific service action is necessary when the event is recoverable.

When the event requires a service action, additional required information is collected to service the fault. For unrecoverable errors or for recoverable events that meet or exceed their service threshold (meaning that a service action point has been reached), a request for service is initiated through an error logging component.

4.3.2 Diagnosing

Using the extensive network of advanced and complementary error detection logic that is built directly into hardware, firmware, and operating systems, the IBM Power Systems servers can perform considerable self-diagnosis.

Boot time

When an IBM Power Systems server powers up, the service processor initializes the system hardware. Boot-time diagnostic testing uses a multi-tier approach for system validation, starting with managed low-level diagnostics that are supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines. Boot-time diagnostic routines include:

- Built-in self-tests (BISTs) for both logic components and arrays ensure the internal integrity of components. Because the service processor assists in performing these tests, the system is enabled to perform fault determination and isolation, whether or not the system processors are operational. Boot-time BISTs can also find faults undetectable by processor-based power-on self-test (POST) or diagnostics.
- Wire-tests discover and precisely identify connection faults between components such as processors, memory, or I/O hub chips.
- Initialization of components such as ECC memory, typically by writing patterns of data and allowing the server to store valid ECC data for each location, can help isolate errors.

To minimize boot time, the system determines which of the diagnostics are required to be started in order to ensure correct operation, based on the way the system was powered off or on the boot-time selection menu.

Run time

All Power Systems servers can monitor critical system components during run time, and they can take corrective actions when recoverable faults occur. IBM hardware error-check architecture provides the ability to report non-critical errors in an *out-of-band* communications path to the service processor without affecting system performance.

A significant part of IBM runtime diagnostic capabilities originates with the service processor. Extensive diagnostic and fault analysis routines have been developed and improved over many generations of POWER processor-based servers, and enable quick and accurate predefined responses to both actual and potential system problems.

The service processor correlates and processes runtime error information, using logic derived from IBM engineering expertise to count recoverable errors (called thresholding) and predict when corrective actions must be automatically initiated by the system. These actions can include these:

- Requests for a part to be replaced
- Dynamic invocation of built-in redundancy for automatic replacement of a failing part
- Dynamic deallocation of failing components so that system availability is maintained

Device drivers

In certain cases diagnostics are best performed by operating-system-specific drivers, most notably I/O devices that are owned directly by a logical partition. In these cases, the operating system device driver often works in conjunction with I/O device microcode to isolate and recover from problems. Potential problems are reported to an operating system device driver, which logs the error. I/O devices can also include specific exercisers that can be invoked by the diagnostic facilities for problem recreation if required by service procedures.

4.3.3 Reporting

In the unlikely event that a system hardware or environmentally induced failure is diagnosed, IBM Power Systems servers report the error through a number of mechanisms. The analysis result is stored in system NVRAM. Error log analysis (ELA) can be used to display the failure cause and the physical location of the failing hardware.

With the integrated service processor, the system has the ability to automatically send out an alert through a phone line to a pager, or call for service in the event of a critical system failure. A hardware fault also illuminates the amber system fault LED located on the system unit to alert the user of an internal hardware problem.

On POWER7 processor-based servers, hardware and software failures are recorded in the system log. When an HMC is attached, an ELA routine analyzes the error, forwards the event to the Service Focal Point (SFP) application running on the HMC or SDMC, and has the capability to notify the system administrator that it has isolated a likely cause of the system problem. The service processor event log also records unrecoverable checkstop conditions, forwards them to the Service Focal Point (SFP) application, and notifies the system administrator. After the information is logged in the SFP application, if the system is properly configured, a call-home service request is initiated and the pertinent failure data with service parts information and part locations is sent to the IBM service organization. This information will also contain the client contact information as defined in the Electronic Service Agent (ESA) guided setup wizard.

Error logging and analysis

When the root cause of an error has been identified by a fault isolation component, an error log entry is created with basic data such as this:

- An error code uniquely describing the error event
- The location of the failing component
- The part number of the component to be replaced, including pertinent data such as engineering and manufacturing levels
- Return codes
- Resource identifiers
- FFDC data

Data containing information about the effect that the repair will have on the system is also included. Error log routines in the operating system and FSP can then use this information and decide whether the fault is a call home candidate. If the fault requires support intervention, then a call will be placed with service and support and a notification sent to the contact defined in the ESA guided set-up wizard.

Remote support

The Remote Management and Control (RMC) subsystem is delivered as part of the base operating system, including the operating system running on the Hardware Management Console. RMC provides a secure transport mechanism across the LAN interface between the operating system and the Hardware Management Console and is used by the operating system diagnostic application for transmitting error information. It performs a number of other functions also, but these are not used for the service infrastructure.

Service Focal Point

A critical requirement in a logically partitioned environment is to ensure that errors are not lost before being reported for service, and that an error should only be reported once, regardless

of how many logical partitions experience the potential effect of the error. The Manage Serviceable Events task on the HMC/SDMC is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management.

When a local or globally reported service request is made to the operating system, the operating system diagnostic subsystem uses the Remote Management and Control Subsystem (RMC) to relay error information to the Hardware Management Console. For global events (platform unrecoverable errors, for example) the service processor will also forward error notification of these events to the Hardware Management Console, providing a redundant error-reporting path in case of errors in the RMC network.

The first occurrence of each failure type is recorded in the Manage Serviceable Events task on the HMC /SDMC. This task then filters and maintains a history of duplicate reports from other logical partitions on the service processor. It then looks at all active service event requests, analyzes the failure to ascertain the root cause, and, if enabled, initiates a call home for service. This methodology ensures that all platform errors will be reported through at least one functional path, ultimately resulting in a single notification for a single problem.

Extended error data

Extended error data (EED) is additional data that is collected either automatically at the time of a failure or manually at a later time. The data collected is dependent on the invocation method but includes information like firmware levels, operating system levels, additional fault isolation register values, recoverable error threshold register values, system status, and any other pertinent data.

The data is formatted and prepared for transmission back to IBM to assist the service support organization with preparing a service action plan for the service representative or for additional analysis.

System dump handling

In certain circumstances, an error might require a dump to be automatically or manually created. In this event, it is off-loaded to the HMC. Specific HMC information is included as part of the information that can optionally be sent to IBM support for analysis. If additional information relating to the dump is required, or if it becomes necessary to view the dump remotely, the HMC dump record notifies the IBM support center regarding on which HMC the dump is located.

4.3.4 Notifying

After a Power Systems server has detected, diagnosed, and reported an error to an appropriate aggregation point, it then takes steps to notify the client, and if necessary the IBM support organization. Depending on the assessed severity of the error and support agreement, this can range from a simple notification to having field service personnel automatically dispatched to the client site with the correct replacement part.

Client Notify

When an event is important enough to report, but does not indicate the need for a repair action or the need to call home to IBM service and support, it is classified as Client Notify. Clients are notified because these events might be of interest to an administrator. The event might be a symptom of an expected systemic change, such as a network reconfiguration or failover testing of redundant power or cooling systems. These are examples of these events:

- Network events such as the loss of contact over a local area network (LAN)
- Environmental events such as ambient temperature warnings
- Events that need further examination by the client (although these events do not necessarily require a part replacement or repair action)

Client Notify events are serviceable events, by definition, because they indicate that something has happened that requires client awareness in the event the client wants to take further action. These events can always be reported back to IBM at the client's discretion.

Call home

A correctly configured POWER processor-based system can initiate an automatic or manual call from a client location to the IBM service and support organization with error data, server status, or other service-related information. The call-home feature invokes the service organization in order for the appropriate service action to begin, automatically opening a problem report, and in certain cases, also dispatching field support. This automated reporting provides faster and potentially more accurate transmittal of error information. Although configuring call-home is optional, clients are strongly encouraged to configure this feature to obtain the full value of IBM service enhancements.

Vital product data and inventory management

Power Systems store vital product data (VPD) internally, which keeps a record of how much memory is installed, how many processors are installed, manufacturing level of the parts, and so on. These records provide valuable information that can be used by remote support and service representatives, enabling them to provide assistance in keeping the firmware and software on the server up-to-date.

IBM problem management database

At the IBM support center, historical problem data is entered into the IBM Service and Support Problem Management database. All of the information that is related to the error, along with any service actions taken by the service representative, is recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

4.3.5 Locating and servicing

The final component of a comprehensive design for serviceability is the ability to effectively locate and replace parts requiring service. POWER processor-based systems use a combination of visual cues and guided maintenance procedures to ensure that the identified part is replaced correctly, every time.

Packaging for service

These service enhancements are included in the physical packaging of the systems to facilitate service:

- Color coding (touch points):
  - Terra-cotta-colored touch points indicate that a component (FRU or CRU) can be concurrently maintained.
  - Blue-colored touch points delineate components that are not concurrently maintained (those that require the system to be turned off for removal or repair).
- Tool-less design: Selected IBM systems support tool-less or simple tool designs. These designs require no tools or simple tools, such as flathead screw drivers to service the hardware components.
- Positive retention: Positive retention mechanisms help to ensure proper connections between hardware components, such as from cables to connectors, and between two cards that attach to each other. Without positive retention, hardware components run the risk of becoming loose during shipping or installation, preventing a good electrical connection. Positive retention mechanisms such as latches, levers, thumb-screws, pop Nylatches (U-clips), and cables are included to help prevent loose connections and aid in installing (seating) parts correctly. These positive retention items do not require tools.

Light Path

The Light Path LED feature is for low-end systems, including Power Systems up to models 750 and 755, that might be repaired by clients. In the Light Path LED implementation, when a fault condition is detected on the POWER7 processor-based system, an amber FRU fault LED is illuminated, which is then rolled up to the system fault LED. The Light Path system pinpoints the exact part by turning on the amber FRU fault LED that is associated with the part to be replaced.

The system can clearly identify components for replacement by using specific component-level LEDs, and can also guide the servicer directly to the component by signaling (staying on solid) the system fault LED, the enclosure fault LED, and the component FRU fault LED.

After the repair, the LEDs shut off automatically if the problem is fixed.

Guiding Light

Midrange and high-end systems, including models 770 and 780 and later, are usually repaired by IBM Support personnel.

The enclosure and system identify LEDs that turn on solid and that can be used to follow the path from the system to the enclosure and down to the specific FRU.

Guiding Light uses a series of flashing LEDs, allowing a service provider to quickly and easily identify the location of system components. Guiding Light can also handle multiple error conditions simultaneously, which might be necessary in certain complex high-end configurations. In these situations, Guiding Light awaits for the servicer's indication of what failure to attend first and then illuminates the LEDs to the failing component.

Data centers can be complex places, and Guiding Light is designed to do more than identify visible components. When a component might be hidden from view, Guiding Light can flash a sequence of LEDs that extend to the frame exterior, clearly *guiding* the service representative to the correct rack, system, enclosure, drawer, and component.

Service labels

Service providers use these labels to assist them in performing maintenance actions. Service labels are found in various formats and positions and are intended to transmit readily available information to the servicer during the repair process.

Several of these service labels and the purpose of each are described in the following list:

- Location diagrams are strategically located on the system hardware, relating information regarding the placement of hardware components. Location diagrams can include location codes, drawings of physical locations, concurrent maintenance status, or other data that is pertinent to a repair. Location diagrams are especially useful when multiple components are installed, such as DIMMs, CPUs, processor books, fans, adapter cards, LEDs, and power supplies.
- Remove or replace procedure labels contain procedures often found on a cover of the system or in other spots that are accessible to the servicer. These labels provide systematic procedures, including diagrams, detailing how to remove and replace certain serviceable hardware components.
- Numbered arrows are used to indicate the order of operation and serviceability direction of components. Various serviceable parts such as latches, levers, and touch points must be pulled or pushed in a certain direction and certain order so that the mechanical mechanisms can engage or disengage. Arrows generally improve the ease of serviceability.

The operator panel

The operator panel on a POWER processor-based system is a four-row by 16-element LCD display that is used to present boot progress codes, indicating advancement through the system power-on and initialization processes. The operator panel is also used to display error and location codes when an error occurs that prevents the system from booting. It includes several buttons, enabling a service support representative (SSR) or client to change various boot-time options and for other limited service functions.

Concurrent maintenance

The IBM POWER7 processor-based systems are designed with the understanding that certain components have higher intrinsic failure rates than others. The movement of fans, power supplies, and physical storage devices naturally make them more susceptible to wearing down or burning out. Other devices such as I/O adapters can begin to wear from repeated plugging and unplugging. For these reasons, these devices have been specifically designed to be concurrently maintainable when properly configured.

In other cases, a client might be in the process of moving or redesigning a data center, or planning a major upgrade. At times like these, flexibility is crucial. The IBM POWER7 processor-based systems are designed for redundant or concurrently maintainable power, fans, physical storage, and I/O towers.

The most recent members of the IBM Power Systems family, based on the POWER7 processor, continue to support concurrent maintenance of power, cooling, PCI adapters, media devices, I/O drawers, GX adapter, and the operator panel. In addition, they support concurrent firmware fix pack updates when possible. The determination of whether a firmware fix pack release can be updated concurrently is identified in the readme file that is released with the firmware.

Firmware updates

System Firmware is delivered as a Release Level or a Service Pack. Release Levels support the general availability (GA) of new function or features, and new machine types or models.

Upgrading to a higher Release Level is disruptive to customer operations. IBM intends to introduce no more than two new Release Levels per year. These Release Levels will be supported by Service Packs. Service Packs are intended to contain only firmware fixes and not to introduce new function. A Service Pack is an update to an existing Release Level.

If a system is HMC managed you will use the HMC for firmware updates. Using the HMC allows you to take advantage of the Concurrent Firmware Maintenance (CFM) option when concurrent service packs are available. CFM is the IBM term used to describe the IBM Power Systems firmware updates that can be partially or wholly concurrent or non-disruptive. With the introduction of CFM, IBM is significantly increasing the client's opportunity to stay on a given release level for longer periods of time. Clients wanting maximum stability can defer until there is a compelling reason to upgrade, such as:

- A release level is approaching its end-of-service date (that is, it has been available for about a year and hence will go out of service support soon).
- The client is moving a system to a more standardized Release Level when there are multiple systems in an environment with similar hardware.
- A new release has new functionality that is needed in the environment.
- A scheduled maintenance action will cause a platform reboot. This also provides an opportunity to upgrade to a new firmware release.

The update and upgrade of system firmware is dependant on several factors, such as whether the system is standalone or managed by a HMC, the current firmware installed, and what operating systems are running on the system. These scenarios and the associated installation instructions are comprehensively outlined in the firmware section of Fix Central, which can be found here:

http://www.ibm.com/support/fixcentral/

You might also want to review the best practice white papers, which can be found here:

http://www14.software.ibm.com/webapp/set2/sas/f/best/home.html

Repair and verify system

Repair and verify (R&V) is a system used to guide a service provider step-by-step through the process of repairing a system and verifying that the problem has been repaired. The steps are customized in the appropriate sequence for the particular repair for the specific system being repaired. Repair scenarios covered by repair and verify include:

- Replacing a defective field-replaceable unit (FRU)
- Reattaching a loose or disconnected component
- Correcting a configuration error
- Removing or replacing an incompatible FRU
- Updating firmware, device drivers, operating systems, middleware components, and IBM applications after replacing a part

Repair and verify procedures can be used by both service representative providers who are familiar with the task and those who are not. Education On Demand content is placed in the procedure at the appropriate locations. Throughout the repair and verify procedure, repair history is collected and provided to the Service and Support Problem Management Database for storage with the serviceable event to ensure that the guided maintenance procedures are operating correctly.

If a server is managed by an HMC, then many of the R&V procedures are performed from the HMC. If the FRU to be replaced is a PCI adaptor or an internal storage device, then

the service action is always performed from the operating system of the partition owning that resource.

Clients can subscribe through the subscription services to obtain notifications about the latest updates available for service-related documentation. The latest version of the documentation is accessible through the internet.

4.4 Manageability

Several functions and tools help manageability and enable you to efficiently and effectively manage your system.

4.4.1 Service user interfaces

The Service Interface allows support personnel or the client to communicate with the service support applications in a server using a console, interface, or terminal. Delivering a clear, concise view of available service applications, the Service Interface allows the support team to manage system resources and service information in an efficient and effective way.

Applications available through the Service Interface are carefully configured and placed to give service providers access to important service functions.

Various service interfaces are used, depending on the state of the system and its operating environment. The primary service interfaces are:

- Light Path and Guiding Light

  For more information, see “Light Path” on page 152 and “Guiding Light” on page 152.
- Service processor, Advanced System Management Interface (ASMI)
- Operator panel
- Operating system service menu
- Service Focal Point on the Hardware Management Console
- Service Focal Point Lite on Integrated Virtualization Manager

Service processor

The service processor is a controller that is running its own operating system. It is a component of the service interface card.

The service processor operating system has specific programs and device drivers for the service processor hardware. The host interface is a processor support interface that is connected to the POWER processor. The service processor is always working, regardless of the main system unit’s state. The system unit can be in the following states:

- Standby (power off)
- Operating, ready-to-start partitions
- Operating with running logical partitions

Functions

The service processor is used to monitor and manage the system hardware resources and devices. The service processor checks the system for errors, ensuring the connection to the HMC for manageability purposes and accepting ASMI Secure Sockets Layer (SSL) network connections. The service processor provides the ability to view and manage the

machine-wide settings by using the ASMI, and enables complete system and partition management from the HMC.

Note: The service processor enables a system that does not boot to be analyzed. The error log analysis can be performed from either the ASMI or the HMC.

The service processor uses two Ethernet 10/100/1000 Mbps ports. Consider this information:

- Both Ethernet ports are visible only to the service processor and can be used to attach the server to an HMC or to access the ASMI. The ASMI options can be accessed through an HTTP server that is integrated into the service processor operating environment.
- Because of firmware-heavy workload, firmware can support only these ports at 10/100 Mbps rate although the Ethernet adapter is capable of 1 Gbps.
- Both Ethernet ports have a default IP address, as follows:
  - Service processor Eth0 or HMC1 port is configured as 169.254.2.147.
  - Service processor Eth1 or HMC2 port is configured as 169.254.3.147.
- When a redundant service processor is present, these default IP addresses are used:
  - Service processor Eth0 or HMC1 port is configured as 169.254.2.146.
  - Service processor Eth1 or HMC2 port is configured as 169.254.3.146.

The functions available through service processor include:

- Call Home
- Advanced System Management Interface (ASMI)
- Error Information (error code, PN, Location Codes) menu
- View of guarded components
- Limited repair procedures
- Generate dump
- LED Management menu
- Remote view of ASMI menus
- Firmware update through USB key

Advanced System Management Interface

ASMI is the interface to the service processor that enables you to manage 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. Various repair procedures require connection to the ASMI.

The ASMI is accessible through the HMC. It is also accessible by using a web browser on a system that is connected directly to the service processor (in this case, either a standard Ethernet cable or a crossed cable) or through an Ethernet network. ASMI can also be accessed from an ASCII terminal. Use the ASMI to change the service processor IP addresses or to apply certain security policies and prevent access from undesired IP addresses or ranges.

You might be able to use the service processor's default settings. In that case, accessing the ASMI is not necessary.

To access ASMI, use one of these steps:

- Access the ASMI by using an HMC.

  If configured to do so, the HMC connects directly to the ASMI for a selected system from this task.

  To connect to the Advanced System Management interface from an HMC, follow these steps:

  a. Open Systems Management from the navigation pane.

  b. From the work pane, select one or more managed systems to work with.

  c. From the System Management tasks list, select **Operations Advanced System Management** (ASM).

- Access the ASMI using a web browser.

  The web interface to the ASMI is accessible by running Microsoft Internet Explorer 7.0, Opera 9.24, or Mozilla Firefox 2.0.0.11 running on a PC or mobile computer that is connected to the service processor. The web interface is available during all phases of system operation, including the initial program load (IPL) and run time. However, a few of the menu options in the web interface are unavailable during IPL or run time to prevent usage or ownership conflicts if the system resources are in use during that phase. The ASMI provides a SSL web connection to the service processor. To establish an SSL connection, open your browser using this address:

  `https://<ip_address_of_service_processor>`

  Where <*ip_address_of_service_processor*> is the address of the service processor of your Power Systems server, such as 9.166.196.7.

  Tip: To make the connection through Internet Explorer, click **Tools Internet Options**. Clear the **Use TLS 1.0** check box, and click **OK**.

- Access the ASMI using an ASCII terminal.

  The ASMI on an ASCII terminal supports a subset of the functions that are provided by the web interface and is available only when the system is in the platform standby state. The ASMI on an ASCII console is not available during various phases of system operation, such as the IPL and run time.

The operator panel

The service processor provides an interface to the operator panel, which is used to display system status and diagnostic information.

The operator panel can be accessed in these ways:

- By using the normal operational front view.
- By pulling it out to access the switches and view the LCD display. Figure 4-6 shows that the operator panel on a Power 720 and Power 740 is pulled out.

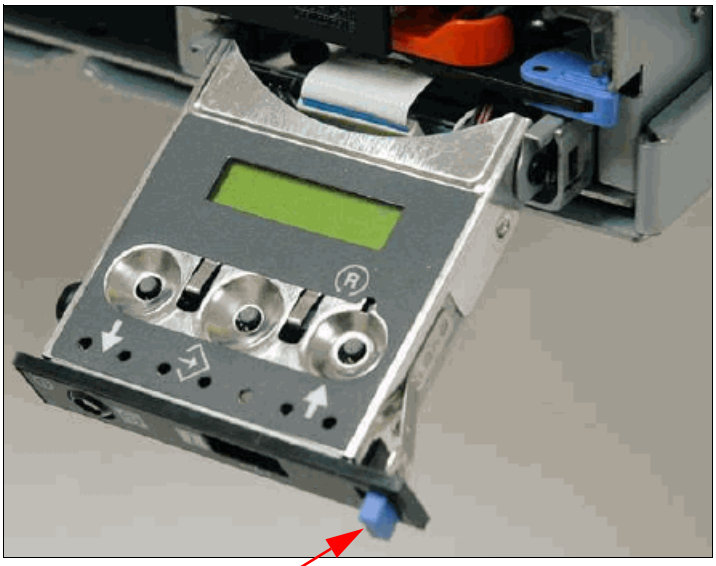


Figure 4-6 Operator panel is pulled out from the chassis

The operator panel includes features such as these:

- A 2 x 16 character LCD display
- Reset, enter, power On/Off, increment and decrement buttons
- Amber System Information/Attention, green Power LED
- Blue Enclosure Identify LED on the Power 720 and Power 740
- Altitude sensor
- USB Port
- Speaker/Beeper

These functions are available through the operator panel:

- Error Information
- Generate dump
- View Machine Type, Model, and Serial Number
- Limited set of repair functions

Operating system service menu

The system diagnostics consist of IBM i service tools, stand-alone diagnostics that are loaded from the DVD drive, and online diagnostics (available in AIX).

Online diagnostics, when installed, are a part of the AIX or IBM i operating system on the disk or server. 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

error log and the AIX configuration data. IBM i has a service tools problem log, IBM i history log (QHST), and IBM i problem log.

The available modes are as follows:

- Service mode

  This requires a service mode boot of the system and enables the checking of system devices and features. Service mode provides the most complete checkout of the system resources. All system resources, except the SCSI adapter and the disk drives used for paging, can be tested.

- Concurrent mode

  This enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, certain devices might require additional actions by the user or diagnostic application before testing can be done.

- Maintenance mode

  This enables the checking of most system resources. Maintenance mode provides the same test coverage as service mode. The difference between the two modes is the way that they are invoked. Maintenance mode requires that all activity on the operating system be stopped. The **shutdown -m** command is used to stop all activity on the operating system and put the operating system into maintenance mode.

The System Management Services (SMS) error log is accessible on the SMS menus. This error log contains errors that are found by partition firmware when the system or partition is booting.

You can access the service processor's error log on the ASMI menus.

You can also access the system diagnostics from an AIX Network Installation Management (NIM) server.

Note: When you order a Power System, a DVD-ROM or DVD-RAM might be optional. An alternate method for maintaining and servicing the system must be available if you do not order the DVD-ROM or DVD-RAM.

The IBM i operating system and associated machine code provide Dedicated Service Tools (DST) as part of the IBM i licensed machine code (Licensed Internal Code) and System Service Tools (SST) as part of the IBM i operating system. DST can be run in dedicated mode (no operating system loaded). DST tools and diagnostics are a superset of those available under SST.

The IBM i **End Subsystem** (ENDSBS *ALL) command can shut down all IBM and customer applications subsystems except the controlling subsystem QTCL. The **Power Down System** (PWRDWNSYS) command can be set to power down the IBM i partition and restart the partition in DST mode.

You can start SST during normal operations, which leaves all applications up and running, using the IBM i **Start Service Tools** (STRSST) command (when signed onto IBM i with the appropriately secured user ID).

With DST and SST, you can look at various logs, run various diagnostics, or take various kinds of system dumps or other options.

Depending on the operating system, the service-level functions that you typically see when using the operating system service menus are as follows:

- Product activity log
- Trace Licensed Internal Code
- Work with communications trace
- Display/Alter/Dump
- Licensed Internal Code log
- Main storage dump manager
- Hardware service manager
- Call Home/Customer Notification
- Error information menu
- LED management menu
- Concurrent/Non-concurrent maintenance (within scope of the OS)
- Managing firmware levels
  - Server
  - Adapter
- Remote support (access varies by OS)

Service Focal Point on the Hardware Management Console

Service strategies become more complicated in a partitioned environment. The Manage Serviceable Events task in the HMC can help to streamline this process.

Each logical partition reports errors that it detects and forwards the event to the Service Focal Point (SFP) application that is running on the HMC, without determining whether other logical partitions also detect and report the errors. For example, if one logical partition reports an error for a shared resource, such as a managed system power supply, other active logical partitions might report the same error.

By using the Manage Serviceable Events task in the HMC, you can avoid long lists of repetitive call-home information by recognizing that these are repeated errors and consolidating them into one error.

In addition, you can use the Manage Serviceable Events task to initiate service functions on systems and logical partitions, including the exchanging of parts, configuring connectivity, and managing dumps.

4.4.2 IBM Power Systems firmware maintenance

The IBM Power Systems Client-Managed Microcode is a methodology that enables you to manage and install microcode updates on Power Systems and associated I/O adapters.

The system firmware consists of service processor microcode, Open Firmware microcode, SPCN microcode, and the POWER Hypervisor.

The firmware and microcode can be downloaded and installed either from an HMC, from a running partition, or from USB port number 1 on the rear of a Power 720 and Power 740, if that system is not managed by an HMC.

Power Systems has a permanent firmware boot side, or A side, and a temporary firmware boot side, or B side. New levels of firmware must be installed on the temporary side first to test the update's compatibility with existing applications. When the new level of firmware has been approved, it can be copied to the permanent side.

For access to the initial web pages that address this capability, see the Support for IBM Systems web page:

`http://www.ibm.com/systems/support`

For Power Systems, select the **Power** link (Figure 4-7).

Figure 4-7 Support for Power servers web page

Although the content under the Popular links section can change, click the **Firmware and HMC updates** link to go to the resources for keeping your system's firmware current.

If there is an HMC to manage the server, the HMC interface can be use to view the levels of server firmware and power subsystem firmware that are installed and are available to download and install.

Each IBM Power Systems server has the following levels of server firmware and power subsystem firmware:

- Installed level

  This level of server firmware or power subsystem firmware has been installed and will be installed into memory after the managed system is powered off and then powered on. It is installed on the temporary side of system firmware.

- Activated level

  This level of server firmware or power subsystem firmware is active and running in memory.

- Accepted level

  This level is the backup level of server or power subsystem firmware. You can return to this level of server or power subsystem firmware if you decide to remove the installed level. It is installed on the permanent side of system firmware.

IBM provides the Concurrent Firmware Maintenance (CFM) function on selected Power Systems. This function supports applying nondisruptive system firmware service packs to the system concurrently (without requiring a reboot operation to activate changes). For systems that are not managed by an HMC, the installation of system firmware is always disruptive.

The concurrent levels of system firmware can, on occasion, contain fixes that are known as *deferred*. These deferred fixes can be installed concurrently but are not activated until the next IPL. Deferred fixes, if any, will be identified in the Firmware Update Descriptions table of the firmware document. For deferred fixes within a service pack, only the fixes in the service pack that cannot be concurrently activated are deferred. Table 4-1 shows the file-naming convention for system firmware.

Table 4-1 Firmware naming convention

PPNNSSS_FFF_DDD			
PP	Package identifier	01	-
		02	-
NN	Platform & Class	AL	Low End
		AM	Mid Range
		AS	Blade Server
		AH	High End
		AP	Bulk Power for IH
		AB	Bulk Power for High End
SSS	Release indicator		
FFF	Current fixpack		
DDD	Last disruptive fixpack		

This example uses the convention:

`01AL710_086 = POWER7 Entry Systems Firmware for 8233-E8B and 8236-E8B`

An installation is disruptive if the following statements are true:

- The release levels (SSS) of currently installed and new firmware differ.
- The service pack level (FFF) and the last disruptive service pack level (DDD) are equal in new firmware.

Otherwise, an installation is concurrent if the service pack level (FFF) of the new firmware is higher than the service pack level currently installed on the system and the conditions for disruptive installation are not met.

4.4.3 Electronic Services and Electronic Service Agent

IBM has transformed its delivery of hardware and software support services to help you achieve higher system availability. Electronic Services is a web-enabled solution that offers an exclusive, no-additional-charge enhancement to the service and support available for IBM servers. These services provide the opportunity for greater system availability with faster problem resolution and preemptive monitoring. The Electronic Services solution consists of two separate, but complementary, elements:

- Electronic Services news page

  The Electronic Services news page is a single internet entry point that replaces the multiple entry points that are traditionally used to access IBM internet services and support. The news page enables you to gain easier access to IBM resources for assistance in resolving technical problems.

- IBM Electronic Service Agent

  The Electronic Service Agent is software that resides on your server. It monitors events and transmits system inventory information to IBM on a periodic, client-defined timetable. The Electronic Service Agent automatically reports hardware problems to IBM.

Early knowledge about potential problems enables IBM to deliver proactive service that can result in higher system availability and performance. In addition, information that is collected through the Service Agent is made available to IBM service support representatives when they help answer your questions or diagnose problems. Installation and use of IBM Electronic Service Agent for problem reporting enables IBM to provide better support and service for your IBM server.

To learn how Electronic Services can work for you, visit:

`https://www.ibm.com/support/electronic/portal`

4.5 Operating system support for RAS features

Table 4-2 gives an overview of a number of features for continuous availability that are supported by the various operating systems running on the Power 720 and Power 740 systems.

Table 4-2 Operating system support for RAS features

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.1	SLES11 SP1
System deallocation of failing components							
Dynamic processor deallocation	X	X	X	X	X	X	X
Dynamic processor sparing	X	X	X	X	X	X	X
Processor instruction retry	X	X	X	X	X	X	X
Alternate processor recovery	X	X	X	X	X	X	X
Partition contained checkstop	X	X	X	X	X	X	X
Persistent processor deallocation	X	X	X	X	X	X	X
GX++ bus persistent deallocation	X	X	X	X	-	-	X
PCI bus extended error detection	X	X	X	X	X	X	X
PCI bus extended error recovery	X	X	X	X	Most	Most	Most
PCI-PCI bridge extended error handling	X	X	X	X	-	-	-
Redundant RIO or 12x Channel link	X	X	X	X	X	X	X
PCI card hot-swap	X	X	X	X	X	X	X
Dynamic SP failover at run time	X	X	X	X	X	X	X
Memory sparing with CoD at IPL time	X	X	X	X	X	X	X
Clock failover run time or IPL	X	X	X	X	X	X	X
Memory availability							
64-byte ECC code	X	X	X	X	X	X	X
Hardware scrubbing	X	X	X	X	X	X	X
CRC	X	X	X	X	X	X	X
Chipkill	X	X	X	X	X	X	X
L1 instruction and data array protection	X	X	X	X	X	X	X
L2/L3 ECC and cache line delete	X	X	X	X	X	X	X
Special uncorrectable error handling	X	X	X	X	X	X	X
Fault detection and isolation							
Platform FFDC diagnostics	X	X	X	X	X	X	X
Runtime diagnostics	X	X	X	X	Most	Most	Most
Storage Protection Keys	-	X	X	X	-	-	-

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.1	SLES11 SP1
Dynamic Trace	X	X	X	X	-	-	X
Operating System FFDC	-	X	X	X	-	-	-
Error log analysis	X	X	X	X	X	X	X
Service Processor support for:							
Built-in-Self-Tests (BIST) for logic and arrays	X	X	X	X	X	X	X
Wire tests	X	X	X	X	X	X	X
Component initialization	X	X	X	X	X	X	X
Serviceability							
Boot-time progress indicators	X	X	X	X	Most	Most	Most
Electronic Service Agent Call Home from HMC[a]	X	X	X	X	-	-	-
Firmware error codes	X	X	X	X	X	X	X
Operating system error codes	X	X	X	X	Most	Most	Most
Inventory collection	X	X	X	X	X	X	X
Environmental and power warnings	X	X	X	X	X	X	X
Hot-plug fans, power supplies	X	X	X	X	X	X	X
Extended error data collection	X	X	X	X	X	X	X
SP “call home” on non-HMC configurations	X	X	X	X	-	-	-
I/O drawer redundant connections	X	X	X	X	X	X	X
I/O drawer hot add and concurrent repair	X	X	X	X	X	X	X
Concurrent RIO/GX adapter add	X	X	X	X	X	X	X
Concurrent cold-repair of GX adapter	X	X	X	X	X	X	X
SP mutual surveillance with POWER Hypervisor	X	X	X	X	X	X	X
Dynamic firmware update with HMC	X	X	X	X	X	X	X
Electronic Service Agent Call Home Application	X	X	X	X	-	-	-
Lightpath LEDs	X	X	X	X	X	X	X
System dump for memory, POWER Hypervisor, SP	X	X	X	X	X	X	X
Infocenter/Systems Support Site service publications	X	X	X	X	X	X	X
System Support Site education	X	X	X	X	X	X	X
Operating system error reporting to HMC SFP	X	X	X	X	X	X	X
RMC secure error transmission subsystem	X	X	X	X	X	X	X
Health check scheduled operations with HMC	X	X	X	X	X	X	X
Operator panel (real or virtual)	X	X	X	X	X	X	X
Concurrent operator panel maintenance	X	X	X	X	X	X	X

RAS feature	AIX 5.3	AIX 6.1	AIX 7.1	IBM i	RHEL 5.7	RHEL 6.1	SLES11 SP1
Redundant HMCs	X	X	X	X	X	X	X
Automated server recovery/restart	X	X	X	X	X	X	X
High availability clustering support	X	X	X	X	X	X	X
Repair and Verify Guided Maintenance	X	X	X	X	Most	Most	Most
Concurrent kernel update	-	X	X	X	X	X	X

a. Electronic Service Agent via a managed HMC will report platform-level information but not Linux operating system detected errors.

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *IBM BladeCenter PS700, PS701, and PS702 Technical Overview and Introduction*, REDP-4655
- *IBM BladeCenter PS703 and PS704 Technical Overview and Introduction*, REDP-4744
- *IBM Power 710 and 730 (8231-E1C, 8231-E2C) Technical Overview and Introduction*, REDP-4796
- *IBM Power 750 and 755 (8233-E8B, 8236-E8C) Technical Overview and Introduction*, REDP-4638
- *IBM Power 770 and 780 (9117-MMC, 9179-MHC) Technical Overview and Introduction*, REDP-4798
- *IBM Power 795 (9119-FHB) Technical Overview and Introduction*, REDP-4640
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *IBM PowerVM Live Partition Mobility*, SG24-7460
- *IBM System p Advanced POWER Virtualization (PowerVM) Best Practices*, REDP-4194
- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM System Storage DS8000: Copy Services in Open Environments*, SG24-6788
- *IBM System Storage DS8700 Architecture and Implementation*, SG24-8786
- *PowerVM and SAN Copy Services*, REDP-4610
- *SAN Volume Controller V4.3.0 Advanced Copy Services*, SG24-7574

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- IBM Power Systems Facts and Features POWER7 Blades and Servers

  http://www.ibm.com/systems/power/hardware/reports/factsfeatures.html
- Specific storage devices supported for Virtual I/O Server

  http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html
- IBM Power 710 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03048usen/POD03048USEN.PDF
- IBM Power 720 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03048usen/POD03048USEN.PDF
- IBM Power 730 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03050usen/POD03050USEN.PDF
- IBM Power 740 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03051usen/POD03051USEN.PDF
- IBM Power 750 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03034usen/POD03034USEN.PDF
- IBM Power 755 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03035usen/POD03035USEN.PDF
- IBM Power 770 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03035usen/POD03035USEN.PDF
- IBM Power 780 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03032usen/POD03032USEN.PDF
- IBM Power 795 server Data Sheet

  http://public.dhe.ibm.com/common/ssi/ecm/en/pod03053usen/POD03053USEN.PDF
- *Active Memory Expansion: Overview and Usage Guide*

  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03037usen/POW03037USEN.PDF
- Migration combinations of processor compatibility modes for active Partition Mobility

  http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm
- Advance Toolchain for Linux website

  http://www.ibm.com/developerworks/wikis/display/hpccentral/How+to+use+Advance+Toolchain+for+Linux+on+POWER

Online resources

These websites are also relevant as further information sources:

- IBM Power Systems Hardware Information Center

  http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp
- IBM System Planning Tool website

  http://www.ibm.com/systems/support/tools/systemplanningtool/
- IBM Fix Central website

  http://www.ibm.com/support/fixcentral/
- Power Systems Capacity on Demand website

  http://www.ibm.com/systems/power/hardware/cod/
- Support for IBM Systems website

  http://www.ibm.com/support/entry/portal/Overview?brandind=Hardware~Systems~Power
- IBM Power Systems website

  http://www.ibm.com/systems/power/
- IBM Storage website

  http://www.ibm.com/systems/storage/

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM Power 720 and 740 Technical Overview and Introduction

Features the 8202-E4C and 8205-E6C based on the latest POWER7 processor technology

Discusses reduced power requirements with power-saving functions

Describes leading midrange performance

This IBM Redpaper publication is a comprehensive guide covering the IBM Power 720 (8202-E4C) and Power 740 (8205-E6C) servers supporting AIX, IBM i, and Linux operating systems. The goal of this paper is to introduce the major innovative Power 720 and Power 740 offerings and their prominent functions, including these:

- The IBM POWER7 processor available at frequencies of 3.0 GHz, 3.55 GHz, and 3.7 GHz.
- The specialized POWER7 Level 3 cache that provides greater bandwidth, capacity, and reliability.
- The 2-port 10/100/1000 Base-TX Ethernet PCI Express adapter included in the base configuration and installed in a PCIe Gen2 x4 slot.
- The integrated SAS/SATA controller for HDD, SSD, tape, and DVD. This controller supports built-in hardware RAID 0, 1, and 10.
- The latest IBM PowerVM virtualization, including PowerVM Live Partition Mobility and PowerVM IBM Active Memory Sharing.
- Active Memory Expansion technology that provides more usable memory than is physically installed in the system.
- IBM EnergyScale technology that provides features such as power trending, power-saving, capping of power, and thermal measurement.

Professionals who want to acquire a better understanding of IBM Power Systems products can benefit from reading this paper.

This paper expands the current set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power 720 and Power 740 systems.

REDP-4797-00

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks