IBM Power Systems S814 and S824 Technical Overview and Introduction

- Outstanding performance based on POWER8 processor technology
- 4U scale-out desktop and rack-mount servers
- Improved reliability, availability, and serviceability features

Alexandre Bicas Caldeira
Bartłomiej Grabowski
Volker Haug
Marc-Eric Kahle
Andrew Laidlaw
Cesar Diniz Maciel
Monica Sanchez
Seulgi Yoppy Sung

Redpaper

ibm.com/redbooks

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000001
IPR2025-00223 (Advanced Micro Devices, Inc., Pensando Systems, Inc. v. Concurrent Ventures, LLC, XtreamEdge, Inc.)

International Technical Support Organization

IBM Power System S814 and S824 Technical Overview and Introduction

August 2014

REDP-5097-00

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000003

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

First Edition (August 2014)

This edition applies to IBM Power System S814 (8286-41A) and IBM Power System S824 (8286-42A) systems.

© Copyright International Business Machines Corporation 2014. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices . . . vii
Trademarks . . . viii

Preface . . . ix
Authors . . . ix
Now you can become a published author, too! . . . xi
Comments welcome . . . xi
Stay connected to IBM Redbooks . . . xii

Chapter 1. General description . . . 1
1.1 Systems overview . . . 2
1.1.1 Power S814 server . . . 2
1.1.2 Power S824 server . . . 3
1.2 Operating environment . . . 4
1.3 Physical package . . . 5
1.3.1 Tower model . . . 5
1.3.2 Rack-mount model . . . 6
1.4 System features . . . 8
1.4.1 Power S814 system features . . . 8
1.4.2 Power S824 system features . . . 9
1.4.3 Minimum features . . . 10
1.4.4 Power supply features . . . 10
1.4.5 Processor module features . . . 10
1.4.6 Memory features . . . 11
1.4.7 PCIe slots . . . 12
1.5 Disk and media features . . . 13
1.6 I/O drawers for Power S814 and Power S824 servers . . . 20
1.6.1 PCIe Gen3 I/O expansion drawer . . . 21
1.6.2 I/O drawers and usable PCI slot . . . 22
1.6.3 EXP24S SFF Gen2-bay drawer . . . 23
1.7 IBM i Solution Edition for Power S814 and Power S824 . . . 24
1.8 Capacity backup . . . 24
1.8.1 Power System S814 (8286-41A) capacity backup offering . . . 24
1.8.2 Power System S824 (8286-42A) capacity backup capability . . . 26
1.9 Server and virtualization management . . . 28
1.10 System racks . . . 29
1.10.1 IBM 7014 Model T00 rack . . . 29
1.10.2 IBM 7014 Model T42 rack . . . 30
1.10.3 IBM 42U Slim Rack 7965-94Y . . . 32
1.10.4 Feature code 0551 rack . . . 32
1.10.5 Feature code 0553 rack . . . 32
1.10.6 Feature code ER05 rack . . . 32
1.10.7 The AC power distribution unit and rack content . . . 33
1.10.8 Rack-mounting rules . . . 35
1.10.9 Useful rack additions . . . 35
1.10.10 OEM rack . . . 38

Chapter 2. Architecture and technical overview . . . 41
2.1 The IBM POWER8 processor . . . 43

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000005

2.1.1 POWER8 processor overview . . . 44
2.1.2 POWER8 processor core . . . 46
2.1.3 Simultaneous multithreading . . . 47
2.1.4 Memory access . . . 48
2.1.5 On-chip L3 cache innovation and Intelligent Cache . . . 48
2.1.6 Level 4 cache and memory buffer . . . 49
2.1.7 Hardware transactional memory . . . 50
2.1.8 Coherent Accelerator Processor Interface . . . 51
2.1.9 Power management and system performance . . . 52
2.1.10 Comparison of the POWER8, POWER7+, and POWER7 processors . . . 53
2.2 Memory subsystem . . . 53
2.2.1 Custom DIMM . . . 54
2.2.2 Memory placement rules . . . 55
2.2.3 Memory bandwidth . . . 57
2.3 System bus . . . 60
2.4 Internal I/O subsystem . . . 61
2.4.1 Slot configuration . . . 61
2.4.2 System ports . . . 64
2.5 PCI adapters . . . 64
2.5.1 PCI Express . . . 64
2.5.2 LAN adapters . . . 66
2.5.3 Graphics accelerator adapters . . . 66
2.5.4 SAS adapters . . . 67
2.5.5 Fibre Channel adapters . . . 67
2.5.6 Fibre Channel over Ethernet . . . 68
2.5.7 InfiniBand Host Channel adapter . . . 69
2.5.8 Asynchronous and USB adapters . . . 70
2.5.9 Cryptographic coprocessor . . . 70
2.5.10 Field-programmable gate array adapters . . . 71
2.5.11 CAPI adapters . . . 71
2.5.12 Flash storage adapters . . . 72
2.6 Internal storage . . . 72
2.6.1 RAID support . . . 77
2.6.2 Easy Tier . . . 79
2.6.3 External SAS port . . . 81
2.6.4 Media bays . . . 82
2.7 External I/O subsystems . . . 82
2.7.1 PCIe Gen3 I/O expansion drawer . . . 82
2.7.2 PCIe Gen3 I/O expansion drawer optical cabling . . . 83
2.7.3 PCIe Gen3 I/O expansion drawer SPCN cabling . . . 86
2.8 External disk subsystems . . . 86
2.8.1 EXP24S SFF Gen2-bay drawer . . . 86
2.8.2 IBM System Storage . . . 88
2.9 Hardware Management Console (optional) . . . 89
2.9.1 HMC code level . . . 90
2.9.2 HMC RAID 1 support . . . 91
2.9.3 HMC connectivity to the POWER8 processor-based systems . . . 91
2.9.4 High availability HMC configuration . . . 93
2.10 Operating system support . . . 94
2.10.1 AIX operating system . . . 94
2.10.2 IBM i . . . 95
2.10.3 Linux operating system . . . 95
2.10.4 Virtual I/O Server . . . 96

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000006

2.10.5 Java 96
2.11 Energy management 96
2.11.1 IBM EnergyScale technology 97
2.11.2 On Chip Controller 100
2.11.3 Energy consumption estimation 101

Chapter 3. Virtualization 103
3.1 IBM POWER Hypervisor 104
3.1.1 Virtual SCSI 105
3.1.2 Virtual Ethernet 105
3.1.3 Virtual Fibre Channel 106
3.1.4 Virtual (TTY) console 106
3.2 POWER processor modes 106
3.3 Active Memory Expansion 109
3.4 Single Root I/O Virtualization 113
3.4.1 Direct access I/O and performance 114
3.4.2 Adapter sharing 115
3.4.3 Adapter resource provisioning (QoS) 115
3.4.4 Flexible deployment 115
3.4.5 Reduced costs 116
3.5 PowerVM 116
3.5.1 PowerVM editions 116
3.5.2 Logical partitions 116
3.5.3 Multiple shared processor pools 121
3.5.4 Virtual I/O Server 122
3.5.5 PowerVM Live Partition Mobility 126
3.5.6 Active Memory Sharing 127
3.5.7 Active Memory Deduplication 128
3.5.8 Operating system support for PowerVM 128
3.5.9 Linux support 130
3.5.10 PowerVM simplification 130
3.6 System Planning Tool 132
3.7 IBM Power Virtualization Center 133
3.8 IBM Power Virtualization Performance 134

Chapter 4. Reliability, availability, and serviceability 137
4.1 Introduction 138
4.1.1 RAS enhancements of POWER8 processor-based scale-out servers 138
4.2 Reliability 139
4.2.1 Designed for reliability 139
4.2.2 Placement of components 140
4.3 Processor/Memory availability details 141
4.3.1 Correctable error introduction 141
4.3.2 Uncorrectable error introduction 142
4.3.3 Processor Core/Cache correctable error handling 142
4.3.4 Processor Instruction Retry and other try again techniques 142
4.3.5 Alternative processor recovery and Partition Availability Priority 143
4.3.6 Core Contained Checkstops and other PowerVM error recovery 143
4.3.7 Cache uncorrectable error handling 143
4.3.8 Other processor chip functions 144
4.3.9 Other fault error handling 144
4.3.10 Memory protection 145
4.3.11 I/O subsystem availability and Enhanced Error Handling 146

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000007

4.4 Serviceability 147
4.4.1 Detecting introduction 147
4.4.2 Error checkers, fault isolation registers, and First-Failure Data Capture 147
4.4.3 Service processor 148
4.4.4 Diagnosing 149
4.4.5 Reporting 150
4.4.6 Notification 153
4.4.7 Locating and servicing 154
4.5 Manageability 157
4.5.1 Service user interfaces 157
4.5.2 IBM Power Systems Firmware maintenance 162
4.5.3 Concurrent firmware maintenance improvements 166
4.5.4 Electronic Services and Electronic Service Agent 166
4.6 Selected POWER8 RAS capabilities by operating system 170

Related publications 173
IBM Redbooks 173
Other publications 173
Online resources 174
Help from IBM 175

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000008

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 grant 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 measurements 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.

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000009

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/us/en/copytrade.shtml

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

Active Memory™
AIX®
DS8000®
Easy Tier®
Electronic Service Agent™
EnergyScale™
eServer™
FlashSystem™
Focal Point™
Global Technology Services®
IBM®
IBM FlashSystem®
Micro-Partitioning®
POWER®
POWER Hypervisor™
Power Systems™
Power Systems Software™
POWER6®
POWER6+™
POWER7®
POWER7+™
POWER8™
PowerHA®
PowerLinux™
PowerPC®
PowerVM®
PowerVP™
PureFlex®
Real-time Compression™
Redbooks®
Redpaper™
Redbooks (logo) ®
RS/6000®
Storwize®
System p®
System Storage®
System z®
Tivoli®
XIV®

The following terms are trademarks of other companies:

Inc., and Inc. device are trademarks or registered trademarks of Kenexa, an IBM Company.

Intel, Intel Xeon, 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.

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

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.

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000010

Preface

This IBM® Redpaper™ publication is a comprehensive guide covering the IBM Power System S814 (8286-41A) and IBM Power System S824 (8286-42A) servers that support IBM AIX®, IBM i, and Linux operating systems. The objective of this paper is to introduce the major innovative Power S814 and Power S824 offerings and their relevant functions:

- The new IBM POWER8™ processor, available at frequencies of 3.02 GHz, 3.52 GHz, 3.72 GHz, 3.89 GHz, and 4.15 GHz
- Significantly strengthened cores and larger caches
- Two integrated memory controllers with improved latency and bandwidth
- Integrated I/O subsystem and hot-pluggable PCIe Gen3 I/O slots
- I/O drawer expansion options offers greater flexibility
- Improved reliability, serviceability, and availability (RAS) functions
- IBM EnergyScale™ technology that provides features such as power trending, power-saving, capping of power, and thermal measurement

This publication is for professionals who want to acquire a better understanding of IBM Power Systems™ products. The intended audience includes the following roles:

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

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 S814 and Power S824 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.

Authors

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 is a Certified IT Specialist and is a member of the Power Systems Advanced Technical Sales Support team for IBM Brazil. He holds a degree in Computer Science from the Universidade Estadual Paulista (UNESP) and an MBA in Marketing. His major areas of focus are competition, sales, and technical sales support. Alexandre has more than 14 years of experience working on IBM Systems & Technology Group Solutions and has worked also as an IBM Business Partner on Power Systems hardware, AIX, and IBM PowerVM® virtualization products.

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000011

Bartłomiej Grabowski is an IBM i and PowerVM Senior Technical Specialist in DHL IT Services in the Czech Republic. He has nine years of experience with IBM i. He holds a Bachelor's degree in Computer Science from the Academy of Computer Science and Management in Bielsko-Biala. His areas of expertise include IBM i administration, PowerHA® solutions that are based on hardware and software replication, Power Systems hardware, and PowerVM. He is an IBM Certified Systems Expert and a coauthor of several PowerVM IBM Redbooks® publications.

Volker Haug is an Open Group Certified IT Specialist within IBM Systems in Germany supporting Power Systems clients and Business Partners. He holds a Diploma degree in Business Management from the University of Applied Studies in Stuttgart. His career includes more than 28 years of experience with Power Systems, AIX, and PowerVM virtualization. He has written several IBM Redbooks publications about Power Systems and PowerVM. Volker is an IBM POWER8 Champion and a member of the German Technical Expert Council, which is an affiliate of the IBM Academy of Technology.

Marc-Eric Kahle is a AIX Software specialist at the IBM Global Technology Services® in Ehningen, Germany. He has worked as a Power Systems Hardware Support specialist in the IBM RS/6000®, Power Systems, and AIX fields since 1993. He has worked at IBM Germany since 1987. His areas of expertise include Power Systems hardware, and he is an AIX certified specialist. He has participated in the development of seven other IBM Redbooks publications.

Andrew Laidlaw is a Client Technical Specialist for IBM working in the UK. He supports Service Provider clients within the UK and Ireland, focusing primarily on Power Systems running AIX and Linux workloads. His expertise extends to open source software package including the KVM hypervisor and various management tools. Andrew holds an Honors degree in Mathematics from the University of Leeds, which includes credits from the University of California in Berkeley.

Cesar Diniz Maciel is an Executive IT Specialist with IBM in the United States. He joined IBM in 1996 as Presales Technical Support for the IBM RS/6000 family of UNIX servers in Brazil, and came to IBM United States in 2005. He is part of the Global Techline team, working on presales consulting for Latin America. He holds a degree in Electrical Engineering from Universidade Federal de Minas Gerais (UFMG) in Brazil. His areas of expertise include Power Systems, AIX, and IBM POWER® Virtualization. He has written extensively on Power Systems and related products. This is his eighth ITSO residency.

Monica Sanchez is an Advisory Software Engineer with more than 13 years of experience in AIX and Power Systems support. Her areas of expertise include AIX, HMC, and networking. She holds a degree in Computer Science from Texas A&M University and is part of the Power HMC Product Engineering team, providing level 2 support for the IBM Power Systems Hardware Management Console.

Seulgi Yoppy Sung is a very passionate Engineer, supporting multi-platform systems as a System Services Representative almost three year, include Power System hardware, AIX, high-end and low-end storage DS8000® and V7000. She is very positive and enthusiastic about Power Systems.

The project that produced this publication was managed by:

Scott Vetter
Executive Project Manager, PMP

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000012

George Ahrens, Kan Bahri, Tamikia Barrow, Terry Brennan, Ron Brown, Carl Burnett, Charlie Burns, Bill Buros, Johnathan Dement, Dan Dumarot, Jessica Erber-Stark, Medha D. Fox, Ben Gibbs, Chuck Graham, Dan Henderson, Tenley Jackson, Kris Kendall, Mark Kressin, Karen Lawrence, Woodrow Lemcke, Edward Liu, Fredrik Lundholm, Pat Mackall, Michael MacPherson, Ricardo Marin Matinata, Bruce Mealey, Dirk Michel, Michael J. Mueller, Thoi Ngyuen, Mark Olson, Bret Olszewski, Neil Palmer, Kanisha Patel, Rajendra Patel, Vani Ramagiri, Audrey Romonosky, Pat O'Rourke, Armin Roell, Todd Rosedahl, Jeff Scheel, Craig Shempert, Guillermo J. Silva, Jeff Stuecheli, Chris Sturgill, Claire Toth, Madeline Vega, Julian Wang
IBM

Louis Bellanger
Bull

Yuki Taniguchi
Hitachi

Mark Applegate
Avnet, Inc.®

Klaus Heer
PROFI Engineering Systems AG

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000013

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000014

- 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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000015

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000016

1

General description

The IBM Power System S814 (8286-41A) and IBM Power System S824 (8286-42A) servers use the latest POWER8 processor technology that is designed to deliver unprecedented performance, scalability, reliability, and manageability for demanding commercial workloads.

The Power S814 is a highly secure architecture providing a stable database and middleware platform for efficient deployment of business processing applications. A one socket system with a 4-core or 6-core POWER8 processor is available in either a rack or tower configuration. The 8-core higher performance system is available only in a rack configuration. All the models support new I/O capabilities, including Coherent Accelerator Processor Interface (CAPI) accelerators, higher internal disk and SSD capacities, and hot-plug PCIe Gen3 slots. The architecture provides a stable highly secure architecture that provides a stable database and middleware platform for efficient deployment of business processing applications.

The 4-core option for the Power S814 is especially attractive to IBM i clients with its P05 software tier, while the other processor options makes the system a P10 tier.

The Power S824 server is designed to put data to work. With a strong commitment to optimizing AIX and IBM i workloads, this system delivers better performance than the prior generation of systems and offers unmatched price and performance value for integrated Linux applications. The one or two socket servers provide the ideal foundation for private and public cloud infrastructure. The Power S824 is available in 6-core, 8-core, 12-core, 16-core, and 24-core configurations.

The Power S814 and Power S824 are ideal for mid-sized business customers consolidating AIX, IBM i, and Linux workloads and for virtualized application servers. The database can run on the same server. Gain faster insights with the IBM POWER8 processor and smart acceleration that is enabled by CAPI technologies, and reduce energy consumption by using advanced energy control.

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000017

1.1 Systems overview

The following sections provide detailed information about the Power S814 and the Power S824 servers.

1.1.1 Power S814 server

The Power S814 (8286-41A) server is a powerful one socket server that ships with up to eight activated cores. It has the I/O configuration flexibility to meet today's growth and tomorrow's processing needs. A one socket system with a 4-core or 6-core POWER8 processor is available in either rack (19-inch rack-mount 4U (EIA units)) or tower configurations. The 8-core higher performance system is available only in a rack configuration.

The Power S814 server supports a maximum of eight DDR3 CDIMM slots. The memory features (two memory DIMMs per feature) that are supported are 16 GB, 32 GB, 64 GB and 128 GB, and run at speeds of 1600 MHz, allowing for a maximum system memory of 1 TB.

If using the 4-core processor feature #EPXK the system is limited to four memory DDR3 CDIMMS slots supporting only the 16 GB CDIMMs (#EM8B), which allows for a maximum of 64 GB RAM per system.

The IBM Active Memory™ Expansion feature enables memory expansion by using compression and decompression of memory content. This can effectively expand the maximum memory capacity if additional server workload capacity and performance are available.

Two different features for the storage backplane are available:

- EJ0N provides twelve *hot-plug* SFF-3 bays and one DVD bay It also gives you the option to split the back plane using #EJ0S.
- EJ0P provides eighteen *hot-plug* SFF-3 bays, one DVD bay, and dual IOA with write cache. For the 42A, it also provides the 1.8" SSD enclosure, #EJTM. This feature is required when EL3U is configured on the 42A. The EJTM feature is not available when EJ0P is configured on the model 41A.

Each of the three backplane options uses leading-edge, integrated SAS RAID controller technology that is designed and patented by IBM.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000018

Figure 1-1 shows the front view of a tower and rack-mount Power S814.

Figure 1-1 Front view of the tower and rack-mount Power S814

1.1.2 Power S824 server

The Power S824 (8284-42A) server is a powerful one or two socket server that ships with up to 24 activated cores. It has the I/O configuration flexibility to meet today's growth and tomorrow's processing needs. This server supports two processor sockets, offering 6-core or 12-core 3.89 GHz, 8-core or 16-core 4.15 GHz, or 24-core 3.52 GHz configurations in a 19-inch rack-mount, 4U (EIA units) drawer configuration. All the cores are active.

The Power S824 server supports a maximum of 16 DDR3 CDIMM slots. The memory features (two memory DIMMs per feature) that are supported are 16 GB, 32 GB, 64 GB and 128 GB, and run at speeds of 1600 MHz, allowing for a maximum system memory of 2 TB.

The Active Memory Expansion feature enables memory expansion by using compression and decompression of memory content. This can effectively expand the maximum memory capacity if additional server workload capacity and performance are available.

Two different features for the storage backplane are available:

- EJ0N provides twelve *hot-plug* SFF-3 bays and one DVD bay It also gives you the option to split the back plane using #EJ0S.
- EJ0P provides eighteen *hot-plug* SFF-3 bays, one DVD bay, and dual IOA with write cache. For the 42A, it also provides the 1.8" SSD enclosure, #EJTM. This feature is required when EL3U is configured on the 42A. The EJTM feature is not available when EJ0P is configured on the model 41A.

Each of the three backplane options uses leading-edge, integrated SAS RAID controller technology that is designed and patented by IBM.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000019

Figure 1-2 shows front view of the Power S824.

Figure 1-2 Front view of the Power S824

1.2 Operating environment

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

Table 1-1 Operating environment for Power S814 and Power S824

Power S814 and Power S824 operating environment				
Description	**Operating**		**Non-operating**	
	Power S814	**Power S824**	**Power S814**	**Power S824**
Temperature	Allowable: 5 - 35[a] degrees C (41 - 95 degrees F) Recommended: 18 - 27 degrees C (64 - 80 degrees F)		5 - 45 degrees C (41 - 113 degrees F)	
Relative humidity	8 - 80%		8 - 80%	
Maximum dew point	28 degrees C (84 degrees F)		N/A	
Operating voltage	900 W Power Supply (tower only): 100 - 127 V AC or 200 - 240 V AC 1400 W Power Supply (rack only): 200 - 240 V AC	900 W Power Supply (rack only): 100 - 127 V AC or 200 - 240 V AC	N/A	
Operating frequency	47 - 63 Hz (AC)		N/A	
Power consumption	1420 watts maximum	2300 watts maximum	N/A	
Power source loading	1.48 kVA maximum	2.38 kVA maximum	N/A	
Thermal output	4845 BTU/hour maximum	7848 BTU/hour maximum	N/A	

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000020

Power S814 and Power S824 operating environment				
Description	**Operating**		**Non-operating**	
	Power S814	**Power S824**	**Power S814**	**Power S824**
Maximum altitude	3,050 m (10,000 ft)		N/A	
Noise level and sound power	Tower System: 5.8 bels (operating) 5.5 bels (idle) Rack system: 6.0 bels (operating) 6.0 bels (idle)	Rack system: 6.5 bels (operating) 6.5 bels (idle)	N/A	

a. Heavy workloads may see some performance degradation above 35 degrees C if internal temperatures trigger a CPU clock reduction.

Tip: 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 operating conditions that are not ideal. Power consumption and heat load vary greatly by server configuration and usage. Use the IBM Systems Energy Estimator to obtain a heat output estimate based on a specific configuration. It is available at the following website:

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

1.3 Physical package

The Power S814 is available in both rack-mount and tower form factors. The Power S824 is available in rack-mount form factor only. The major physical attributes for each are described in the following sections.

1.3.1 Tower model

The Power S814 can be configured as tower models by selecting the features that are shown in Table 1-2.

Table 1-2 Features for selecting tower models

Tower description	Feature code and description
Power S814 tower with #EJ0N storage backplane	Front door and covers for 12-bay backplane (#EJTG)
Power S814 tower with #EJ0P storage backplane	Front door and covers for 18-bay backplane (#EJTH)

Table 1-3 on page 6 shows the physical dimensions of the Power S814 tower chassis.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000021

Table 1-3 Physical dimensions of the Power S814 tower chassis

Dimension	Power S814 (8286-41A)
Width	486.1 mm (19.1 in.)
Depth	751.7 mm (29.6 in.)
Height	182.4 mm (7.2 in.)
Weight (maximum configuration)	51.89 kg (114.4 lbs)

Figure 1-3 shows the rear view of the tower Power S814 system.

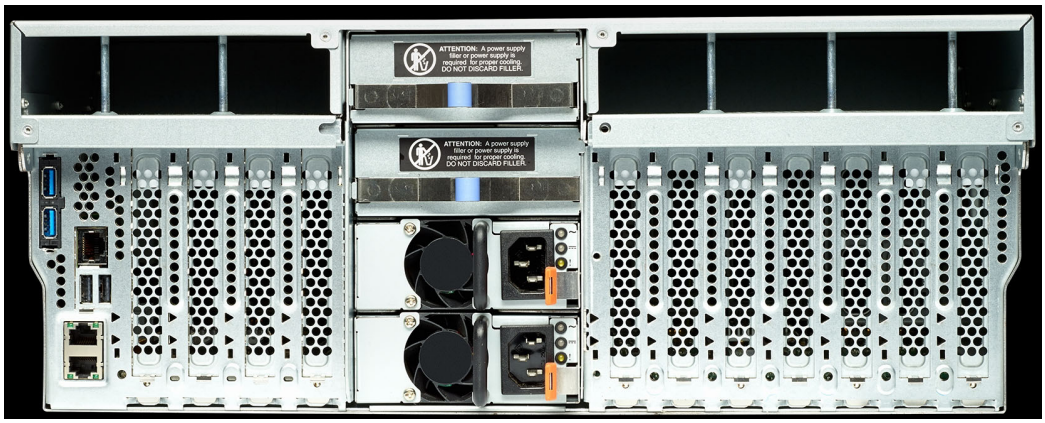

Figure 1-3 Rear view of a tower Power S814 system

1.3.2 Rack-mount model

The Power S814 and Power S824 can be configured as a 4U (4 EIA) rack-mount model by selecting the features that are shown in Table 1-4.

Table 1-4 Features for selecting rack-mount models

Rack-mount server description	Feature code and description
Power S814 and Power S824 rack-mount with #EJ0N storage backplane	Front bezel for 12-bay backplane (#EJT8)
Power S814 and Power S824 rack-mount with #EJ0P storage backplane	Front bezel for 18-bay backplane (#EJT9)

Table 1-5 shows the physical dimensions of the rack-mount Power S814 and Power S824 chassis. The server is available only in a rack-mounted form factor and takes 4U (4 EIA units) of rack space.

Table 1-5 Physical dimensions of the rack-mount Power S814 and Power S824 chassis

Dimension	Power S814 (8286-41A)	Power S824 (8286-42A)
Width	443 mm (17.5 in.)	443 mm (17.5 in.)
Depth	756 mm (29.8 in.)	756 mm (29.8 in.)
Height	173 mm (6.9 in.)	173 mm (6.9 in.)
Weight (maximum configuration)	43.8 kg (97 lbs)	43.8 kg (97 lbs)

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000022

Figure 1-4 shows the front view of a rack-mount Power S814 system.

Figure 1-4 Front view of a rack-mount Power S814 system

Figure 1-5 shows the front view of the Power S824 system.

Figure 1-5 Front view of the Power S824 system

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000023

1.4 System features

The system chassis contains one processor module (Power S814) or up to two processor modules (Power S824). Each POWER8 processor module is either 4-core, 6-core, or 8-core for the Power S814, and 6-core, 8-core, or 12-core for the Power S824. Each of the POWER8 processor chips in the server has a 64-bit architecture, up to 512 KB of L2 cache per core, and up to 8 MB of L3 cache per core.

1.4.1 Power S814 system features

This summary describes the standard features of the Power S814:

- Tower or 19-inch rack-mount (4U) chassis.
- Configuration with one of the following POWER8 processor modules:
  - 4-core 3.02 GHz.
  - 6-core 3.02 GHz.
  - 8-core 3.72 GHz (only available with the rack configuration).
- High-performance 1600 MHz DDR3 error-correcting code (ECC) memory with 16 GB, 32 GB, 64 GB, or 128 B memory features options.
- Up to 1 TB memory with one processor DCM.

Note: When using 4-core processor feature #EPXK, the maximum available memory is 64 GB

- Choice of two storage features:
  - Choice one:
    - Twelve SFF-3 bays, one DVD bay, one integrated SAS controller without cache, and JBOD, RAID 0, 5, 6, or 10.
    - Optionally, split the SFF bays and add a second integrated SAS controller without cache.
  - Choice two:
    - Eighteen SFF-3 bays, one DVD bay, a pair of integrated SAS controllers with cache, RAID 0, 5, 6, 10, 5T2, 6T2, and 10T2.
    - Optionally, attach an EXP24S SAS HDD/SSD Expansion Drawer to the dual IOA.
- Hot swap PCIe Gen3 slots. Seven slots with one POWER8 processor DCM: two x16 slots and five x8 slots.

Note:

- One of the x8 PCIe slots on the Power S814 and Power S824 servers is used for a PCIe2 4-port 1 Gb Ethernet Adapter (#5899).
- One fewer PCIe slot is available with the dual IOA storage backplane feature EJ0P.

- One DVD-RAM drive.
- Integrated features:
  - Service processor.
  - EnergyScale technology.
  - Hot-swap and redundant cooling.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000024

- Four USB 3.0 ports for general use.
- Two USB 2.0 ports for non-general use.
- Two HMC 1 Gb Ethernet RJ45 ports.
- One System port with RJ45 connector.

- Two hot-plug, redundant power supplies supporting a tower.
- Two hot-plug, redundant power supplies supporting a rack (AC or DC).

1.4.2 Power S824 system features

This summary describes the standard features of the Power S824:

- 19-inch rack-mount (4U) chassis
- Configuration with one or two of the following POWER8 processor modules:
  - 6-core or 12-core 3.89 GHz
  - 8-core or 16-core 4.15 GHz
  - 24-core 3.52 GHz
- High-performance 1600 MHz DDR3 ECC memory with 16 GB, 32 GB, 64 GB, or 128 GB memory features options
- Up to 1 TB memory with one POWER8 processor DCM
- Up to 2 TB memory with two POWER8 processor DCMs
- Choice of two storage features:
  - Choice one:
    - Twelve SFF bays, one DVD bay, one integrated SAS controller without cache, and JBOD, RAID 0, 5, 6, or 10.
    - Optionally, split the twelve SFF bays and add a second integrated SAS controller without cache.
  - Choice two:
    - Eighteen SFF bays; one DVD bay, a pair of integrated SAS controllers with cache, RAID 0, 5, 6, 10, 5T2, 6T2, and 10T2, and an 8-bay 1.8-inch SSD cage.
    - Optionally, attach an EXP24S SAS HDD/SSD Expansion Drawer to the dual IOA.
- Hot swap PCIe Gen3 Slots:
  - Eleven slots with two Power System processor DCMs: Four x16 slots and seven x8 slots
  - Seven slots with one Power System processor DCM: Two x16 slots and five x8 slots

Note:

- One of the x8 PCIe slots on the Power S814 and Power S824 servers is used for a PCIe2 4-port 1 Gb Ethernet Adapter (#5899).
- One fewer PCIe slot is available with the dual IOA storage backplane feature EJ0P.

- Integrated features:
  - Service processor
  - EnergyScale technology
  - Hot-swap and redundant cooling
  - Four USB 3.0 ports for general use
  - Two USB 2.0 ports for non-general use

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000025

- Two HMC 1 Gb Ethernet RJ45 ports
- One System port with RJ45 connector

- Four hot-plug, redundant power supplies (AC or DC)

1.4.3 Minimum features

The minimum Power S814 initial order must include a processor module, processor activations, 16 GB of memory, a storage backplane, one HDD or SDD DASD device (unless Fibre Channel adapters are configured to boot from SAN), a PCIe2 4-port 1 Gb Ethernet adapter (#5899), two power supplies and power cords, an operating system indicator, a cover set indicator, and a Language Group Specify.

The minimum Power S824 initial order must include a processor module, processor activations, 32 GB of memory, a storage backplane, one HDD or SDD DASD device (unless Fibre Channel adapters are configured to boot from SAN), a PCIe2 4-port 1 Gb Ethernet adapter (#5899), four power supplies and power cords, an operating system indicator, a cover set indicator, and a Language Group Specify.

1.4.4 Power supply features

The Power S814 supports the following power supplies:

- Two 900 watt 100-127 volt or 200-240 Volt AC options (#EB2L) power supplies
- Two 1400 watt 200-240 volt (#EB2M) optional power supplies
- Two 1400 watt 180-400 volt (#EB2N) optional DC power supplies

The Power S824 supports four 900 Watt 100-127 Volt or 200-240 Volt AC options (#EB2L) power supplies supporting a rack chassis.

One power supply is required on the one socket system for normal system operation; the second is for redundancy.

Two power supplies are required for normal system operation on a two socket system; the third and fourth power supplies are for redundancy.

1.4.5 Processor module features

The following section describes the processor modules that are available for the Power S814 and Power S824 servers.

Power S814 processor modules

The Power S814 supports a maximum of one processor of either four processor cores (#EPXK), six processor cores (#EPX0) or eight processor cores (#EPX6). All processor cores must be activated unless they are factory deconfigured by using feature code 2319.

Note:

- The four core processor card (#EPXK) limits the maximum memory to 64 GB per system.
- The eight core processor card (#EPX6) is available only with the rack configuration.

Table 1-6 on page 11 summarizes the processor features that are available for the Power S814.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000026

Table 1-6 Processor features for the Power S814

Feature code	Processor module description
EPXK	4-core 3.02 GHz POWER8 processor card
EPX0	6-core 3.02 GHz POWER8 processor card
EPX6	8-core 3.72 GHz POWER8 processor card

Power S824 processor modules

The Power S824 supports a configuration of a one or two processor modules with six, eight, or twelve processor cores. All processor cores must be activated.

The following listing defines the allowed quantities of processor activation entitlements:

- One 6-core, 3.89 GHz processor (#EPXE) requires that six processor activation codes be ordered. A maximum of six processor activation code features (6 x #EPYE, or 3 x #EPYE and 3 x #EPZE) is allowed per processor.
- Two 6-core, 3.89 GHz processors (#EPXE) require that twelve processor activation codes be ordered. A maximum of twelve processor activation code features (12 x #EPYE) is allowed per processor.
- One 8-core, 4.15 GHz processor (#EPXF) requires that eight processor activation codes be ordered. A maximum of eight processor activation code features (8 x #EPYF) is allowed per processor.
- Two 8-core, 4.15 GHz processors (#EPXF) require that 16 processor activation codes be ordered. A maximum of 16 processor activation code features (16 x #EPYF) is allowed per processor.
- Two 12-core, 3.52 GHz processor (#EPXZ) requires that 24 processor activation codes be ordered. A maximum of 24 processor activation code features (12 x #EPYZ) is allowed per processor.

Table 1-7 summarizes the processor features that are available for the Power S824.

Table 1-7 Processor features for the Power S824

Feature code	Processor module description
EPXE	6-core 3.89 GHz POWER8 processor card
EPXF	8-core 4.15 GHz POWER8 processor card
EPXZ	12-core 3.52 GHz POWER8 processor card

1.4.6 Memory features

A minimum of 16 GB of memory is required on the Power S814 system. The initial order for an 8286-41A system can be with a single memory feature. Miscellaneous equipment specification (MES) memory upgrades must pair the first memory feature and pair additional memory features.

Table 1-8 on page 12 lists the memory features that are available for the Power S814 and S824 system. DDR4 features (EM91, EM92, EM93, and EM94) operate at DDR3 speeds.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000027

Table 1-8 Summary of memory features for Power S814 and Power S824

Feature code	DIMM capacity	Access rate	Maximum quantity
EM8B	16 GB	1600 MHz	8
EM8Z	32 GB	1600 MHz	8
EM8D	64 GB	1600 MHz	8
EM8E	128 GB	1600 MHz	8
EM91	16 GB	1600 MHz	8
EM92	32 GB	1600 MHz	8
EM93	64 GB	1600 MHz	8
EM94	128 GB	1600 MHz	8

Note: If using the 4-core processor feature #EPXK, the only allowed memory feature code is #EM8B (16 GB CDIMMs) with a maximum quantity of four.

A minimum 32 GB of memory is required on the Power S824 server. Memory is required to be ordered in pairs. Base memory is two 16 GB, 1600 MHz DDR3 memory modules (#EM8B).

Plans for future memory upgrades should be taken into account when deciding which memory feature size to use at the time of initial system order.

1.4.7 PCIe slots

The following section describes the available PCIe slots:

- With one POWER8 processor DCM, the Power S814 has seven PCIe Gen3 slots. Two are x16 full-height and full-length slots. Five are x8 Gen3 full-height, half-length slots.
- The Power S824 has up to 11 PCIe hot-plug Gen3 slots, providing excellent configuration flexibility and expandability.

  With one POWER8 processor DCM, seven PCIe Gen3 slots are available. Two are x16 full-height and full-length slots. Five are x8 Gen3 full-height, half-length slots. Four other PCIe slots are physically present, but are not available for use until a second processor DCM is installed.

  With two POWER8 processor DCMs, 11 PCIe Gen3 slots are available. Four are x16 full-height and full-length slots. Seven are x8 Gen3 full-height, half-length slots.

The x16 slots can provide up to twice the bandwidth of x8 slots because they offer twice as many PCIe lanes. PCIe Gen3 slots can support up to twice the bandwidth of a PCIe Gen2 slot and up to four times the bandwidth of a PCIe Gen 1 slot, assuming an equivalent number of PCIe lanes.

Note:

- One of the x8 PCIe slots on the Power S814 and Power S824 servers is used for a PCIe2 4-port 1 Gb Ethernet Adapter (#5899).
- One fewer PCIe slot is available with the dual IOA storage backplane feature EJ0P.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000028

The new servers are smarter about energy efficiency for cooling a PCIe adapter environment. They sense which IBM PCIe adapters are installed in their PCIe slots, and if an adapter requires higher levels of cooling, they automatically speed up fans to increase airflow across the PCIe adapters.

If it is required to have more PCIe I/O slots in a systems, you can attach external I/O drawers to the Power S814 and Power S824.

Note: The PCIe expansion drawer is not supported on the 4-core Power S814.

For more information, see 1.6, “I/O drawers for Power S814 and Power S824 servers” on page 21.

1.5 Disk and media features

Three backplane options are available for the Power S814 and Power S824 servers and provide a great deal of flexibility and capability. One of the three options must be configured:

1. Storage backplane with 12 SFF-3 bays and one DVD bay (#EJ0N)
2. Storage backplane with 12 SFF-3 bays and one DVD bay (#EJ0N) and feature EJ0S provides split backplane functionality
3. Storage backplane with 18 SFF-3 bays, 1.8-inch SSD cage bays (42A only), one DVD bay, dual integrated SAS controllers with write cache and IBM Easy Tier® functionality (#EJ0P)

Note: When the 8286-41A is configured with the 4-core processor feature, #EPXK, a maximum of ten disk drives are allowed. This is true with either the EJ0N or the EJ0P backplane options.

Each of the three backplane options provides SFF-3 SAS bays in the system unit. These 2.5-inch or small form factor (SFF) SAS bays can contain SAS drives (HDD or SSD) mounted on a Gen3 tray or carrier. Thus the drives are designated SFF-3. SFF-1 or SFF-2 drives do not fit in an SFF-3 bay. All SFF-3 bays support concurrent maintenance or hot-plug capability.

In addition to supporting HDDs and SSDs in the SFF-3 SAS bays of the Power S824, the storage backplane feature EJ0P supports a mandatory 8-bay, 1.8-inch SSD Cage (#EJTM). All eight bays are accessed by both of the integrated SAS controllers. The bays support concurrent maintenance (hot-plug). The SSD 1.8-inch drive, such as the 387 GB capacity feature ES16 (AIX, Linux) or ES17 (IBM i), are supported.

The high performance SAS controllers provide RAID 0, RAID 5, RAID 6, and RAID 10 support. The dual SAS controllers can automatically move hot data to an attached SSD and cold data to an attached HDD for AIX and Linux environments by using the Easy Tier function.

Table 1-9 shows the available disk drive feature codes that can be installed in the Power S814.

Table 1-9 Disk drive feature code description for Power S814

Feature code	CCIN	Description	Max	OS support
ES0G		775 GB SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES0H		775 GB SFF-2 SSD for IBM i	336	IBM i

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000029

Feature code	CCIN	Description	Max	OS support
ES0L		387 GB SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES0M		387 GB SFF-3 SSD for IBM i	18	IBM i
ES0N		775 GB SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES0P		775 GB SFF-3 SSD for IBM i	18	IBM i
ES0Q		387 GB SFF-2 4 K SSD for AIX/Linux	336	AIX, Linux
ES0R		387 GB SFF-2 4 K SSD for IBM i	336	IBM i
ES0S		775 GB SFF-2 4 K SSD for AIX/Linux	336	AIX, Linux
ES0T		775 GB SFF-2 4 K SSD for IBM i	336	IBM i
ES0U		387 GB SFF-3 4 K SSD AIX/Linux	18	AIX, Linux
ES0V		387 GB SSF-3 4 K SSD for IBM i	18	IBM i
ES0W		775 GB SFF-3 4 K SSD for AIX/Linux	18	AIX, Linux
ES0X		775 GB SFF-3 4 K SSD for IBM i	18	IBM i
ES19		387 GB SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES1A		387 GB SFF-2 SSD for IBM i	336	IBM i
ES62		3.86-4.0 TB 7200 RPM 4K SAS LFF-1 Nearline Disk Drive (AIX/Linux)	336	AIX, Linux
ES64		7.72-8.0 TB 7200 RPM 4K SAS LFF-1 Nearline Disk Drive (AIX/Linux)	336	AIX, Linux
ES78		387 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	336	AIX, Linux
ES79		387 GB SFF-2 SSD 5xx eMLC4 for IBM i	336	IBM i
ES7E		775 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	336	AIX, Linux
ES7F		775 GB SFF-2 SSD 5xx eMLC4 for IBM i	336	IBM i
ES7K		387 GB SFF-3 SSD 5xx eMLC4 for AIX/Linux	18	AIX, Linux
ES7L		387 GB SFF-3 SSD 5xx eMLC4 for IBM i	18	IBM i
ES7P		775 GB SFF-3 SSD 5xx eMLC4 for AIX/Linux	18	AIX, Linux
ES7Q		775 GB SFF-3 SSD 5xx eMLC4 for IBM i	18	IBM i
ES80		1.9 TB Read Intensive SAS 4k SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES81		1.9 TB Read Intensive SAS 4k SFF-2 SSD for IBM i	336	IBM i
ES85		387 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux
ES86		387 GB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8C		775 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux
ES8D		775 GB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8F		1.55 TB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000030

Feature code	CCIN	Description	Max	OS support
ES8G		1.55 TB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8J		1.9 TB Read Intensive SAS 4k SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES8K		1.9 TB Read Intensive SAS 4k SFF-3 SSD for IBM i	18	IBM i
ES8N		387 GB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8P		387 GB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
ES8Q		775 GB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8R		775 GB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
ES8V		1.55 TB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8W		1.55 TB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
1948	19B1	283 GB 15 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
1953		300 GB 15 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
1962	19B3	571 GB 10 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
1964		600 GB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
ESD2	59CD	1.1 TB 10 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
ESD3		1.2 TB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
ESD4		571 GB 10 K RPM SAS SFF-3 Disk Drive (IBM i)	18	IBM i
ESD5		600 GB 10 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux
ESD8		1.1 TB 10 K RPM SAS SFF-3 DIsk Drive (IBM i)	18	IBM i
ESD9		1.2 TB 10 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux
ESDV		283 GB 15 K RPM SAS SFF-3 Disk Drive (IBM i)	18	IBM i
ESDB		300 GB 15 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux
ESDE		571 GB 15 K RPM SAS SFF-3 Disk Drive - 528 Block (IBM i)	18	IBM i
ESDF		600 GB 15 K RPM SAS SFF-3 Disk Drive - 5xx Block (AIX/Linux)	18	AIX, Linux
ESDN		571 GB 15 K RPM SAS SFF-2 Disk Drive - 528 Block (IBM i)	672	IBM i
ESDP		600 GB 15 K RPM SAS SFF-2 Disk Drive - 5xx Block (AIX/Linux)	672	AIX, Linux
ESEU	59D2	571 GB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESEV		600 GB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESEY		283 GB 15 K RPM SAS SFF-2 4 K Block - 4224 Disk Drive	672	IBM i

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000031

Feature code	CCIN	Description	Max	OS support
ESEZ		300 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	672	AIX, Linux
ESF2	58DA	1.1 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESF3		1.2 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESF4	58D3	571 GB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESF5		600 GB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux
ESF8	58DB	1.1 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESF9		1.2 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux
ESFG		283 GB 15 K RPM SAS SFF-3 4 K Block - 4224 Disk Drive	18	IBM i
ESFB		300 GB 15 K RPM SAS SFF-3 4 K Block - 4096 Disk Drive	18	AIX, Linux
ESFE		571 GB 15 K RPM SAS SFF-3 4 K Block - 4224 Disk Drive	18	IBM i
ESFF		600 GB 15 K RPM SAS SFF-3 4 K Block - 4096 Disk Drive	18	AIX, Linux
ESFN		571 GB 15 K RPM SAS SFF-2 4 K Block - 4224 Disk Drive	672	IBM i
ESFP		600 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	672	AIX, Linux
ESFS		1.7 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESFT		1.8 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESFU		1.7 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESFV		1.8 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux

Note: If using the 4-core processor feature #EPXK, the only available disk options are:

- #ESDT - 146 GB 15k rpm disk drive
- #ESDU - 139 GB 15k rpm disk drive
- #ESDB - 300 GB 15k rpm disk drive
- #ESDA - 283 GB 15k rpm disk drive
- #ES0L - 387 GB Solid State disk
- #ES0M - 387 GB Solid State disk

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000032

Table 1-10 shows the available disk drive feature codes that can be installed in the Power S824.

Table 1-10 Disk drive feature code description for Power S824

Feature code	CCIN	Description	Max	OS support
ES0G		775 GB SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES0H		775 GB SFF-2 SSD for IBM i	336	IBM i
ES0L		387 GB SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES0M		387 GB SFF-3 SSD for IBM i	18	IBM i
ES0N		775 GB SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES0P		775 GB SFF-3 SSD for IBM i	18	IBM i
ES0Q		387 GB SFF-2 4 K SSD for AIX/Linux	336	AIX, Linux
ES0R		387 GB SFF-2 4 K SSD for IBM i	336	IBM i
ES0S		775 GB SFF-2 4 K SSD for AIX/Linux	336	AIX, Linux
ES0T		775 GB SFF-2 4 K SSD for IBM i	336	IBM i
ES0U		387 GB SFF-3 4 K SSD AIX/Linux	18	AIX, Linux
ES0V		387 GB SSF-3 4 K SSD for IBM i	18	IBM i
ES0W		775 GB SFF-3 4 K SSD for AIX/Linux	18	AIX, Linux
ES0X		775 GB SFF-3 4 K SSD for IBM i	18	IBM i
ES0Y		177 GB 1.8" 4 K SSD Read Intensive for AIX/Linux	8	AIX, Linux
ES0Z		177 GB 1.8" SSD Read Intensive for AIX/Linux (528 byte)	8	AIX, Linux
ES16		387 GB 1.8" SSD for AIX/Linux	8	AIX, Linux
ES17		387 GB 1.8" SSD for IBM i	8	IBM i
ES1C	**5B32**	**387 GB 1.8" SAS 5xx SSD eMLC4 for AIX/Linux**	**8**	AIX, Linux
ES1D		387 GB 1.8" SAS 5xx SSD eMLC4 for IBM i	8	IBM i
ES2V	5B30	387 GB 1.8" SAS 4k SSD eMLC4 for AIX/Linux	18	AIX, Linux
ES2W		387 GB 1.8" SAS 4k SSD eMLC4 for IBM i	8	IBM i
ES2X	5B33	775 GB 1.8" SAS 5xx SSD eMLC4 for AIX/Linux	8	AIX, Linux
ES2Y		775 GB 1.8" SAS 5xx SSD eMLC4 for IBM i	8	IBM i
ES4K	5B31	775 GB 1.8" SAS 4k SSD eMLC4 for AIX/Linux	8	AIX, Linux
ES4L		775 GB 1.8" SAS 4k SSD eMLC4 for IBM i	8	IBM i
ES19		387 GB SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES1A		387 GB SFF-2 SSD for IBM i	336	IBM i
ES78		387 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	336	AIX, Linux
ES79		387 GB SFF-2 SSD 5xx eMLC4 for IBM i	336	IBM i

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000033

Feature code	CCIN	Description	Max	OS support
ES7E		775 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	336	AIX, Linux
ES7F		775 GB SFF-2 SSD 5xx eMLC4 for IBM i	336	IBM i
ES7K		387 GB SFF-3 SSD 5xx eMLC4 for AIX/Linux	18	AIX, Linux
ES7L		387 GB SFF-3 SSD 5xx eMLC4 for IBM i	18	IBM i
ES7P		775 GB SFF-3 SSD 5xx eMLC4 for AIX/Linux	18	AIX, Linux
ES7Q		775 GB SFF-3 SSD 5xx eMLC4 for IBM i	18	IBM i
ES80		1.9 TB Read Intensive SAS 4k SFF-2 SSD for AIX/Linux	336	AIX, Linux
ES81		1.9 TB Read Intensive SAS 4k SFF-2 SSD for IBM i	336	IBM i
ES85		387 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux
ES86		387 GB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8C		775 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux
ES8D		775 GB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8F		1.55 TB SFF-2 SSD 4k eMLC4 for AIX/Linux	336	AIX, Linux
ES8G		1.55 TB SFF-2 SSD 4k eMLC4 for IBM i	336	IBM i
ES8J		1.9 TB Read Intensive SAS 4k SFF-3 SSD for AIX/Linux	18	AIX, Linux
ES8K		1.9 TB Read Intensive SAS 4k SFF-3 SSD for IBM i	18	IBM i
ES8N		387 GB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8P		387 GB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
ES8Q		775 GB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8R		775 GB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
ES8V		1.55 TB SFF-3 SSD 4k eMLC4 for AIX/Linux	18	AIX, Linux
ES8W		1.55 TB SFF-3 SSD 4k eMLC4 for IBM i	18	IBM i
1738	19B4	856 GB 10 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
1752		900 GB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
1948	19B1	283 GB 15 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
1953		300 GB 15 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
1962	19B3	571 GB 10 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
1964		600 GB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
ESD2	59CD	1.1 TB 10 K RPM SAS SFF-2 Disk Drive (IBM i)	672	IBM i
ESD3		1.2 TB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	672	AIX, Linux
ESD4		571 GB 10 K RPM SAS SFF-3 Disk Drive (IBM i)	18	IBM i
ESD5		600 GB 10 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000034

Feature code	CCIN	Description	Max	OS support
ESD8		1.1 TB 10 K RPM SAS SFF-3 DIsk Drive (IBM i)	18	IBM i
ESD9		1.2 TB 10 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux
ESDA		283 GB 15 K RPM SAS SFF-3 Disk Drive (IBM i)	18	IBM i
ESDB		300 GB 15 K RPM SAS SFF-3 Disk Drive (AIX/Linux)	18	AIX, Linux
ESDE		571 GB 15 K RPM SAS SFF-3 Disk Drive - 528 Block (IBM i)	18	IBM i
ESDF		600 GB 15 K RPM SAS SFF-3 Disk Drive - 5xx Block (AIX/Linux)	18	AIX, Linux
ESDN		571 GB 15 K RPM SAS SFF-2 Disk Drive - 528 Block (IBM i)	672	IBM i
ESDP		600 GB 15 K RPM SAS SFF-2 Disk Drive - 5xx Block (AIX/Linux)	672	AIX, Linux
ESEU	59D2	571 GB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESEV		600 GB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESEY		283 GB 15 K RPM SAS SFF-2 4 K Block - 4224 Disk Drive	672	IBM i
ESEZ		300 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	672	AIX, Linux
ESF2	58DA	1.1 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESF3		1.2 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESF4	58D3	571 GB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESF5		600 GB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux
ESF8	58DB	1.1 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESF9		1.2 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux
ESFA		283 GB 15 K RPM SAS SFF-3 4 K Block - 4224 Disk Drive	18	IBM i
ESFB		300 GB 15 K RPM SAS SFF-3 4 K Block - 4096 Disk Drive	18	AIX, Linux
ESFE		571 GB 15 K RPM SAS SFF-3 4 K Block - 4224 Disk Drive	18	IBM i
ESFF		600 GB 15 K RPM SAS SFF-3 4 K Block - 4096 Disk Drive	18	AIX, Linux

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000035

Feature code	CCIN	Description	Max	OS support
ESFN		571 GB 15 K RPM SAS SFF-2 4 K Block - 4224 Disk Drive	672	IBM i
ESFP		600 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	672	AIX, Linux
ESFS		1.7 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4224	672	IBM i
ESFT		1.8 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	672	AIX, Linux
ESFU		1.7 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4224	18	IBM i
ESFV		1.8 TB 10 K RPM SAS SFF-3 Disk Drive 4 K Block - 4096	18	AIX, Linux

Included in the feature EJ0N or EJ0P backplanes is a slimline media bay that can optionally house a SATA DVD-RAM (#5771). The DVD drive is run by the integrated SAS controllers, and a separate PCIe adapter is not required.

Note: When IBM i is the primary operating system (#2145), a DVD device does need to be ordered. A DVD device is needed for normal system operations, but not required on all systems. Clients with write-sensitive disk/HDD workloads should upgrade from the #EJ0N backplane to the #EJ0P backplane to gain the performance advantage of write cache.

The Power S814 and the Power S824 support the RDX USB External Docking Station for Removable Disk Cartridge (#EUA4). The USB External Docking Station accommodates RDX removable disk cartridge of any capacity. The disks are in a protective rugged cartridge enclosure that plugs into the docking station. The docking station holds one removable rugged disk drive/cartridge at a time. The rugged removable disk cartridge and docking station backs up similar to a tape drive. This can be an excellent alternative to DAT72, DAT160, 8mm, and VXA-2 and VXA-320 tapes.

The Power S814 tower additionally supports a top mounted RDX docking station. The enclosure is firmly mounted on top of a Power S814 tower configuration. It is not available for a rack-mounted configuration. The enclosure adds about 5 cm (2 in) height to the tower. The RDX docking station can be removed from the enclosure using concurrent maintenance.

Table 1-11 shows the available media device feature codes for Power S814 and Power S824.

Table 1-11 Media device feature code description for Power S814 and Power S824

Feature code	Description
5771	SATA Slimline DVD-RAM Drive
EUA3	RDX USB Top Mount Docking Station for Removable Cartridge for S814 Tower Model
EUA4	RDX USB External Docking Station for Removable Disk Cartridge

SCSI disks are not supported in the Power S814 and Power S824 disk bays. Also, because no PCIe LP SCSI adapter is available, you cannot attach existing SCSI disk subsystems.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000036

If you need more disks than are available with the internal disk bays, you can attach additional external disk subsystems. For more information about the available external disk subsystems, see 2.7, “External I/O subsystems” on page 82.

For more information about the internal disk features, see 2.6, “Internal storage” on page 72.

1.6 I/O drawers for Power S814 and Power S824 servers

If additional Gen3 PCIe slots beyond the system node slots are required, PCIe Gen3 I/O drawers can be attached to the Power S814 and Power S824 servers.

Disk-only I/O drawers (#5887) are also supported, providing storage capacity.

Similarly, the GX++ attached EXP30 Ultra SSD Drawer (#EDR1 or #5888) is not supported. Also, the 3.5-inch-based feature 5886 EXP12S SAS Disk Drawer and feature 5786 EXP24 SCSI Disk Drawer are not supported.

IBM offers the IBM System Storage® 7226 Model 1U3 Multi-Media Enclosure that can hold one or more DVDs, tape drive, or RDX docking stations. For more information about the multimedia drawer, see “IBM System Storage 7226 Model 1U3 Multi-Media Enclosure” on page 35.

1.6.1 PCIe Gen3 I/O expansion drawer

The 19-inch 4 EIA (4U) PCIe Gen3 I/O expansion drawer (#EMX0) and two PCIe Fan Out Modules (#EMXF) provide twelve PCIe I/O full-length, full-height slots. One Fan Out Module provides six PCIe slots labeled C1 through C6. C1 and C4 are x16 slots and C2, C3, C5, and C6 are x8 slots. PCIe Gen1, Gen2, and Gen3 full-high adapter cards are supported.

A blind swap cassette (BSC) is used to house the full-high adapters that go into these slots. The BSC is the same BSC as used with the previous generation server's 12X attached I/O drawers (#5802, #5803, #5877, #5873). The drawer is shipped with a full set of BSC, even if the BSC is empty.

Concurrent repair and add/removal of PCIe adapter cards is done by HMC guided menus or by operating system support utilities.

A PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer (#EJ08) and 3.0 m (#ECC7) or 10.0 m (#ECC8) CXP 16X Active Optical cables (AOC) connect the system node to a PCIe Fan Out module in the I/O expansion drawer. One feature #ECC7 or one #ECC8 ships two AOC cables. Each PCIe Gen3 I/O expansion drawer has two power supplies.

A maximum of one PCIe Gen3 I/O expansion drawer is supported on the one socket Power S814 and Power S824 systems.

Note: The expansion drawer is not supported on 4-core Power S814 servers.

A maximum of two PCIe Gen3 I/O expansion drawer is supported on the two sockets Power S824 system.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000037

Figure 1-6 shows a PCIe Gen3 I/O expansion drawer.

Figure 1-6 PCIe Gen3 I/O expansion drawer

1.6.2 I/O drawers and usable PCI slot

Figure 1-7 shows the rear view of the PCIe Gen3 I/O expansion drawer equipped with two PCIe3 6-slot Fan-Out modules with the location codes for the PCIe adapter slots.

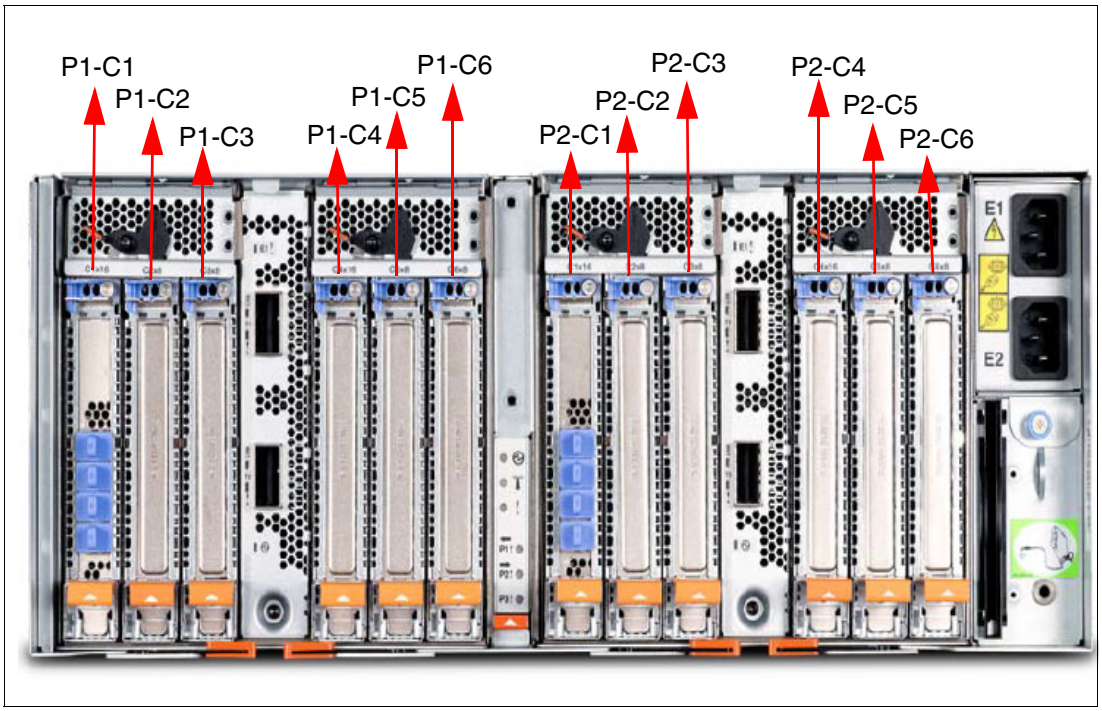

Figure 1-7 Rear view of a PCIe Gen3 I/O expansion drawer with PCIe slots location codes

Table 1-12 provides details of the PCI slots in the PCIe Gen3 I/O expansion drawer equipped with two PCIe3 6-slot Fan-Out modules.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000038

Table 1-12 PCIe slot locations for the PCIe Gen3 I/O expansion drawer with two Fan Out modules

Slot	Location code	Description
Slot 1	P1-C1	PCIe3, x16
Slot 2	P1-C2	PCIe3, x8
Slot 3	P1-C3	PCIe3, x8
Slot 4	P1-C4	PCIe3, x16
Slot 5	P1-C5	PCIe3, x8
Slot 6	P1-C6	PCIe3, x8
Slot 7	P2-C1	PCIe3, x16
Slot 8	P2-C2	PCIe3, x8
Slot 9	P2-C3	PCIe3, x8
Slot 10	P2-C4	PCIe3, x16
Slot 11	P2-C5	PCIe3, x8
Slot 12	P2-C6	PCIe3, x8

- All slots support full-length, regular-height adapter or short (low-profile) with a regular-height tailstock in single-wide, Gen3, blind-swap cassettes.
- Slots C1 and C4 in each PCIe3 6-slot Fan Out module are x16 PCIe3 buses and slots C2, C3, C5, and C6 are x8 PCIe buses.
- All slots support enhanced error handling (EEH).
- All PCIe slots are hot swappable and support concurrent maintenance.

Table 1-13 summarizes the maximum number of I/O drawers supported and the total number of PCI slots that are available.

Table 1-13 Maximum number of I/O drawers supported and total number of PCI slots

System	Maximum #EMX0 drawers	Total number of slots	
		PCIe3, x16	PCIe3, x8
Power S814 (1-socket)	1	4	8
Power S824 (1-socket)	1	4	8
Power S824 (2-sockets)	2	8	16

1.6.3 EXP24S SFF Gen2-bay drawer

If you need more disks than are available with the internal disk bays, you can attach additional external disk subsystems, such as the EXP24S SAS HDD/SSD Expansion Drawer (#5887). The EXP24S SFF Gen2-bay drawer is an expansion drawer supporting up to twenty-four 2.5-inch hot-swap SFF SAS HDDs on IBM POWER6®, IBM POWER6+™, IBM POWER7®, IBM POWER7+™, or POWER8 servers in 2U of 19-inch rack space. The EXP24S bays are controlled by SAS adapters or controllers that are attached to the I/O drawer by SAS X or Y cables.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000039

The EXP24S drawer is attached to SAS ports on either a PCIe SAS adapter in the server or to the SAS ports on the rear of the server. Two SAS ports on the rear of the server are enabled with the expanded-function storage backplane with dual IOA support (#EJ0P).

A maximum of 14 EXP24S drawers are supported on the Power S814 and Power S824.

The SFF bays of the EXP24S differ from the SFF bays of the POWER8 system units. The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the SFF-3 bays of the POWER8 system unit.

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

Figure 1-8 shows the EXP24S SFF drawer.

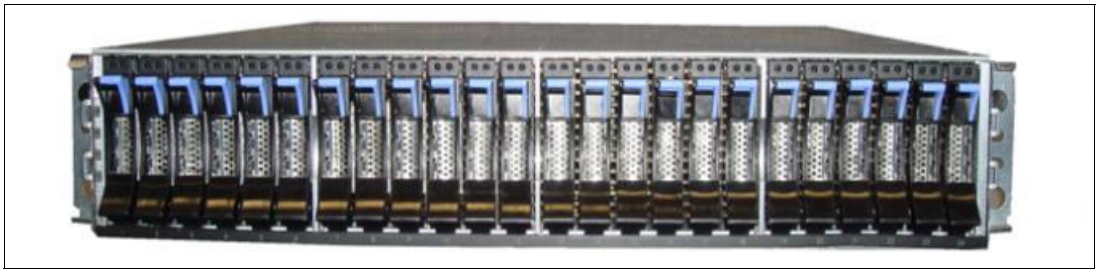

Figure 1-8 EXP24S SFF drawer

1.7 IBM i Solution Edition for Power S814 and Power S824

The IBM i Solution Editions are designed to help you take advantage of the combined experience and expertise of IBM and 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 S814 Solution Edition 4-code (#4928) and 6-core and 8-core ($4927) configurations.

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

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

The Power S814 Solution Edition and Power S824 Solution Edition include 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.

Here are the requirements to be eligible to purchase a Power S814 Solution Edition or an Power S824 Solution Edition order:

- The offering must include new or upgrade software licenses or software maintenance from the ISV for the qualifying IBM server. Services or 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.
- A Solution Edition is not eligible for registration as a capacity backup (CBU).

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000040

1.8 Capacity backup

The following describes the different CBU features for the Power S814 and Power S824.

1.8.1 Power System S814 (8286-41A) capacity backup offering

The Power S814 (8286-41A) CBU designation enables you to temporarily transfer IBM i processor license entitlements and IBM i user license entitlements that are purchased for a primary machine to a secondary CBU-designated system for high availability disaster recovery (HADR) operations. Temporarily transferring these resources instead of purchasing them for your secondary system might result in significant savings. Processor activations cannot be transferred.

The CBU specify feature 0444 is available only as part of a new server purchase. Certain system prerequisites must be met and system registration and approval are required before the CBU specify feature can be applied on a new server. Standard IBM i terms and conditions do not allow either IBM i processor license entitlements or IBM i user license entitlements to be transferred permanently or temporarily. These entitlements remain with the machine for which they were ordered. When you register the association between your primary and on-order CBU system, you must agree to certain terms and conditions regarding the temporary transfer.

After a new CBU system is registered along with the proposed primary and the configuration is approved, you can temporarily move your optional IBM i processor license entitlement and IBM i user license entitlements from the primary system to the CBU system when the primary system is down or while the primary system processors are inactive. The CBU system can then support failover and role swapping for a full range of test, disaster recovery, and high availability scenarios. Temporary entitlement transfer means that the entitlement is a property that is transferred from the primary system to the CBU system and may remain in use on the CBU system while the registered primary and CBU systems are in deployment for the high availability or disaster recovery operation.

The intent of the CBU offering is to enable regular role-swap operations.

The primary systems for a single socket Power S814 (8286-41A) server with its IBM i P05 (with the 4-core processor feature #EPXK) and P10 software tier can be a POWER7, POWER7+, or POWER8 server with a P10, P20, or P30 software tier, which is listed below:

- Power S824: 8286-42A
- Power S814: 8286-41A
- Power 770: 9117-MMC
- Power 770: 9117-MMD
- Power 760: 9109-RMD
- Power 750: 8408-E8D
- Power 740: 8205-E6C
- Power 740: 8205-E6D
- Power 730: 8231-E2D
- Power 720: 8202-E4C with features #EPC6, #EPC7
- Power 720: 8202-E4D with features #EPCL, #EPCM
- Power 710: 8231-E1D with features #EPCG, #EPCJ
- Power 710: 8231-E1D with features #EPCJ, #EPCG
- P460: 7895-42X
- P260: 7895-22X
- PS704: 7891-74X
- PS703: 7891-73X

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000041

The primary machine must be in the same enterprise as the CBU system. The solution editions are not eligible for CBU status.

Before you can temporarily transfer IBM i processor license entitlements from the registered primary system, you must have more than one IBM i processor license on the primary machine and at least one IBM i processor license on the CBU server. An activated processor must be available on the CBU server to use the transferred entitlement. You can then transfer any IBM i processor entitlements above the minimum one, assuming the total IBM i workload on the primary system does not require the IBM i entitlement that you want to transfer during the time of the transfer. During this temporary transfer, the CBU system's internal records of its total number of IBM i processor license entitlements are not updated, and you might see IBM i license noncompliance warning messages from the CBU system. These warning messages in this situation do not mean that you are not in compliance.

Before you can temporarily transfer IBM i user entitlements, you must have more than the minimum number of IBM i user entitlements on the primary server. You can then transfer any IBM i user entitlements above the minimum, assuming the total IBM i users on the primary system do not require the IBM i entitlement that you want to transfer during the time of the transfer. The servers with P20 or higher software tiers do not have user entitlements that can be transferred, and only processor license entitlements can be transferred. The minimum number of IBM i users on the POWER7, POWER7+, and POWER8 with i user entitlements are as follows:

- Power 720 4-core (8202-E4B, 8202-E4C, 8202-E4D): Five users
- Power 720 6-core or 8-core (8202-E4B, 8202-E4C, 8202-E4D): 30 users
- Power S814 (8284-41A): 30 users

For example, if you have a 4-core Power 720 (8202-E4D) as your primary system with two IBM i processor license entitlements (one above the minimum) and 50 IBM i user entitlements (20 above the minimum), you can temporarily transfer up to one IBM i entitlement and up to 20 user entitlements. During this temporary transfer, the CBU system's internal records of its total number of IBM i processor and user license entitlements is not updated, and you might see IBM i license noncompliance warning messages from the CBU system.

Note: If an IBM Power 720 or Power S814 server has just one partition and if it is using the default parameters that are set by IBM Manufacturing, the IBM i licensing manager ignores the temporary transfer and does not use additional processor cores. To work around this restriction for valid CBU situations, you can add a partition with no resource on your server or implement a shared processor pool.

If your primary or CBU machine is sold or discontinued from use, any temporary entitlement transfers must be returned to the machine on which they were originally acquired. For CBU registration and further information, visit the following website:

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

1.8.2 Power System S824 (8286-42A) capacity backup capability

The Power S824 (8286-42A) CBU designation enables you to temporarily transfer IBM i processor license entitlements and IBM i user license entitlements that are purchased for a primary machine to a secondary CBU-designated system for HADR operations. Temporarily transferring these resources instead of purchasing them for your secondary system might result in significant savings. Processor activations cannot be transferred.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000042

The CBU specify feature 0444 is available only as part of a new server purchase. Certain system prerequisites must be met and system registration and approval are required before the CBU specify feature can be applied on a new server. Standard IBM i terms and conditions do not allow either IBM i processor license entitlements or IBM i user license entitlements to be transferred permanently or temporarily. These entitlements remain with the machine for which they were ordered. When you register the association between your primary and on-order CBU system, you must agree to certain terms and conditions regarding the temporary transfer.

The CBU status is granted only after the primary and CBU systems are registered and the pairing is approved. After a CBU system designation is approved and the system is installed, you can temporarily move your optional IBM i processor license entitlements and 5250 Enterprise Enablement entitlements from the primary system to the CBU system when the primary system is down or while the primary system processors are inactive. The CBU system can then support failover and role swapping for a full range of test, disaster recovery, and high availability scenarios. Temporary entitlement transfer means that the entitlement is a property that is transferred from the primary system to the CBU system and may remain in use on the CBU system while the registered primary and CBU systems are in deployment for the high availability or disaster recovery operation.

The primary systems for a Power S824 (8286-42A) server with its P20 software tier can be as follows:

- Power S824: 8286-42A
- Power 730: 8231-E2D
- Power 740: 8205-E6C, 8205-E6D
- Power 750: 8408-E8D
- Power 760: 9109-RMD
- Power 770: 9117-MMC, 9117-MMD

These systems have IBM i software licenses with an IBM i P20 or P30 software tier. The primary machine must be in the same enterprise as the CBU system. The Solution Editions are not eligible for CBU status.

Before you can temporarily transfer IBM i processor license entitlements from the registered primary system, you must have more than one IBM i processor license on the primary machine and at least one IBM i processor license on the CBU server. An activated processor must be available on the CBU server to use the transferred entitlement. You can then transfer any IBM i processor entitlements above the minimum one, assuming that the total IBM i workload on the primary system does not require the IBM i entitlement that you want to transfer during the time of the transfer. During this temporary transfer, the CBU system's internal records of its total number of IBM i processor license entitlements are not updated, and you might see IBM i license noncompliance warning messages from the CBU system. These warning messages in this situation do not mean that you are not in compliance.

Before you can temporarily transfer 5250 entitlements, you must have more than one 5250 Enterprise Enablement entitlement on the primary server and at least one 5250 Enterprise Enablement entitlement on the CBU system. You can then transfer the entitlements that are not required on the primary server during the time of transfer and that are above the minimum of one entitlement.

For example, if you have a 12-core Power 740 as your primary system with six IBM i processor license entitlements (five above the minimum) and two 5250 Enterprise Enablement entitlements (one above the minimum), you can temporarily transfer up to five IBM i entitlements and one 5250 Enterprise Enablement entitlement. During the temporary transfer, the CBU system's internal records of its total number of IBM i processor entitlements

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000043

are not updated, and you might see IBM i license noncompliance warning messages from the CBU system.

> **Note:** If a Power 740 or S824 CBU server has just one partition and if it is using the default parameters that are set by IBM Manufacturing, the IBM i licensing manager ignores the temporary transfer and does not use additional processor cores. To work around this restriction for valid CBU situations, you can add a partition with no resource on your server or implement a shared processor pool.

If your primary or CBU machine is sold or discontinued from use, any temporary entitlement transfers must be returned to the machine on which they were originally acquired. For CBU registration and further information, visit the following website:

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

1.9 Server and virtualization management

If you want to implement logical partitions, a Hardware Management Console (HMC) or the Integrated Virtualization Manager (IVM) is required to manage the Power S814 and Power S824 servers. In general, multiple IBM POWER6, POWER6+, POWER7, POWER7+, and POWER8 processor-based servers can be supported by a single HMC.

> **Remember:** If you do not use an HMC or IVM, the Power S814 and Power S824 runs in full system partition mode, which means 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 S814 and Power S824, the HMC must be a rack-mount CR5 or later, or deskside C08 or later.

In April 2015, IBM announced a new HMC model, machine type 7042-CR9 Hardware features on the CR9 model include a second disk drive (#1998) for RAID 1 data mirroring, and the option of a redundant power supply. If you prefer not to have RAID 1 enabled on the HMC, you can override it in the ordering system and remove the additional HDD from the order. RAID 1 is also offered on the 7042-CR6, 7042-CR7, 7042=CR8, and 7042-CR9 models as a miscellaneous equipment specification (MES) upgrade option.

Starting with HMC V8R8.1.0 code the HMC can manage more LPARs per processor core. A core can be partitioned in up to 20 LPARs (0.05 of a core).

Several HMC models are supported to manage POWER8 processor-based systems. The 7042-CR9 is the only HMCs that are available for ordering at the time of writing, but you can also use one of the withdrawn models that are listed in Table 1-14.

Table 1-14 HMC models that support POWER8 processor technology-based servers

Type-model	Availability	Description
7042-C08	Withdrawn	IBM 7042 Model C08 Deskside Hardware Management Console
7042-CR5	Withdrawn	IBM 7042 Model CR5 Rack-Mounted Hardware Management Console
7042-CR6	Withdrawn	IBM 7042 Model CR6 Rack mounted Hardware Management Console
7042-CR7	Withdrawn	IBM 7042 Model CR7 Rack mounted Hardware Management Console

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000044

Type-model	Availability	Description
7042-CR8	Withdrawn	IBM 7042 Model CR8 Rack mounted Hardware Management Console
7042-CR9	Available	IBM 7042 Model CR9 Rack mounted Hardware Management Console

At the time of writing the IBM POWER8 processor-based Power S822 server requires HMC V8R8.3.0.

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

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

If you are attaching an HMC to a new server or adding a function to an existing server that requires a firmware update, the HMC machine code might need to be updated because HMC code must always be equal to or higher than the managed server's firmware. Access to firmware and machine code updates is conditioned on entitlement and license validation in accordance with IBM policy and practice. IBM may verify entitlement through customer number, serial number, electronic restrictions, or any other means or methods that are employed by IBM at its discretion.

1.10 System racks

The Power S814 and Power S824 systems are designed to mount in the 36U 7014-T00 (#0551), the 42U 7014-T42 (#0553), or the IBM 42U Slim Rack (7965-94Y) rack. These racks are built to the 19-inch EIA 310D standard.

Order information: The racking approach for the initial order must be either a 7014-T00, 7014-T42, or 7965-94Y. If an additional rack is required for I/O expansion drawers as an MES to an existing system, either a feature 0551, 0553, or ER05 rack must be ordered.

You must leave 2U of space at either the bottom or top of the rack, depending on the client's cabling preferences, to allow for cabling to exit the rack

If a system is to be installed in a rack or cabinet that is not IBM, ensure that the rack meets the requirements that are described in 1.10.10, “OEM rack” on page 38.

Responsibility: The client is responsible for ensuring 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.

1.10.1 IBM 7014 Model T00 rack

The 1.8-meter (71-inch) model T00 is compatible with past and present IBM Power Systems servers. The features of the T00 rack are as follows:

- Has 36U (EIA units) of usable space.
- Has optional removable side panels.
- Has optional side-to-side mounting hardware for joining multiple racks.
- Has increased power distribution and weight capacity.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000045

- Supports both AC and DC configurations.
- Up to four power distribution units (PDUs) can be mounted in the PDU bays (see Figure 1-10 on page 33), but others can fit inside the rack. For more information, see 1.10.7, “The AC power distribution unit and rack content” on page 33.
- For the T00 rack, three door options are available:
  - Front Door for 1.8 m Rack (#6068).

    This feature provides an attractive black full height rack door. The door is steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack.
  - A 1.8 m Rack Acoustic Door (#6248).

    This feature provides a front and rear rack door that is designed to reduce acoustic sound levels in a general business environment.
  - A 1.8 m Rack Trim Kit (#6263).

    If no front door is used in the rack, this feature provides a decorative trim kit for the front.
- Ruggedized Rack Feature

  For enhanced rigidity and stability of the rack, the optional Ruggedized Rack Feature (#6080) provides additional hardware that reinforces the rack and anchors it to the floor. This hardware is designed primarily for use in locations where earthquakes are a concern. The feature includes a large steel brace or truss that bolts into the rear of the rack.

  It is hinged on the left side so it can swing out of the way for easy access to the rack drawers when necessary. The Ruggedized Rack Feature also includes hardware for bolting the rack to a concrete floor or a similar surface, and bolt-in steel filler panels for any unoccupied spaces in the rack.
- Weights are as follows:
  - T00 base empty rack: 244 kg (535 lb).
  - T00 full rack: 816 kg (1795 lb).
  - Maximum weight of drawers is 572 kg (1260 lb).
  - Maximum weight of drawers in a zone 4 earthquake environment is 490 kg (1080 lb). This number equates to 13.6 kg (30 lb) per EIA.

Important: If additional weight is added to the top of the rack, for example, adding #6117, the 490 kg (1080 lb) must be reduced by the weight of the addition. As an example, #6117 weighs approximately 45 kg (100 lb) so the new maximum weight of drawers that the rack can support in a zone 4 earthquake environment is 445 kg (980 lb). In the zone 4 earthquake environment, the rack must be configured starting with the heavier drawers at the bottom of the rack.

1.10.2 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 are for the model T42 rack (which differ from the model T00):

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000046

The available door options for the Model T42 rack are shown in Figure 1-9.

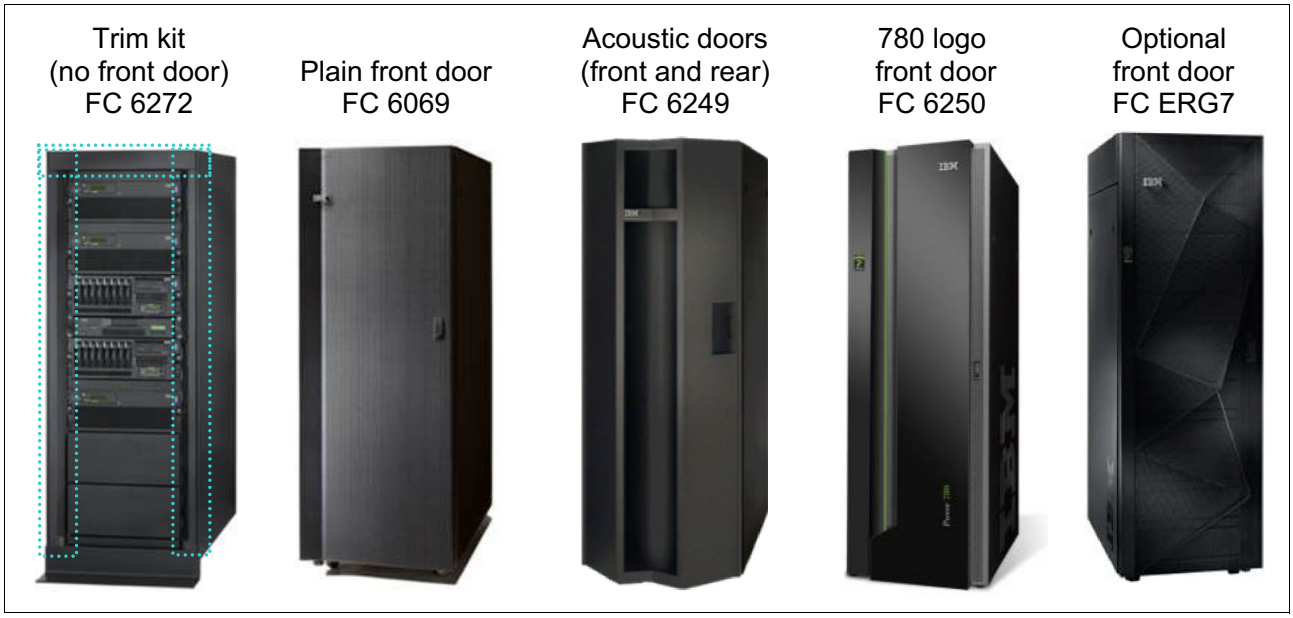

Figure 1-9 Door options for the T42 rack

- The 2.0 m Rack Trim Kit (#6272) is used, if no front door is used in the rack.
- The Front Door for a 2.0 m Rack (#6069) is made of steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack. This door is non-acoustic and has a depth of about 25 mm (1 in.).
- The 2.0 m Rack Acoustic Door (#6249) consists of a front and rear door to reduce noise by approximately 6 dB(A). It has a depth of approximately 191 mm (7.5 in.).
- The High-End Appearance Front Door (#6250) provides a front rack door with a field-installed Power 780 logo indicating that the rack contains a Power 780 system. The door is not acoustic and has a depth of about 90 mm (3.5 in.).

> **High end:** For the High-End Appearance Front Door (#6250), use the High-End Appearance Side Covers (#6238) to make the rack appear as though it is a high-end server (but in a 19-inch rack format instead of a 24-inch rack).

- The #ERG7 provides an attractive black full height rack door. The door is steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack. The non-acoustic door has a depth of about 134 mm (5.3 in.).

Rear Door Heat Exchanger

To lead away more heat, a special door that is named the Rear Door Heat Exchanger (#6858) is available. This door replaces the standard rear door on the rack. Copper tubes that are attached to the rear door circulate chilled water, provided by the customer. The chilled water removes heat from the exhaust air being blown through the servers and attachments that are mounted in the rack. With industry standard quick couplings, the water lines in the door attach to the customer-supplied secondary water loop.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000047

For details about planning for the installation of the IBM Rear Door Heat Exchanger, see the following website:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphad_p5/iphadexchangeroverview.html

1.10.3 IBM 42U Slim Rack 7965-94Y

The 2.0-meter (79-inch) model 7965-94Y is compatible with past and present IBM Power Systems servers and provides an excellent 19-inch rack enclosure for your data center. Its 600 mm (23.6 in.) width combined with its 1100 mm (43.3 in.) depth plus its 42 EIA enclosure capacity provides great footprint efficiency for your systems and allows it to be easily located on standard 24-inch floor tiles.

The IBM 42U Slim Rack has a lockable perforated front steel door, providing ventilation, physical security, and visibility of indicator lights in the installed equipment within. In the rear, either a lockable perforated rear steel door (#EC02) or a lockable Rear Door Heat Exchanger (RDHX)(1164-95X) is used. Lockable optional side panels (#EC03) increase the rack's aesthetics, help control airflow through the rack, and provide physical security. Multiple 42U Slim Racks can be bolted together to create a rack suite (indicate feature code #EC04).

Up to six optional 1U PDUs can be placed vertically in the sides of the rack. Additional PDUs can be located horizontally, but they each use 1U of space in this position.

1.10.4 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 that is delivered when you order the 7014-T00 rack. The included features might vary. Certain features that are delivered as part of the 7014-T00 must be ordered separately with the #0551.

1.10.5 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 that is delivered when you order the 7014-T42 rack. The included features might vary. Certain features that are delivered as part of the 7014-T42 must be ordered separately with the #0553.

1.10.6 Feature code ER05 rack

This feature provides a 19-inch, 2.0-meter high rack with 42 EIA units of total space for installing rack-mounted Central Electronics Complex or expansion units. The 600 mm wide rack fits within a data center's 24-inch floor tiles and provides better thermal and cable management capabilities. The following features are required on the #ER05:

- #EC01 front door
- #EC02 rear door or #EC05 Rear Door Heat Exchanger (RDHX) indicator

PDUs on the rack are optional. Each #7196 and #7189 PDU consumes one of six vertical mounting bays. Each PDU beyond four consumes 1U of rack space.

If ordering Power Systems equipment in an MES order, use the equivalent rack feature ER05 instead of 7965-94Y so IBM Manufacturing can ship the hardware in the rack.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000048

1.10.7 The AC power distribution unit and rack content

For rack models T00 and T42, 12-outlet PDUs are available. These include the AC power distribution units #9188 and #7188 and the AC Intelligent PDU+ #5889 and #7109. The Intelligent PDU+ (#5889 and #7109) is 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 PDUs have 12 client-usable IEC 320-C13 outlets. There are six groups of two outlets that are 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.

Four PDUs can be mounted vertically in the back of the T00 and T42 racks. Figure 1-10 shows 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 are 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, the best approach is to use fillers in the EIA units that are occupied by these PDUs to facilitate proper air-flow and ventilation in the rack.

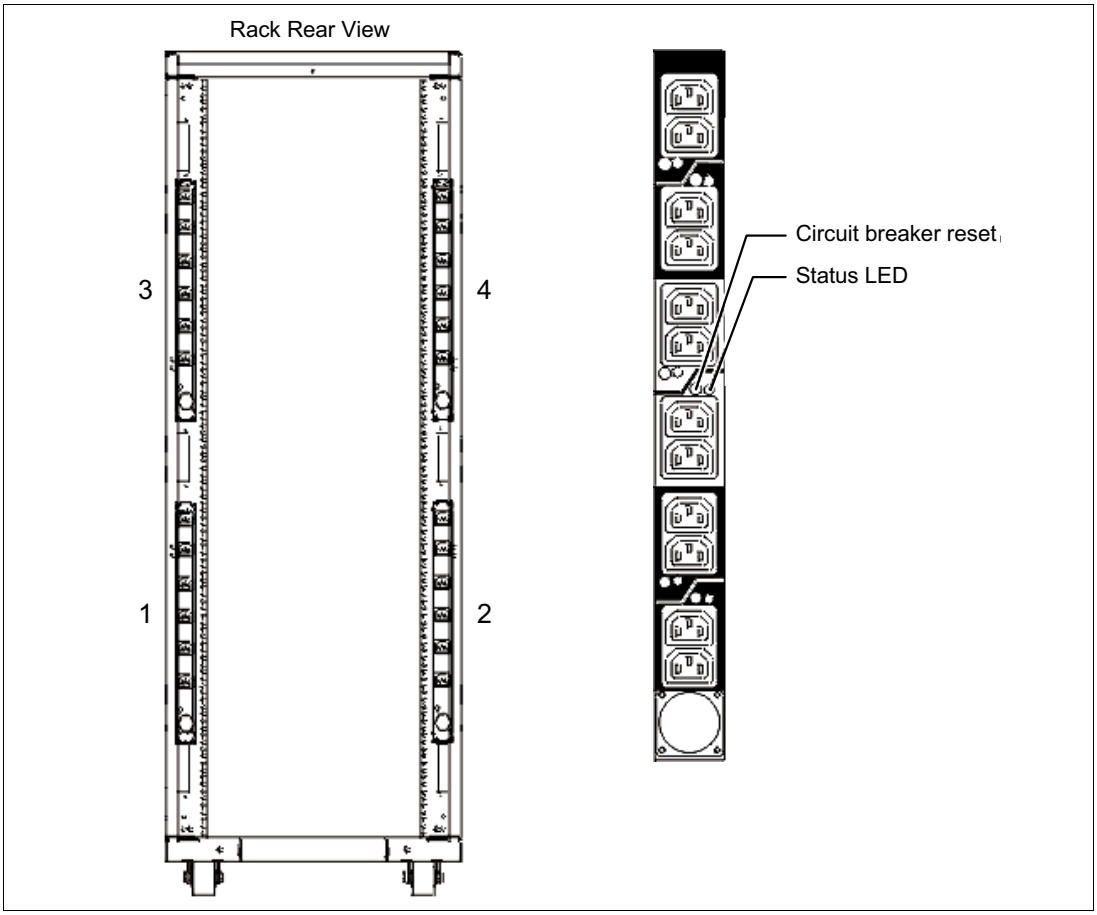

Figure 1-10 PDU placement and PDU view

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000049

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.

Table 1-15 shows the available wall power cord options for the PDU and iPDU features, which must be ordered separately.

Table 1-15 Wall power cord options for the PDU and iPDU features

Feature code	Wall plug	Rated voltage (Vac)	Phase	Rated amperage	Geography
6653	IEC 309, 3P+N+G, 16A	230	3	16 amps/phase	Internationally available
6489	IEC309 3P+N+G, 32A	230	3	32 amps/phase	EMEA
6654	NEMA L6-30	200-208, 240	1	24 amps	US, Canada, LA, Japan
6655	RS 3750DP (watertight)	200-208, 240	1	24 amps	US, Canada, LA, Japan
6656	IEC 309, P+N+G, 32A	230	1	24 amps	EMEA
6657	PDL	230-240	1	32 amps	Australia, New Zealand
6658	Korean plug	220	1	30 amps	North and South Korea
6492	IEC 309, 2P+G, 60A	200-208, 240	1	48 amps	US, Canada, LA, Japan
6491	IEC 309, P+N+G, 63A	230	1	63 amps	EMEA

Notes: Ensure that the appropriate power cord feature is configured to support the power that is being supplied. Based on the power cord that is used, the PDU can supply 4.8 - 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.

To better enable electrical redundancy, each server has two power supplies that must be connected to separate PDUs, which are not included in the base order.

For maximum availability, a preferred approach is to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to independent power sources.

For detailed power requirements and power cord details about the 7014 racks, see the “Planning for power” section in the IBM Power Systems Hardware Information Center website, found at:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/topic/p7had/p7hadrpower.htm

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000050

For detailed power requirements and power cord details about the 7965-94Y rack, see the "Planning for power" section in the IBM Power Systems Hardware Information Center website, found at:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/topic/p7had/p7hadkickoff795394x.htm

1.10.8 Rack-mounting rules

Consider the following primary rules when you mount the system into a rack:

- The system is designed to be placed at any location in the rack. For rack stability, start filling a rack from the bottom.
- Any remaining space in the rack can be used to install other systems or peripheral devices, if 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, be sure to follow the rack manufacturer's safety instructions regarding rack stability.

Order information: The racking approach for the initial order must be either a 7014-T00, 7014-T42, or 7965-94Y. If an additional rack is required for I/O expansion drawers as an MES to an existing system, either a feature 0551, 0553, or ER05 rack must be ordered.

You must leave 2U of space at either the bottom or top of the rack, depending on the client's cabling preferences, to allow for cabling to exit the rack

1.10.9 Useful rack additions

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

IBM System Storage 7226 Model 1U3 Multi-Media Enclosure

The IBM System Storage 7226 Model 1U3 Multi-Media Enclosure can accommodate up to two tape drives, two RDX removable disk drive docking stations, or up to four DVD-RAM drives.

The IBM System Storage 7226 Multi-Media Enclosure supports LTO Ultrium and DAT160 Tape technology, DVD-RAM, and RDX removable storage requirements on the following IBM systems:

- IBM POWER6 processor-based systems
- IBM POWER7 processor-based systems
- IBM POWER8 processor-based systems
- IBM POWER processor-based BladeCenters (supports SAS tape drive devices only)
- IBM POWER processor-based PureFlex® Systems (supports Fibre Channel and USB devices only)

The IBM System Storage 7226 Multi-Media Enclosure offers an expansive list of drive feature options, as shown in Table 1-16 on page 36.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000051

Table 1-16 Supported drive features for the 7226-1U3

Feature code	Description	Status
5619	DAT160 SAS Tape Drive	Available
EU16	DAT160 USB Tape Drive	Available
1420	DVD-RAM SAS Optical Drive	Available
1422	DVD-RAM Slim SAS Optical Drive	Available
5762	DVD-RAM USB Optical Drive	Available
5763	DVD Front USB Port Sled with DVD-RAM USB Drive	Available
5757	DVD RAM Slim USB Optical Drive	Available
8248	LTO Ultrium 5 Half High Fibre Tape Drive	Available
8241	LTO Ultrium 5 Half High SAS Tape Drive	Available
8348	LTO Ultrium 6 Half High Fibre Tape Drive	Available
8341	LTO Ultrium 6 Half High SAS Tape Drive	Available
EU03	RDX 3.0 Removable Disk Docking Station	Available

Option descriptions are as follows:

- DAT160 160 GB Tape Drives: With SAS or USB interface options and a data transfer rate up to 12 MBps (assumes 2:1 compression), the DAT160 drive is read/write compatible with DAT160, and DDS4 data cartridges.
- LTO Ultrium 5 Half-High 1.5 TB SAS and FC Tape Drive: With a data transfer rate up to 280 MBps (assuming a 2:1 compression), the LTO Ultrium 5 drive is read/write compatible with LTO Ultrium 5 and 4 data cartridges, and read-only compatible with Ultrium 3 data cartridges. By using data compression, an LTO-5 cartridge can store up to 3 TB of data.
- LTO Ultrium 6 Half-High 2.5 TB SAS and FC Tape Drive: With a data transfer rate up to 320 MBps (assuming a 2.5:1 compression), the LTO Ultrium 6 drive is read/write compatible with LTO Ultrium 6 and 5 media, and read-only compatibility with LTO Ultrium 4. By using data compression, an LTO-6 cartridge can store up to 6.25 TB of data.
- DVD-RAM: The 9.4 GB SAS Slim Optical Drive with an SAS and USB interface option is compatible with most standard DVD disks.
- RDX removable disk drives: The RDX USB docking station is compatible with most RDX removable disk drive cartridges when it is used in the same operating system. The 7226 offers the following RDX removable drive capacity options:
  - 500 GB (#1107)
  - 1.0 TB (#EU01)
  - 2.0 TB (#EU2T)

Removable RDX drives are in a rugged cartridge that inserts in to an RDX removable (USB) disk docking station (#1103 or #EU03). RDX drives are compatible with docking stations, which are installed internally in IBM POWER6, POWER6+, POWER7, POWER7+, and POWER8 servers, where applicable.

Media that is used in the 7226 DAT160 SAS and USB tape drive features are compatible with DAT160 tape drives that are installed internally in IBM POWER6, POWER6+, POWER7, POWER7+, and POWER8 servers, and in IBM BladeCenter systems.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000052

Media that is used in LTO Ultrium 5 Half-High 1.5 TB tape drives are compatible with Half High LTO5 tape drives that are installed in the IBM TS2250 and TS2350 external tape drives, IBM LTO5 tape libraries, and half-high LTO5 tape drives that are installed internally in IBM POWER6, POWER6+, POWER7, POWER7+, and POWER8 servers.

Figure 1-11 shows the IBM System Storage 7226 Multi-Media Enclosure.

Figure 1-11 IBM System Storage 7226 Multi-Media Enclosure

The IBM System Storage 7226 Multi-Media Enclosure offers customer-replaceable unit (CRU) maintenance service to help make the installation or replacement of new drives efficient. Other 7226 components are also designed for CRU maintenance.

The IBM System Storage 7226 Multi-Media Enclosure is compatible with most IBM POWER6, POWER6+, POWER7, POWER7+, and POWER8 systems, and also with the IBM BladeCenter models (PS700, PS701, PS702, PS703, and PS704) that offer current level AIX, IBM i, and Linux operating systems.

Unsupported: IBM i does not support 7226 USB devices.

For a complete list of host software versions and release levels that support the IBM System Storage 7226 Multi-Media Enclosure, see the following System Storage Interoperation Center (SSIC) website, found at:

http://www.ibm.com/systems/support/storage/config/ssic/index.jsp

Note: Any of the existing 7216-1U2, 7216-1U3, and 7214-1U2 multimedia drawers are also supported.

Flat panel display options

The IBM 7316 Model TF4 is a rack-mountable flat panel console kit that can also be configured with the tray pulled forward and the monitor folded up, providing full viewing and keying capability for the HMC operator.

The Model TF4 is a follow-on product to the Model TF3 and offers the following features:

- A slim, sleek, and lightweight monitor design that occupies only 1U (1.75 in.) in a 19-inch standard rack.
- A 18.5-inch (409.8 mm x 230.4 mm) flat panel TFT monitor with truly accurate images and virtually no distortion.
- The ability to mount the IBM Travel Keyboard in the 7316-TF4 rack keyboard tray.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000053

- Support for the IBM 1x8 Rack Console Switch (#4283) IBM Keyboard/Video/Mouse (KVM) switches.

  Feature 4283 is a 1x8 Console Switch that fits in the 1U space behind the TF4. It is a CAT5 based switch containing eight rack interface (ARI) ports for connecting either PS/2 or USB console switch cables. It supports chaining of servers using IBM Conversion Options switch cable feature 4269. This feature provides four cables that connect a KVM switch to a system, or can be used in a daisy-chain scenario to connect up to 128 systems to a single KVM switch. It also supports server-side USB attachments.

1.10.10 OEM rack

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

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

The website mentions the following key points:

- 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 1-12 is a top view showing the specification dimensions.

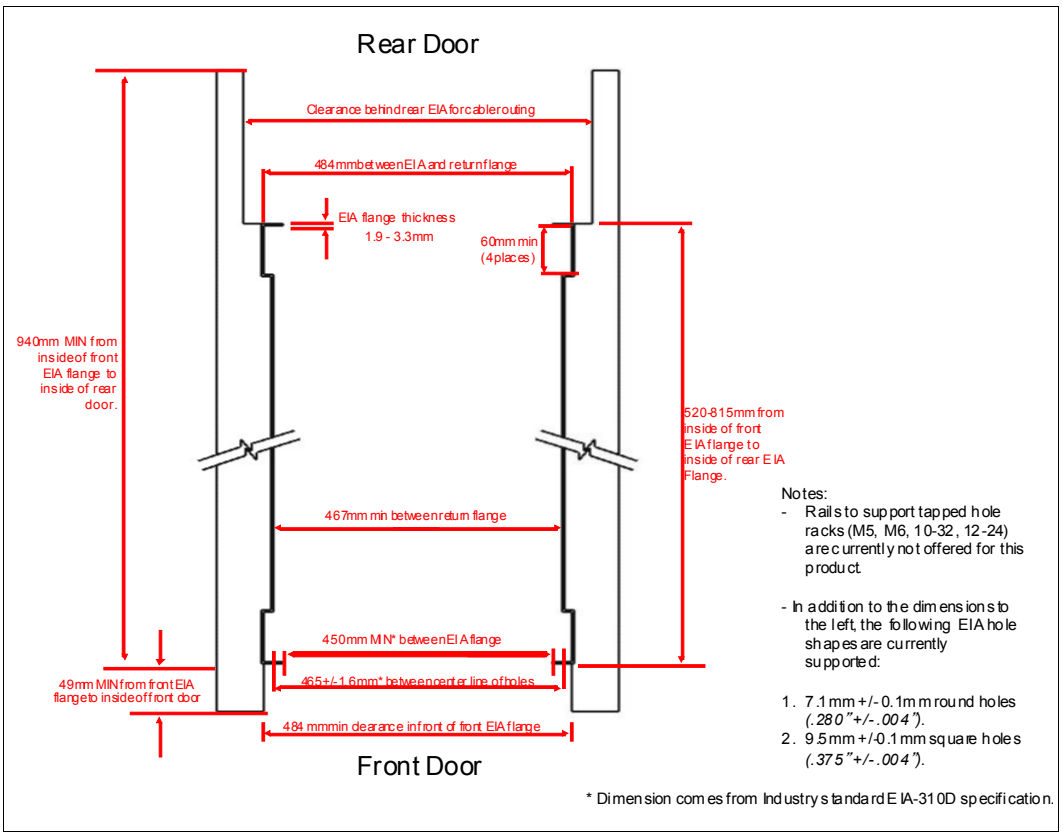

Figure 1-12 Top view of rack specification dimensions (not specific to IBM)

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000054

- 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 1-13 shows the top front specification dimensions.

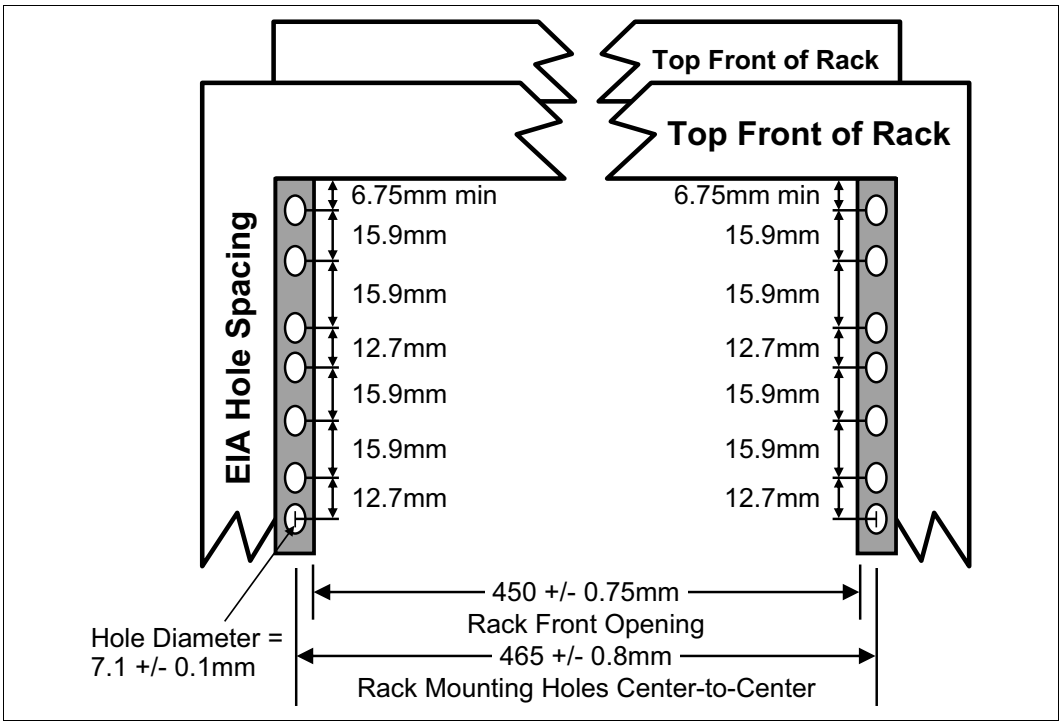

Figure 1-13 Rack specification dimensions - top front view

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000055

2

Architecture and technical overview

This chapter describes the overall system architecture for the IBM Power System S814 (8286-41A) and IBM Power System S824 (8286-42A) servers. The bandwidths that are provided throughout the section are theoretical maximums that are used for reference.

The speeds that are 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 by using real-world performance measurements and production workloads.

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000057

Figure 2-1 shows the logical system diagram for the Power S814.

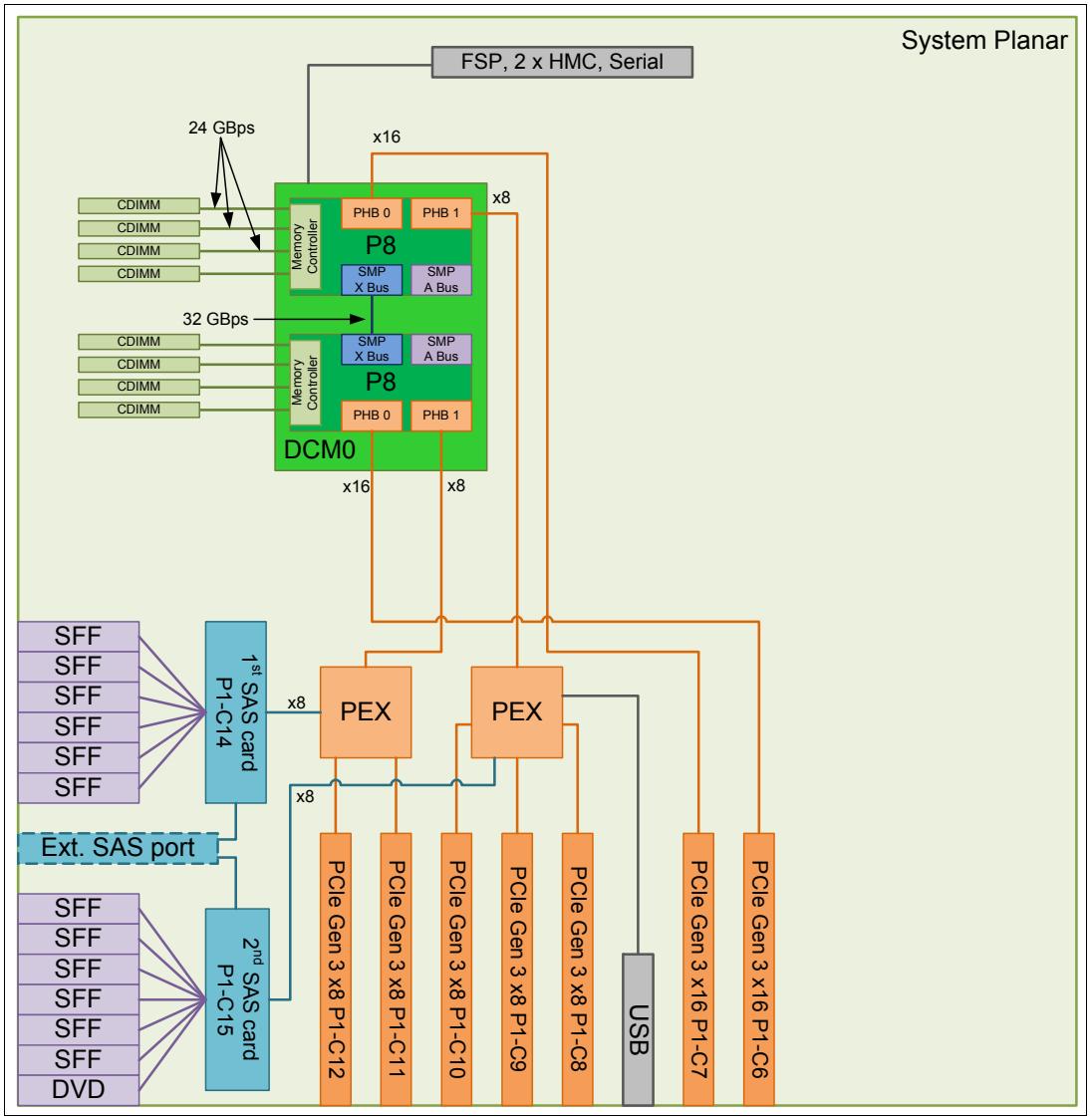

Figure 2-1 IBM Power S814 logical system diagram

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000058

Figure 2-2 shows the logical system diagram for the Power S824.

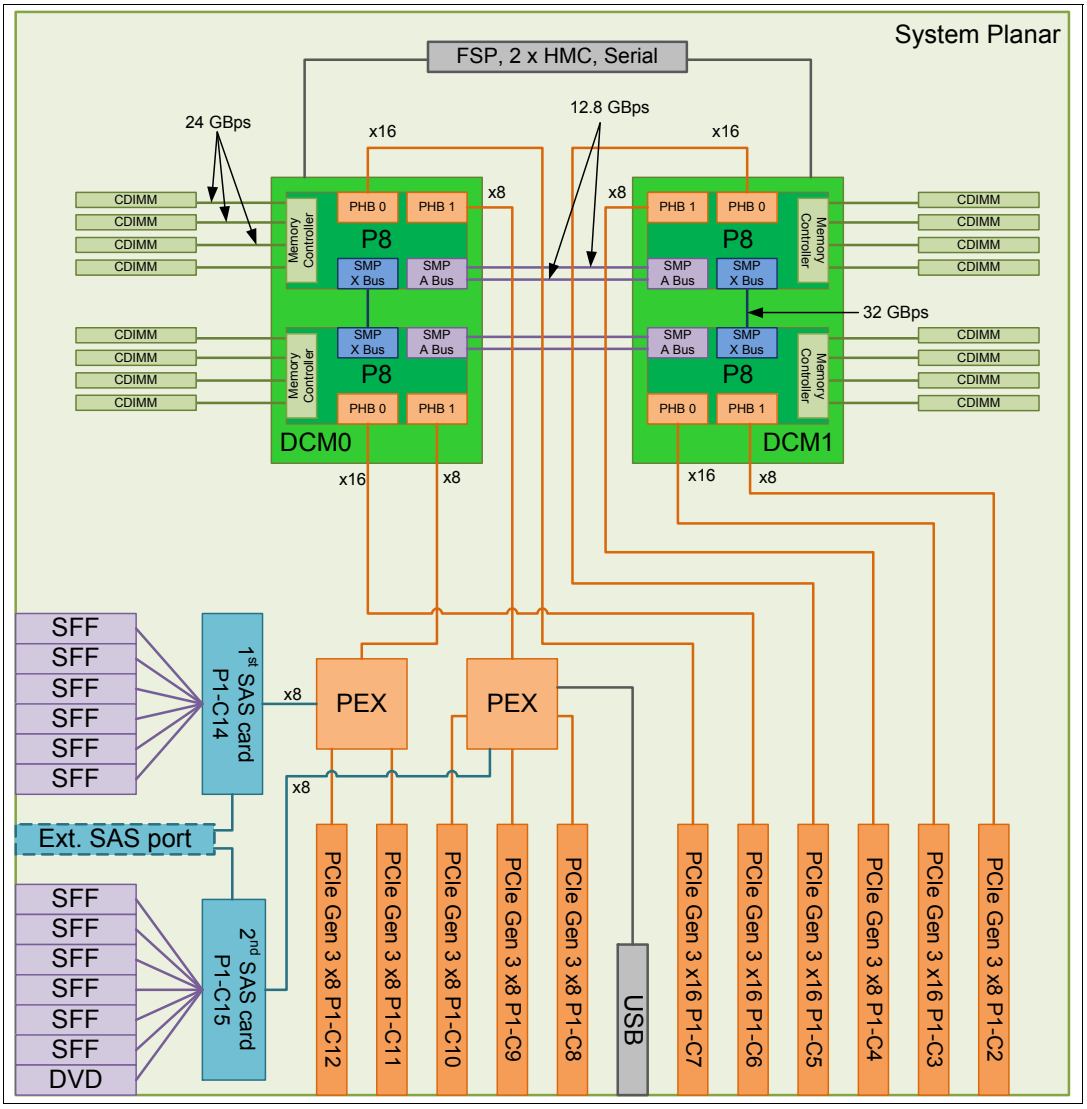

Figure 2-2 IBM Power S824 logical system diagram

2.1 The IBM POWER8 processor

This section introduces the latest processor in the IBM Power Systems product family, and describes its main characteristics and features in general.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000059

2.1.1 POWER8 processor overview

The POWER8 processor is manufactured by the IBM 22 nm Silicon-On-Insulator (SOI) technology. Each chip is 649 mm^2 and contains 4.2 billion transistors. As shown in Figure 2-3, the chip contains 12 cores, two memory controllers, PCIe Gen3 I/O controllers, and an interconnection system that connects all components within the chip. Each core has 512 KB of L2 cache, and all cores share 96 MB of L3 embedded DRAM (eDRAM). The interconnect also extends through module and board technology to other POWER8 processors in addition to DDR3 memory and various I/O devices.

POWER8 systems use memory buffer chips to interface between the POWER8 processor and DDR3 or DDR4 memory.[1] Each buffer chip also includes an L4 cache to reduce the latency of local memory accesses.

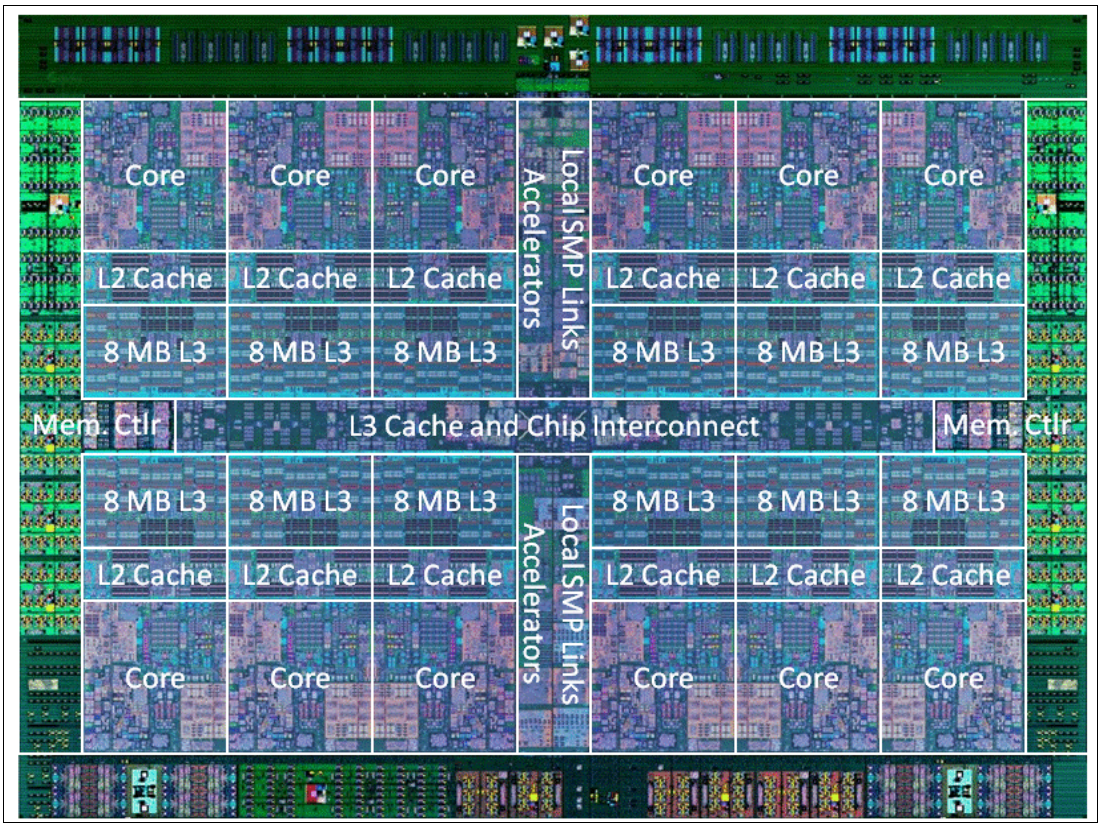

Figure 2-3 The POWER8 processor chip

The POWER8 processor is for system offerings from single-socket servers to multi-socket Enterprise servers. It incorporates a triple-scope broadcast coherence protocol over local and global SMP links to provide superior scaling attributes. Multiple-scope coherence protocols reduce the amount of SMP link bandwidth that is required by attempting operations on a limited scope (single chip or multi-chip group) when possible. If the operation cannot complete coherently, the operation is reissued using a larger scope to complete the operation.

[1] At the time of the publication, the available POWER8 processor-based systems use DDR3 memory.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000060

Here are additional features that can augment the performance of the POWER8 processor:

- Support for DDR3 and DDR4 memory through memory buffer chips that offload the memory support from the POWER8 memory controller.
- L4 cache within the memory buffer chip that reduces the memory latency for local access to memory behind the buffer chip; the operation of the L4 cache is not apparent to applications running on the POWER8 processor. Up to 128 MB of L4 cache can be available for each POWER8 processor.
- Hardware transactional memory.
- On-chip accelerators, including on-chip encryption, compression, and random number generation accelerators.
- Coherent Accelerator Processor Interface, which allows accelerators plugged into a PCIe slot to access the processor bus using a low latency, high-speed protocol interface.
- Adaptive power management.

There are two versions of the POWER8 processor chip. Both chips use the same building blocks. The scale-out systems use a 6-core version of POWER8. The 6-core chip is installed in pairs in a Dual Chip Module (DCM) that plugs into a socket in the system board of the systems. Functionally, it works as a single chip.

Figure 2-4 shows a graphic representation of the 6-core processor.

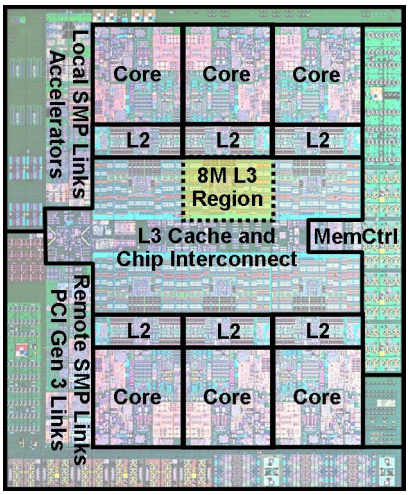

Figure 2-4 6-core POWER8 processor chip

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000061

Table 2-1 summarizes the technology characteristics of the POWER8 processor.

Table 2-1 Summary of POWER8 processor technology

Technology	POWER8 processor
Die size	649 mm^2
Fabrication technology	► 22 nm lithography ► Copper interconnect ► SOI ► eDRAM
Maximum processor cores	6 or 12
Maximum execution threads core/chip	8/96
Maximum L2 cache core/chip	512 KB/6 MB
Maximum On-chip L3 cache core/chip	8 MB/96 MB
Maximum L4 cache per chip	128 MB
Maximum memory controllers	2
SMP design-point	16 sockets with IBM POWER8 processors
Compatibility	With prior generation of POWER processor

2.1.2 POWER8 processor core

The POWER8 processor core is a 64-bit implementation of the IBM Power Instruction Set Architecture (ISA) Version 2.07 and has the following features:

- Multi-threaded design, which is capable of up to eight-way simultaneous multithreading (SMT)
- 32 KB, eight-way set-associative L1 instruction cache
- 64 KB, eight-way set-associative L1 data cache
- Enhanced prefetch, with instruction speculation awareness and data prefetch depth awareness
- Enhanced branch prediction, using both local and global prediction tables with a selector table to choose the best predictor
- Improved out-of-order execution
- Two symmetric fixed-point execution units
- Two symmetric load/store units and two load units, all four of which can also run simple fixed-point instructions
- An integrated, multi-pipeline vector-scalar floating point unit for running both scalar and SIMD-type instructions, including the Vector Multimedia eXtension (VMX) instruction set and the improved Vector Scalar eXtension (VSX) instruction set, and capable of up to sixteen floating point operations per cycle (eight double precision or sixteen single precision)
- In-core Advanced Encryption Standard (AES) encryption capability
- Hardware data prefetching with 16 independent data streams and software control
- Hardware decimal floating point (DFP) capability

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000062

More information about Power ISA Version 2.07 can be found at the following website:

https://www.power.org/documentation/power-isa-version-2-07/

Figure 2-5 shows a picture of the POWER8 core, with some of the functional units highlighted.

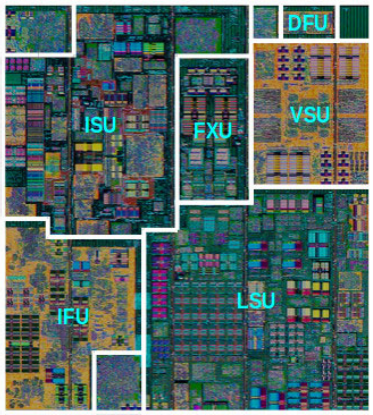

Figure 2-5 POWER8 processor core

2.1.3 Simultaneous multithreading

POWER8 processor advancements in multi-core and multi-thread scaling are remarkable. A significant performance opportunity comes from parallelizing workloads to enable the full potential of the microprocessor, and the large memory bandwidth. Application scaling is influenced by both multi-core and multi-thread technology.

Simultaneous Multithreading (SMT) allows a single physical processor core to simultaneously dispatch instructions from more than one hardware thread context. With SMT, each POWER8 core can present eight hardware threads. Because there are multiple hardware threads per physical processor core, additional instructions can run at the same time. SMT is primarily beneficial in commercial environments where the speed of an individual transaction is not as critical as the total number of transactions that are performed. SMT typically increases the throughput of workloads with large or frequently changing working sets, such as database servers and web servers.

Table 2-2 shows a comparison between the different POWER processors in terms of SMT capabilities that are supported by each processor architecture.

Table 2-2 SMT levels that are supported by POWER processors

Technology	Cores/system	Maximum SMT mode	Maximum hardware threads per partition
IBM POWER4	32	Single Thread (ST)	32
IBM POWER5	64	SMT2	128
IBM POWER6	64	SMT2	128
IBM POWER7	256	SMT4	1024
IBM POWER8	192	SMT8	1536

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000063

The architecture of the POWER8 processor, with its larger caches, larger cache bandwidth, and faster memory, allows threads to have faster access to memory resources, which translates in to a more efficient usage of threads. Because of that, POWER8 allows more threads per core to run concurrently, increasing the total throughput of the processor and of the system.

2.1.4 Memory access

On the Power S814 and Power S824, each POWER8 module has two memory controllers, each connected to four memory channels. Each memory channel operates at 1600 MHz and connects to a DIMM. Each DIMM on a POWER8 system has a memory buffer that is responsible for many functions that were previously on the memory controller, such as scheduling logic and energy management. The memory buffer also has 16 MB of level 4 (L4) cache.

Each memory channel can address up to 128 GB. Therefore, the Power S814 can address up to 1 TB of total memory, and the Power S824 can address up to 2 TB of total memory.

Figure 2-6 gives a simple overview of the POWER8 processor memory access structure in the Power S814 and Power S824.

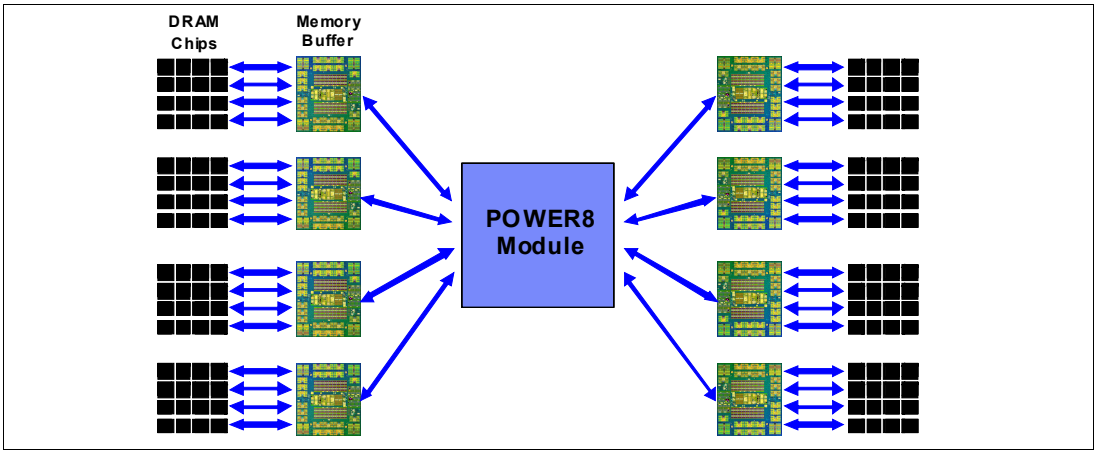

Figure 2-6 Overview of POWER8 memory access structure

2.1.5 On-chip L3 cache innovation and Intelligent Cache

Similar to POWER7 and POWER7+, the POWER8 processor uses a breakthrough in material engineering and microprocessor fabrication to implement the L3 cache in eDRAM and place it on the processor die. L3 cache is critical to a balanced design, as is the ability to provide good signaling 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 8 MB 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 that is associated with another core, depending on the reference patterns. Data can also be cloned and stored in more than one core's FLR-L3 cache, again depending on the reference patterns. This Intelligent Cache management enables the POWER8 processor to optimize the access to L3 cache lines and minimize overall cache latencies.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000064

Figure 2-3 on page 44 and Figure 2-4 on page 45 show the on-chip L3 cache, and highlight one fast 8 MB L3 region closest to a processor core.

The innovation of using eDRAM on the POWER8 processor die is significant for several 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 performance of eDRAM when implemented on-chip is similar to conventional SRAM but requires far less physical space. IBM on-chip eDRAM uses only a third of the components that conventional SRAM uses, 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.

2.1.6 Level 4 cache and memory buffer

POWER8 processor-based systems introduce an additional level in memory hierarchy. The Level 4 (L4) cache is implemented together with the memory buffer in the Custom DIMM (CDIMM). Each memory buffer contains 16 MB of L4 cache. On a Power S814, you can have up to 128 MB of L4 cache, and on a Power S824 you can have up to 256 MB of L4 cache.

Figure 2-7 shows a picture of the memory buffer, where you can see the 16 MB L4 cache, and processor links and memory interfaces.

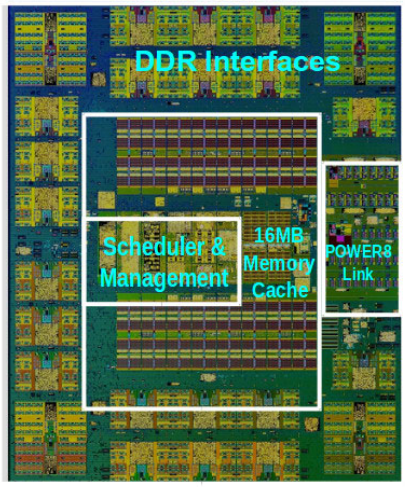

Figure 2-7 Memory buffer chip

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000065

Table 2-3 shows a comparison of the different levels of cache in the POWER7, POWER7+, and POWER8 processors.

Table 2-3 POWER8 cache hierarchy

Cache	POWER7	POWER7+	POWER8
L1 instruction cache: Capacity/associativity	32 KB, 4-way	32 KB, 4-way	32 KB, 8-way
L1 data cache: Capacity/associativity bandwidth	32 KB, 8-way Two 16 B reads or one 16 B writes per cycle	32 KB, 8-way Two 16 B reads or one 16 B writes per cycle	64 KB, 8-way Two 16 B reads or one 16 B writes per cycle
L2 cache: Capacity/associativity bandwidth	256 KB, 8-way Private 32 B reads and 16 B writes per cycle	256 KB, 8-way Private 32 B reads and 16 B writes per cycle	512 KB, 8-way Private 32 B reads and 16 B writes per cycle
L3 cache: Capacity/associativity bandwidth	On-Chip 4 MB/core, 8-way 16 B reads and 16 B writes per cycle	On-Chip 10 MB/core, 8-way 16 B reads and 16 B writes per cycle	On-Chip 8 MB/core, 8-way 32 B reads and 32 B writes per cycle
L4 cache: Capacity/associativity bandwidth	N/A	N/A	Off-Chip 16 MB/buffer chip, 16-way Up to 8 buffer chips per socket

For more information about the POWER8 memory subsystem, see 2.2, “Memory subsystem” on page 53.

2.1.7 Hardware transactional memory

Transactional memory is an alternative to lock-based synchronization. It attempts to simplify parallel programming by grouping read and write operations and running them like a single operation. Transactional memory is like database transactions where all shared memory accesses and their effects are either committed all together or discarded as a group. All threads can enter the critical region simultaneously. If there are conflicts in accessing the shared memory data, threads try accessing the shared memory data again or are stopped without updating the shared memory data. Therefore, transactional memory is also called a lock-free synchronization. Transactional memory can be a competitive alternative to lock-based synchronization.

Transactional Memory provides a programming model that makes parallel programming easier. A programmer delimits regions of code that access shared data and the hardware runs these regions atomically and in isolation, buffering the results of individual instructions, and retrying execution if isolation is violated. Generally, transactional memory allows programs to use a programming style that is close to coarse-grained locking to achieve performance that is close to fine-grained locking.

Most implementations of transactional memory are based on software. The POWER8 processor-based systems provide a hardware-based implementation of transactional memory that is more efficient than the software implementations and requires no interaction with the processor core, therefore allowing the system to operate in maximum performance.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000066

2.1.8 Coherent Accelerator Processor Interface

The Coherent Accelerator Interface Architecture (CAIA) defines a coherent accelerator interface structure for attaching especially to the Power Systems.

The Coherent Accelerator Processor Interface (CAPI) can attach accelerators that have coherent shared memory access to the processors in the server and share full virtual address translation with these processors, using a standard PCIe Gen3 bus.

Applications can have customized functions in Field Programmable Gate Arrays (FPGA) and be able to enqueue work requests directly in shared memory queues to the FPGA, and using the same effective addresses (pointers) it uses for any of its threads running on a host processor. From the practical perspective, CAPI allows a specialized hardware accelerator to be seen as an additional processor in the system, with access to the main system memory, and coherent communication with other processors in the system.

The benefits of using CAPI include the ability to access shared memory blocks directly from the accelerator, perform memory transfers directly between the accelerator and processor cache, and reduction on the code path length between the adapter and the processors because the adapter is not operating as a traditional I/O device, and there is no device driver layer to perform processing. It also presents a simpler programming model.

Figure 2-8 shows a high-level view of how an accelerator communicates with the POWER8 processor through CAPI. The POWER8 processor provides a Coherent Attached Processor Proxy (CAPP), which is responsible for extending the coherence in the processor communications to an external device. The coherency protocol is tunneled over standard PCIe Gen3, effectively making the accelerator part of the coherency domain.

The accelerator adapter implements the Power Service Layer (PSL), which provides address translation and system memory cache for the accelerator functions. The custom processors on the board, consisting of an FPGA or an Application Specific Integrated Circuit (ASIC) use this layer to access shared memory regions, and cache areas as through they were a processor in the system. This ability greatly enhances the performance of the data access for the device and simplifies the programming effort to use the device. Instead of treating the hardware accelerator as an I/O device, it is treated as a processor. This eliminates the requirement of a device driver to perform communication, and the need for Direct Memory Access that requires system calls to the operating system kernel. By removing these layers, the data transfer operation requires fewer clock cycles in the processor, greatly improving the I/O performance.

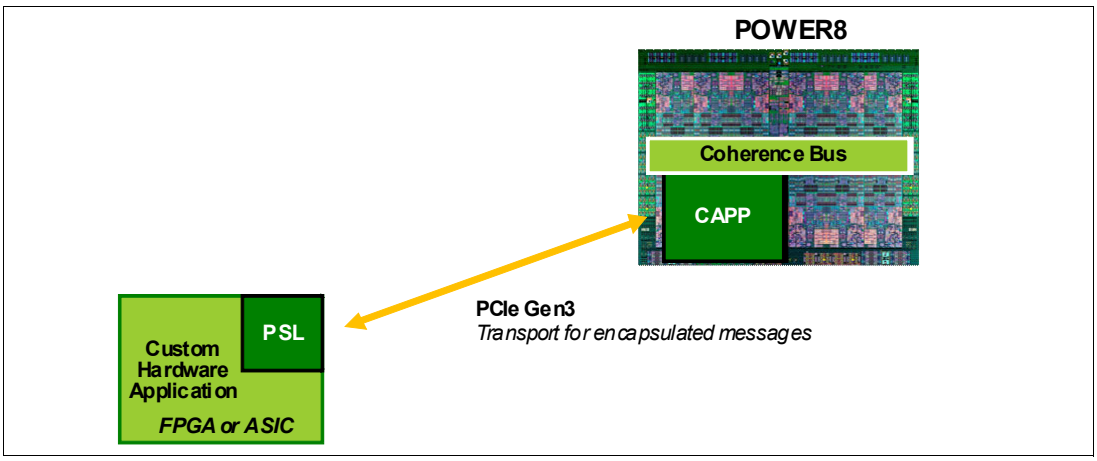

Figure 2-8 CAPI accelerator that is attached to the POWER8 processor

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000067

The implementation of CAPI on the POWER8 processor allows hardware companies to develop solutions for specific application demands and use the performance of the POWER8 processor for general applications and a custom acceleration of specific functions using a hardware accelerator, with a simplified programming model and efficient communication with the processor and memory resources.

For a list of available CAPI adapters, see 2.5.11, “CAPI adapters” on page 71.

2.1.9 Power management and system performance

The POWER8 processor has power saving and performance enhancing features that can be used to lower overall energy usage, while yielding higher performance when needed. The following modes can be enabled and modified to use these features.

Dynamic Power Saver: Favor Performance

This mode is intended to provide the best performance. If the processor is being used even moderately, the frequency is raised to the maximum frequency possible to provide the best performance. If the processors are lightly used, the frequency is lowered to the minimum frequency, which is potentially far below the nominal shipped frequency, to save energy. The top frequency that is achieved is based on system type and is affected by environmental conditions. Also, when running at the maximum frequency, more energy is being consumed, which means this mode can potentially cause an increase in overall energy consumption.

Dynamic Power Saver: Favor Power

This mode is intended to provide the best performance per watt consumed. The processor frequency is adjusted based on the processor usage to maintain the workload throughput without using more energy than required to do so. At high processor usage levels, the frequency is raised above nominal, as in the Favor Performance mode. Likewise, at low processor usage levels, the frequency is lowered to the minimum frequency. The frequency ranges are the same for the two Dynamic Power Saver modes, but the algorithm that determines which frequency to set is different.

Dynamic Power Saver: Tunable Parameters

The Dynamic Power Saver: Favor Performance and Dynamic Power Saver: Favor Power modes are tuned to provide both energy savings and performance increases. However, there might be situations where only top performance is of concern, or, conversely, where peak power consumption is an issue. The tunable parameters can be used to modify the setting of the processor frequency in these modes to meet these various objectives. Modifying these parameters should be done only by advanced users. If you must address any issues concerning the Tunable Parameters, IBM support personal should be directly involved in the parameter value selection.

Idle Power Saver

This mode is intended to save the maximum amount of energy when the system is nearly idle. When the processors are found to be nearly idle, the frequency of all processors is lowered to the minimum. Additionally, workloads are dispatched onto a smaller number of processor cores so that the other processor cores can be put into a low energy usage state. When processor usage increases, the process is reversed: The processor frequency is raised back up to nominal, and the workloads are spread out once again over all of the processor cores. There is no performance boosting aspect in this mode, but entering or exiting this mode might affect overall performance. The delay times and usage levels for entering and exiting this mode can be adjusted to allow for more or less aggressive energy savings.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000068

The controls for all of these modes are available on the Advanced System Management Interface (ASMI) and are described in more detail in a white paper that is found at the following link:

http://public.dhe.ibm.com/common/ssi/ecm/en/pow03039usen/POW03039USEN.PDF

For more information, see 2.11, “Energy management” on page 96.

2.1.10 Comparison of the POWER8, POWER7+, and POWER7 processors

Table 2-4 shows comparable characteristics between the generations of POWER8, POWER7+, and POWER7 processors.

Table 2-4 Comparison of technology for the POWER8 processor and the prior generations

Characteristics	POWER8	POWER7+	POWER7
Technology	22 nm	32 nm	45 nm
Die size	649 mm^2	567 mm^2	567 mm^2
Number of transistors	4.2 billion	2.1 billion	1.2 billion
Maximum cores	12	8	8
Maximum SMT threads per core	8 threads	4 threads	4 threads
Maximum frequency	4.15 GHz	4.4 GHz	4.25 GHz
L2 Cache	512 KB per core	256 KB per core	256 KB per core
L3 Cache	8 MB of FLR-L3 cache per core with each core having access to the full 96 MB of L3 cache, on-chip eDRAM	10 MB of FLR-L3 cache per core with each core having access to the full 80 MB of L3 cache, on-chip eDRAM	4 MB or 8 MB of FLR-L3 cache per core with each core having access to the full 32 MB of L3 cache, on-chip eDRAM
Memory support	DDR3 and DDR4	DDR3	DDR3
I/O bus	PCIe Gen3	GX++	GX++

2.2 Memory subsystem

The Power S814 is a one-socket system that supports a single POWER8 processor module. The server supports a maximum of eight DDR3 CDIMM slots. Memory features that are supported are 16 GB, 32 GB, 64 GB, and 128 GB, and run at speeds of 1600 MHz, allowing for a maximum system memory of 1 TB.

Note: If using the 4-core memory feature #EPXK the maximum number of CDIMMs available are four and the only memory size allowed is 16 GB, allowing the server to a maximum of 64 GB RAM.

The Power S824 is a two-socket system that supports up to two POWER8 processor modules. The server supports a maximum of 16 DDR3 CDIMM slots, with eight DIMM slots per installed processor. Memory features that are supported are 16 GB, 32 GB, 64 GB, and 128 GB, and run at speeds of 1600 MHz, allowing for a maximum system memory of 2 TB.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000069

These servers support an optional feature that is called Active Memory Expansion that allows the effective maximum memory capacity to be much larger than the true physical memory. This feature runs innovative compression and decompression of memory content by using a dedicated coprocessor to provide memory expansion up to 125%, depending on the workload type and its memory usage. As an example, a server with 256 GB RAM physically installed can effectively be expanded over 512 GB RAM. This approach can enhance virtualization and server consolidation by allowing a partition to do more work with the same physical amount of memory or a server to run more partitions and do more work with the same physical amount of memory.

2.2.1 Custom DIMM

Custom DIMMs (CDIMMs) are innovative memory DIMMs that do house industry-standard DRAM memory chips and a set of components that allow for higher bandwidth and lower latency communications:

- Memory Scheduler
- Memory Management (RAS Decisions & Energy Management)
- Buffer Cache

By adopting this architecture for the memory DIMMs, several decisions and processes regarding memory optimizations are run internally into the CDIMM, saving bandwidth and allowing for faster processor to memory communications. This also allows for a more robust RAS. For more information, see Chapter 4, “Reliability, availability, and serviceability” on page 137.

A detailed diagram of the CDIMM that is available for the Power S814 and Power S824 can be seen in Figure 2-9.

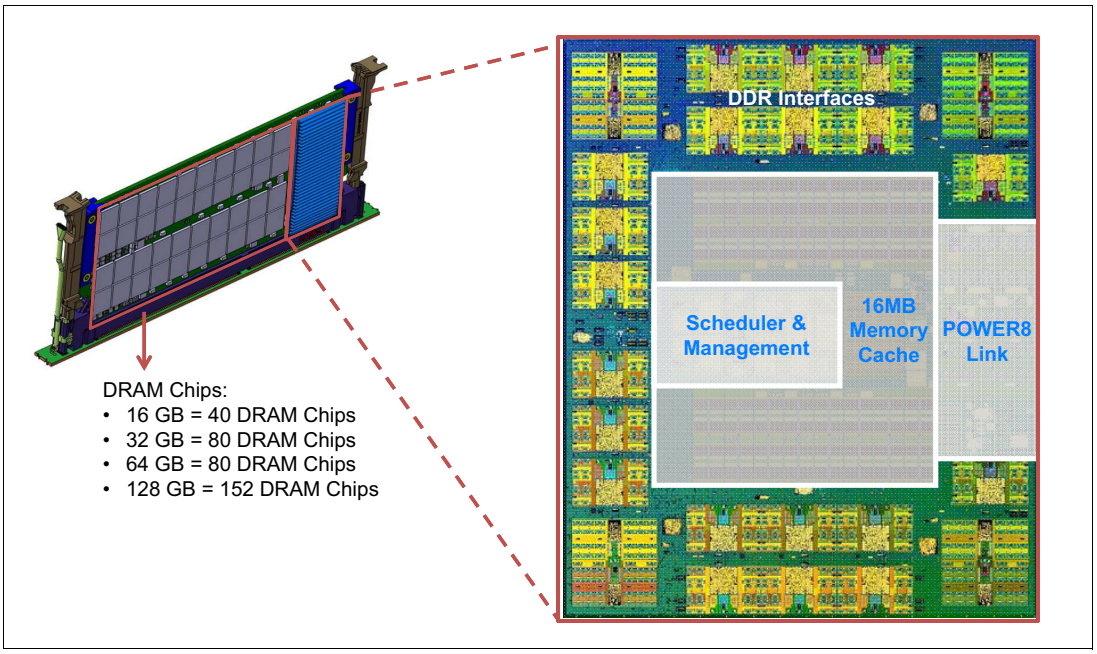

Figure 2-9 A short CDIMM on a tall CDIMM slot diagram

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000070

The CDIMMs slots for the Power S814 and Power S824 are tall CDIMMs slots. At the time of writing, only short CDIMMs are available and they are compatible with tall CDIMMs slots by having an integrated extender that allows for proper airflow and ease of handling. Tall CDIMMs slots allow for larger DIMM sizes and the adoption of future technologies more seamlessly.

The Buffer Cache is a L4 cache and is built on eDRAM technology (same as the L3 cache), which has a lower latency than regular SRAM. Each CDIMM has 16 MB of L4 cache and a fully populated Power S824 server (two processors and 16 CDIMMs) has 128 MB of L4 Cache. The L4 Cache performs several functions that have direct impact in performance and brings a series of benefits for the Power S814 and Power S824:

- Reduces energy consumption by reducing the number of memory requests.
- Increases memory write performance by acting as a cache and by grouping several random writes into larger transactions.
- Partial write operations that target the same cache block are “gathered” within the L4 cache before being written to memory, becoming a single write operation.
- Reduces latency on memory access. Memory access for cached blocks has up to 55% lower latency than non-cached blocks.

2.2.2 Memory placement rules

The following memory options are orderable:

- 16 GB CDIMM, 1600 MHz DDR3 DRAM (#EM8B)
- 32 GB CDIMM, 1600 MHz DDR3 DRAM (#EM8Z)
- 64 GB CDIMM, 1600 MHz DDR3 DRAM (#EM8D)
- 128 GB CDIMM, 1600 MHz DDR3 DRAM (#EM8E)

Note: If using the 4-core memory feature #EPXK the only memory size allowed is 16 GB (#EM8B)

All memory must be ordered in pairs with a minimum of 16 GB (S814) or 32 GB (S824) per system.

The supported maximum memory is as follows:

- One processor module installed (4-core): 64 GB (four 16 GB CDIMMs)
- One processor module installed (6-core or 8-core): 1 TB (eight 128 GB CDIMMs)
- Two processors modules installed: 2 TB (sixteen 128 GB CDIMMs)

The basic rules for memory placement follow:

- Each feature code equates to a single physical CDIMM.
- All memory features must be ordered in pairs.
- All memory CDIMMs must be installed in pairs.
- Each CDIMM within a pair must be of the same capacity.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000071

Figure 2-10 shows the physical memory DIMM topology.

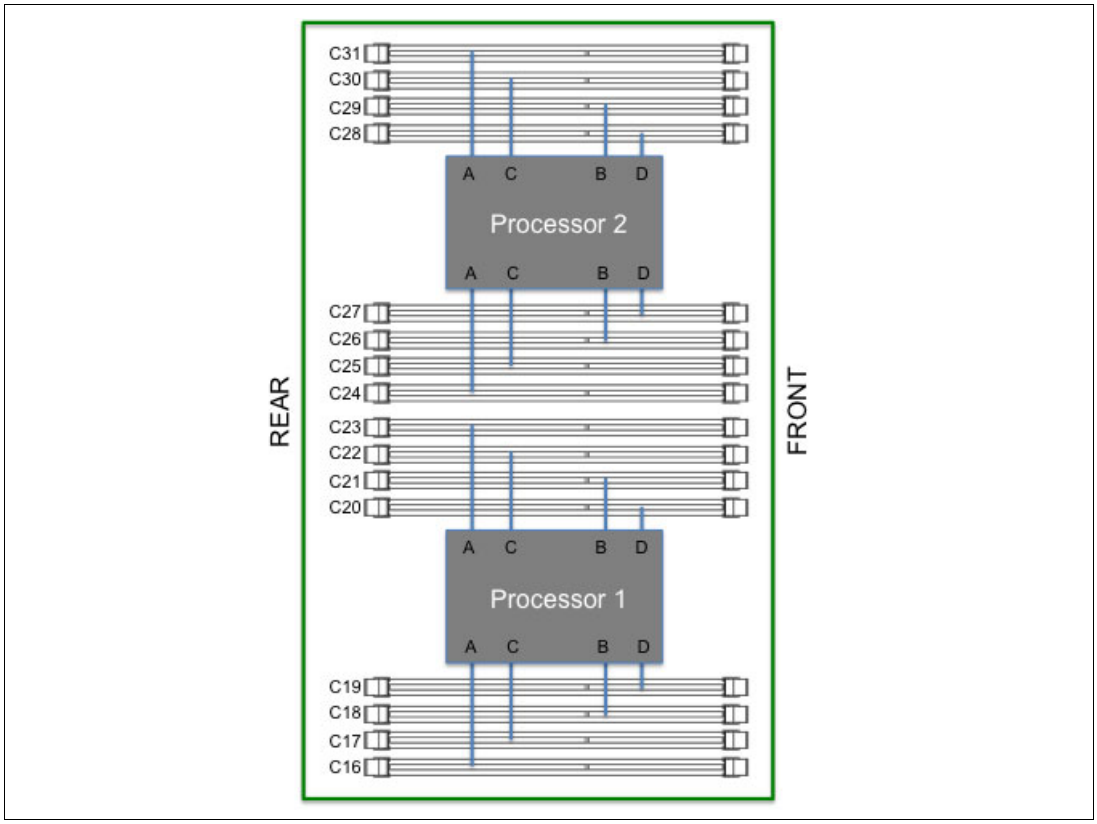

Figure 2-10 Memory DIMM topology for the Power S814 and Power S824

In general, the best approach is to install memory evenly across all processors in the system. Balancing memory across the installed processors allows memory access in a consistent manner and typically results in the best possible performance for your configuration. You should account for any plans for future memory upgrades when you decide which memory feature size to use at the time of the initial system order.

For systems with a single processor module installed, the plugging order for memory DIMMS is as follows:

- The first CDIMM pair is identical and installed at C16 and C18.
- The next CDIMM pair is identical and installed at C21 and C23.
- The next CDIMM pair is identical and installed at C17 and C19.
- The next CDIMM pair is identical and installed at C20 and C22.

For systems with two processors modules that are installed, the plugging order for memory DIMMS is as follows:

- The first CDIMM pair is identical and installed at C16 and C18.
- The next CDIMM pair is identical and installed at C24 and C26.
- The next CDIMM pair is identical and installed at C21 and C23.
- The next CDIMM pair is identical and installed at C29 and C31.
- The next CDIMM pair is identical and installed at C17 and C19.
- The next CDIMM pair is identical and installed at C25 and C27.
- The next CDIMM pair is identical and installed at C20 and C22.
- The next CDIMM pair is identical and installed at C28 and C30.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000072

2.2.3 Memory bandwidth

The POWER8 processor has exceptional cache, memory, and interconnect bandwidths. The next sections show the bandwidth of the Power S814 and Power S824.

Power S814

Table 2-5 shows the maximum bandwidth estimates for a single core on the Power S814 system.

Table 2-5 Power S814 single core bandwidth estimates

Single core	Power S814	Power S814
	1 core @ 3.02 GHz	1 core @ 3.72 GHz
L1 (data) cache	144.96 GBps	178.56 GBps
L2 cache	144.96 GBps	178.56 GBps
L3 cache	193.28 GBps	238.08 GBps

The bandwidth figures for the caches are calculated as follows:

- L1 cache: In one clock cycle, two 16-byte load operations and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core and the formula is as follows:
  - 3.02 GHz Core: (2 * 16 B + 1 * 16 B) * 3.02 GHz = 144.96 GBps
  - 3.72 GHz Core: (2 * 16 B + 1 * 16 B) * 3.72 GHz = 178.56 GBps
- L2 cache: In one clock cycle, one 32-byte load operation and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core and the formula is as follows:
  - 3.02 GHz Core: (1 * 32 B + 1 * 16 B) * 3.02 GHz = 144.96 GBps
  - 3.72 GHz Core: (1 * 32 B + 1 * 16 B) * 3.72 GHz = 178.56 GBps
- L3 cache: One 32-byte load operation and one 32-byte store operation can be accomplished at half-clock speed and the formula is as follows:
  - 3.02 GHz Core: (1 * 32 B + 1 * 32 B) * 3.02 GHz = 193.28 GBps
  - 3.72 GHz Core: (1 * 32 B + 1 * 32 B) * 3.72 GHz = 238.08 GBps

For the entire Power S814 system populated with one processor module, the overall bandwidths are what is shown in Table 2-6.

Table 2-6 Power S814 total bandwidth estimates

Total bandwidths	Power S814	Power S814
	6 cores @ 3.02 GHz	8 cores @ 3.72 GHz
L1 (data) cache	870 GBps	1,428 GBps
L2 cache	870 GBps	1,428 GBps
L3 cache	1,160 GBps	1,905 GBps
Total Memory	192 GBps	192 GBps
PCIe Interconnect	96 GBps	96 GBps

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000073

Total memory bandwidth: Each POWER8 processor has eight memory channels running at 8 GBps capable of writing 2 bytes and reading 1 byte at a time. The bandwidth formula is calculated as follows:

8 channels * 8 GBps * 3 Bytes = 192 GBps per processor module

PCIe Interconnect: Each POWER8 processor has 48 PCIe lanes running at 8 Gbps full-duplex. The bandwidth formula is calculated as follows:

48 lanes * 1 processor * 8 Gbps * 2 = 96 GBps

Power S824

The Power S824 has different processor modules than the Power S814. Table 2-7 shows the maximum bandwidth estimates for a single core on the Power S824 system.

Table 2-7 Power S824 single core bandwidth estimates

Single core	Power S824	Power S824	Power S824
	1 core @ 3.89 GHz	**1 core @ 4.15 GHz**	**1 core @ 3.52 GHz**
L1 (data) cache	186.72 GBps	199.20 GBps	168.96 GBps
L2 cache	186.72 GBps	199.20 GBps	168.96 GBps
L3 cache	248.96 GBps	265.60 GBps	225.28 GBps

The bandwidth figures for the caches are calculated as follows:

- L1 cache: In one clock cycle, two 16-byte load operations and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core and the formula is as follows:
  - 3.89 GHz Core: (2 * 16 B + 1 * 16 B) * 3.89 GHz = 186.72 GBps
  - 4.15 GHz Core: (2 * 16 B + 1 * 16 B) * 4.15 GHz = 199.20 GBps
  - 3.52 GHz Core: (2 * 16 B + 1 * 16 B) * 3.52 GHz = 168.96 GBps
- L2 cache: In one clock cycle, one 32-byte load operation and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core and the formula is as follows:
  - 3.89 GHz Core: (1 * 32 B + 1 * 16 B) * 3.89 GHz = 186.72 GBps
  - 4.15 GHz Core: (1 * 32 B + 1 * 16 B) * 4.15 GHz = 199.20 GBps
  - 3.52 GHz Core: (2 * 16 B + 1 * 16 B) * 3.52 GHz = 168.96 GBps
- L3 cache: One 32-byte load operation and one 32-byte store operation can be accomplished at half-clock speed and the formula is as follows:
  - 3.89 GHz Core: (1 * 32 B + 1 * 32 B) * 3.89 GHz = 248.96 GBps
  - 4.15 GHz Core: (1 * 32 B + 1 * 32 B) * 4.15 GHz = 265.60 GBps
  - 3.52 GHz Core: (1 * 32 B + 1 * 32 B) * 3.52 GHz = 225.28 GBps

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000074

For the entire Power S824 system populated with two processor modules, the overall bandwidths are shown in Table 2-8:

Table 2-8 Power S824 total bandwidth estimates

Total bandwidths	Power S824	Power S824	Power S824
	12 cores @ 3.89 GHz	**16 cores @ 4.15 GHz**	**24 cores @ 3.52 GHz**
L1 (data) cache	2,241 GBps	3,187 GBps	4,055 GBps
L2 cache	2,241 GBps	3,187 GBps	4,055 GBps
L3 cache	2,988 GBps	4,250 GBps	5,407 GBps
Total Memory	384 GBps	384 GBps	384 GBps
A Bus SMP	25.6 GBps	25.6 GBps	25.6 GBps
X Bus SMP	32 GBps	32 GBps	32 GBps
Total SMP	115 GBps	115 GBps	115 GBps
PCIe Interconnect	192 GBps	192 GBps	192 GBps

Total memory bandwidth: Each POWER8 processor has eight memory channels running at 8 GBps capable of writing 2 bytes and reading 1 byte at a time. The bandwidth formula is calculated as follows:

8 channels * 8 GBps * 3 Bytes = 192 GBps per processor module (384 total for 2 modules)

SMP interconnect: There are two types of buses (see Figure 2-2 on page 43), the X Bus which runs between processors on the same DCM, and the A Bus between processors on different DCMs.

The A Bus is 2-byte, 3-lane (two active lanes and one spare lane) working at 6.4 GHz. The bandwidth formula is calculated as follows:

2 A busses * 2 bytes * 6.4 GHz = 25.6 GBps

The X Buses are rated at 32 GBps.

The formula for the Total SMP Bandwidth is calculated as follows:

2 X Busses + 4 A buses = 115 GBps

PCIe Interconnect: Each POWER8 processor has 48 PCIe lanes running at 8 Gbps full-duplex. The bandwidth formula is calculated as follows:

48 lanes * 2 processors * 8 Gbps * 2 = 192 GBps

Note: There are several POWER8 design points to take into consideration when comparing between hardware designs using SMP communication bandwidths as a unique measurement: POWER8 provides:

- More cores per socket leading to lower inter-CPU communication
- More RAM density (up to 1 TB per socket) allowing for less inter-CPU communication
- Greater RAM bandwidth for less dependence on L3 and L4 cache
- Intelligent hypervisor scheduling that places RAM usage close to CPU
- New SMP routing that so multiple channels are available when congestion occurs

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000075

2.3 System bus

This section provides more information that is related to the internal buses.

The Power S814 and Power S824 systems have internal I/O connectivity through Peripheral Component Interconnect Express Gen3 (PCI Express Gen3, or PCIe Gen3) slots and also external connectivity through SAS adapters.

The internal I/O subsystem on the Power S814 and Power S824 is connected to the PCIe Controllers on a POWER8 processor in the system. Each POWER8 processor module has a bus that has 48 PCIe lanes running at 8 Gbps full-duplex and provides 96 GBps of I/O connectivity to the PCIe slots, SAS internal adapters, and USB ports.

Some PCIe devices are connected directly to the PCIe Gen3 buses on the processors, and other devices are connected to these buses through PCIe Gen3 Switches. The PCIe Gen3 Switches are high-speed devices (512 - 768 GBps each) that allow for the optimal usage of the processors PCIe Gen3 x16 buses by grouping slower x8 or x4 devices that plug in to a x16 slot and not use its full bandwidth. For more information about which slots are connected directly to the processor and which ones are attached to a PCIe Gen3 Switch (referred as PEX), see 2.1, “The IBM POWER8 processor” on page 43.

Figure 2-11 compares the POWER7 and POWER8 I/O buses architecture.

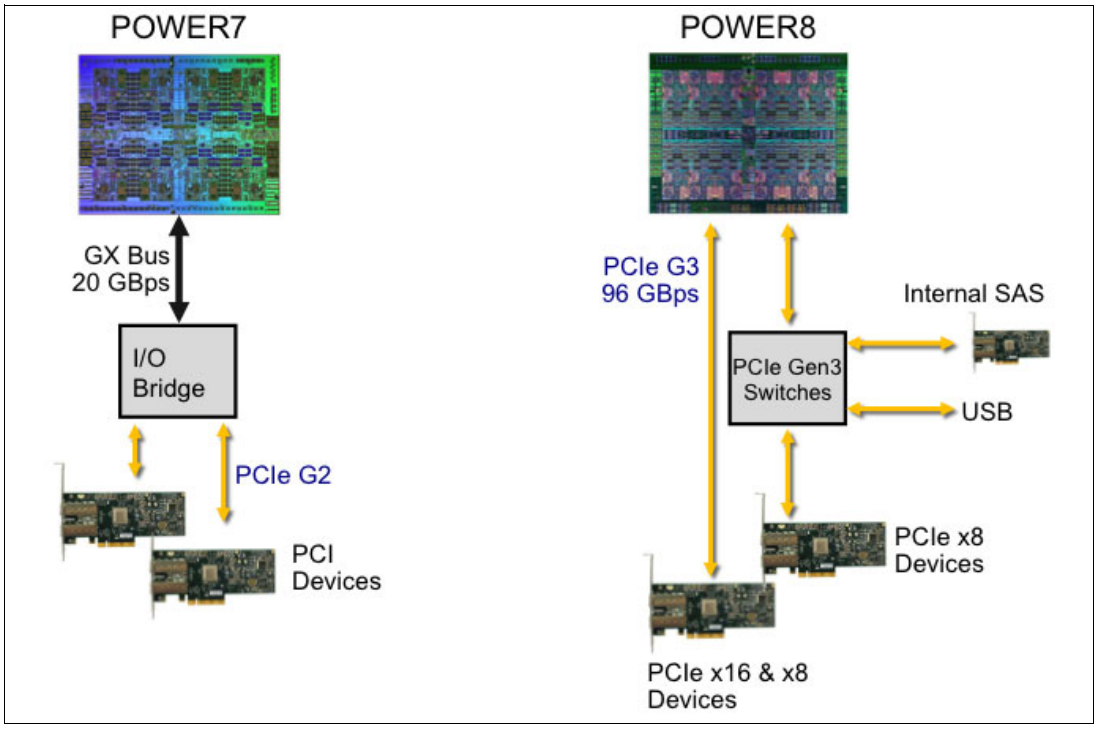

Figure 2-11 Comparison of POWER7 and OWER8 I/O buses architectures

Table 2-9 on page 61 lists the I/O bandwidth of Power S814 and Power S824 processor configurations.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000076

Table 2-9 I/O bandwidth

I/O	I/O bandwidth (maximum theoretical)
Total I/O bandwidth	Power S814 - one processor: ► 48 GBps simplex ► 96 GBps duplex Power S824 - two processors: ► 96 GBps simplex ► 192 GBps duplex

PCIe Interconnect: Each POWER8 processor module has 48 PCIe lanes running at 8 Gbps full-duplex. The bandwidth formula is calculated as follows:

48 lanes * 2 processors * 8 Gbps * 2 = 192 GBps

2.4 Internal I/O subsystem

The internal I/O subsystem is on the system board, which supports PCIe slots. PCIe adapters on the Power S814 and Power S824 are hot-pluggable.

All PCIe slots support Enhanced Error Handling (EEH). PCI EEH-enabled adapters respond to a special data packet that is generated from the affected PCIe 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 most frequently used devices, although certain third-party PCI devices might not provide native EEH support.

2.4.1 Slot configuration

The number of PCIe slots that are available on the Power S814 and Power S824 depends on the storage backplane that is used (#EJ0N or #EJ0P) and the number of installed processors. The amount and speed of the available slots are shown in Table 2-10.

Table 2-10 PCIe slots versus server configuration

	One processor	Two processors
Storage Backplane #EJ0N	Qty 2 - x16 Gen3 low profile slots plus Qty 5 - x8 Gen3 low profile slots	Qty 4 - x16 Gen3 low profile slots plus Qty 7 - x8 Gen3 low profile slots
Storage Backplane #EJ0P	Qty 2 - x16 Gen3 low profile slots plus Qty 4 - x8 Gen3 low profile slots	Qty 4 - x16 Gen3 low profile slots plus Qty 6 - x8 Gen3 low profile slots

Table 2-11 shows the PCIe Gen3 slot configuration for the Power S824.

The slot configuration on the S814 and S842 is the same, except slots P1-C2 to P1-C5 are not available on the S814. These slots are only available to the S842 when the second processor slot is populated.

Table 2-11 Slot configuration of a Power S824

Slot	Description	Location code	Card size	Installed processors required to enable
Slot 1	PCIe Gen3 x8	P1-C2	Full Height, Half Length	2

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000077

Slot	Description	Location code	Card size	Installed processors required to enable
Slot 2	PCIe Gen3 x16	P1-C3	Full Height, Full Length	2
Slot 3	PCIe Gen3 x8	P1-C4	Full Height, Half Length	2
Slot 4	PCIe Gen3 x16	P1-C5	Full Height, Full Length	2
Slot 5	PCIe Gen3 x16	P1-C6	Full Height, Full Length	1
Slot 6	PCIe Gen3 x16	P1-C7	Full Height, Full Length	1
Slot 7	PCIe Gen3 x8	P1-C8	Full Height, Half Length	1
Slot 8	PCIe Gen3 x8	P1-C9	Full Height, Half Length	1
Slot 9[a]	PCIe Gen3 x8	P1-C10	Full Height, Half Length	1
Slot 10[b]	PCIe Gen3 x8	P1-C11	Full Height, Half Length	1
Slot 11	PCIe Gen3 x8	P1-C12	Full Height, Half Length	1

a. Included on all base configurations, this slot (P1-C10) comes populated with a PCIe2 4-port 1 Gb Ethernet Adapter (#5899).

b. Slot 10 (P1-C11) becomes obstructed by external SAS Ports and cannot be used by PCIe devices when #EJ0P is present. If a PCIe adapter is plugged into slot 10, then #EJ0P cannot be used or the adapter must be moved to another suitable slot.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000078

Figure 2-12 and Figure 2-13 show the top view diagrams of the available PCIe slots on the server with the available backplane options.

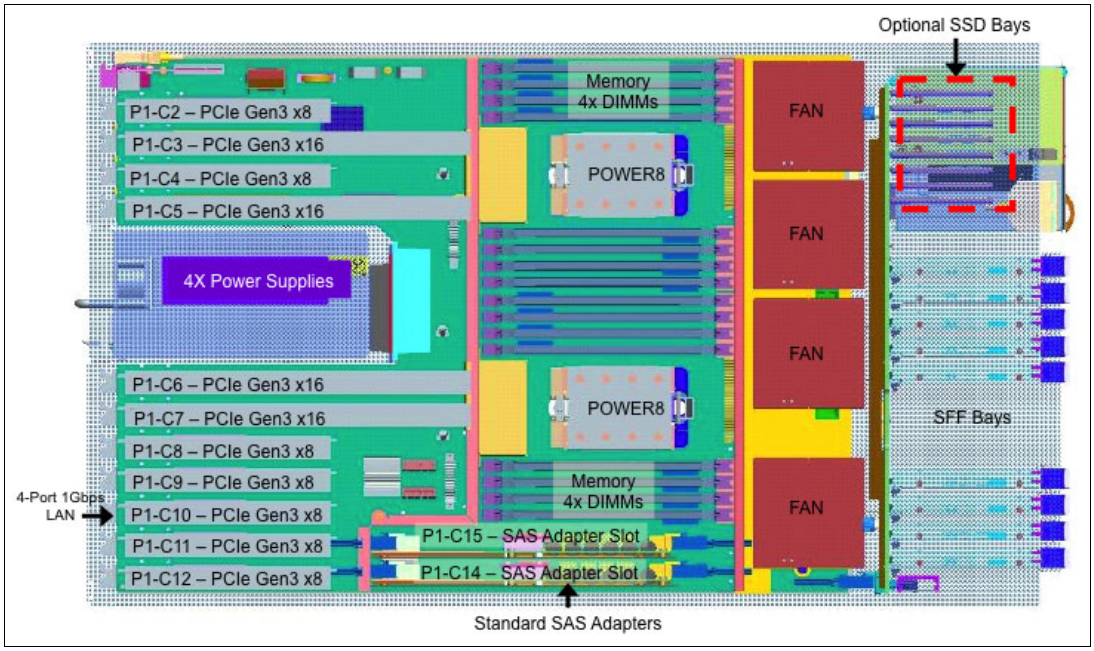

Figure 2-12 Top view diagram for Power S824 with #EJ0N backplane option

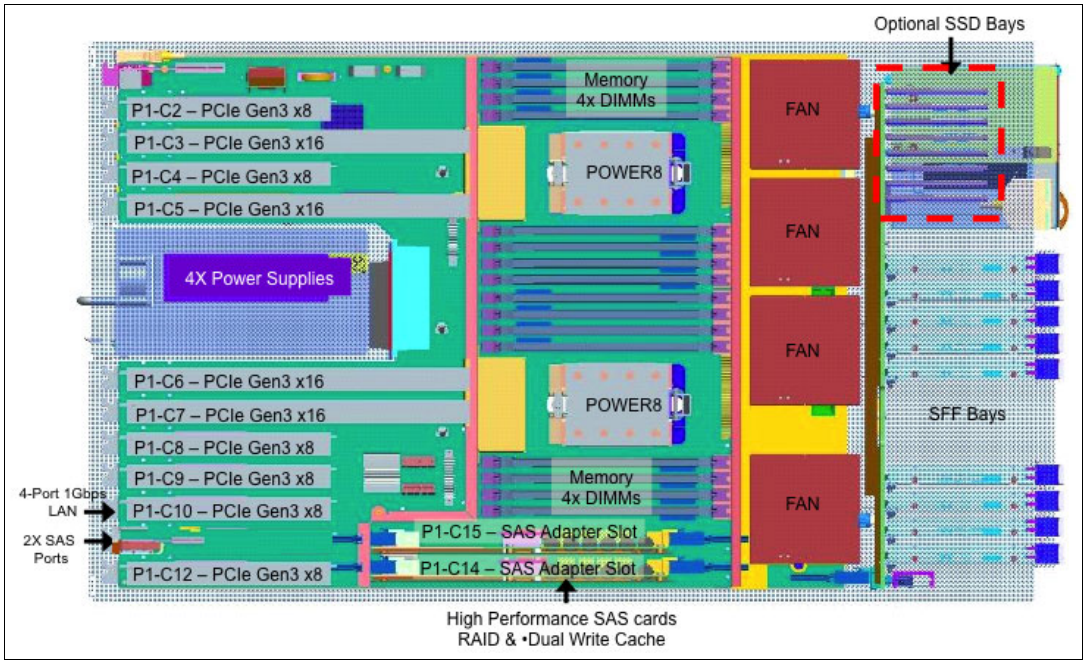

Figure 2-13 Top view diagram for Power S824 with #EJ0P backplane option

At least one PCIe Ethernet adapter is required on the server and therefore one of the x8 PCIe slots is used for this required adapter, which is identified as the P1-C10 slot. Included on all base configurations, as of the time of writing, this adapter is the PCIe2 4-port 1 Gb Ethernet Adapter (#5899). It can be replaced or moved in field or as part of an upgrade if the server still contains at least one Ethernet adapter.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000079

Remember: Slot 9 (P1-C10) comes with PCIe2 4-port 1 Gb Ethernet Adapter (#5899) installed. In case this slot is needed, the adapter must be moved in the field to another suitable slot or replaced by another Ethernet adapter.

2.4.2 System ports

The system board has one serial port that is called a *system port*. When an HMC is connected to the server, the integrated system port of the server is rendered non-functional. In this case, you must install an asynchronous adapter, which is described in Table 2-18 on page 70, for serial port usage:

- Integrated system port is not supported under AIX or Linux when the HMC ports are connected to an HMC. Either the HMC ports or the integrated system port can be used, but not both.
- The integrated system port is 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 integrated system port, including basic port settings (baud rate and so on), modem selection, and Call Home and call-in policy, can be performed with the Advanced Systems Management Interface (ASMI).

Remember: The integrated console/modem port usage that is described is for systems that are configured as a single, system-wide partition. When the system 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.

2.5 PCI adapters

This section covers the various types and functions of the PCI adapters that are supported by the Power S814 and Power S824 servers.

2.5.1 PCI Express

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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000080

The PCIe interfaces that are supported on this server are PCIe Gen3, which are capable of 16 GBps simplex (32 GBps duplex) on a single x16 interface. PCIe Gen3 slots also support previous generations (Gen2 and Gen1) adapters, which operate at lower speeds, according to the following rules:

- Place x1, x4, x8, and x16 speed adapters in the same size connector slots first, before mixing adapter speed with connector slot size.
- Adapters with smaller speeds are allowed in larger sized PCIe connectors but larger speed adapters are not compatible in smaller connector sizes (that is, a x16 adapter cannot go in an x8 PCIe slot connector).

All adapters support Enhanced Error Handling (EEH). PCIe adapters use a different type of slot than PCI adapters. If you attempt to force an adapter into the wrong type of slot, you might damage the adapter or the slot.

IBM POWER8 processor-based servers can support two different form factors of PCIe adapters:

- PCIe low profile (LP) cards, which are used with the Power S822 PCIe slots. These cards are not compatible with Power S814 and Power S824 servers because of its low height, but have its similar cards in another form factors.
- PCIe full height and full high cards are not compatible with Power S822 and are designed for the following servers:
  - Power S814
  - Power S824

Before adding or rearranging adapters, use the System Planning Tool to validate the new adapter configuration. For more information, 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 that is 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 describe the supported adapters and provide tables of orderable feature numbers. The tables indicate operating system support (AIX, IBM i, and Linux) for each of the adapters.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000081

2.5.2 LAN adapters

To connect the Power S814 and Power S824 servers to a local area network (LAN), you can use the LAN adapters that are supported in the PCIe slots of the system unit. Table 2-12 lists the available LAN adapters.

Table 2-12 Available LAN adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
5767	5767	2-Port 10/100/1000 Base-TX Ethernet PCI Express Adapter	6	10	AIX, IBM i, Linux
5899	576F	PCIe2 4-port 1 GbE Adapter	6	10	AIX, IBM i, Linux
EC2N		PCIe3 2-port 10 GbE NIC&RoCE SR Adapter	6	30	AIX, IBM i, Linux
EC33		PCIe3 2-port 56Gb FDR IB Adapter x16	6	4	AIX, IBM i, Linux
EC38		PCIe3 2-port 10 GbE NIC&RoCE SFP+ Copper Adapter	2	30	AIX, IBM i, Linux
EC3B	57B6	PCIe3 2-Port 40 GbE NIC RoCE QSFP+ Adapter	16	10	AIX, IBM i, Linux
EC3M		PCIe3 2-port 100GbE (NIC&RoCE) QSFP28 Adapter x16	2	4	AIX, IBM i, Linux
EN0S	2CC3	PCIe2 4-Port (10Gb+1 GbE) SR+RJ45 Adapter	16	10	AIX, IBM i, Linux
EN0U	2CC3	PCIe2 4-port (10Gb+1 GbE) Copper SFP+RJ45 Adapter	6	10	AIX, IBM i, Linux
EN0W	2CC4	PCIe2 2-port 10/1 GbE BaseT RJ45 Adapter	6	10	AIX, IBM i, Linux
EN15	2CE3	PCIe3 4-port 10 GbE SR Adapter	6	10	AIX, IBM i, Linux
EN17	2CE4	PCIe3 4-port 10 GbE SFP+ Copper Adapter	6	10	AIX, IBM i, Linux

Note: Adapters #EN15 and #EN17 support SR-IOV when minimum firmware and software levels are met. See 3.4, “Single Root I/O Virtualization” on page 113 for more information.

2.5.3 Graphics accelerator adapters

Table 2-13 lists the available graphics accelerator adapters. An adapter can be configured to operate in either 8-bit or 24-bit color modes. The adapter supports both analog and digital monitors.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000082

Table 2-13 Available graphics accelerator adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
5748	5748	POWER GXT145 PCI Express Graphics Accelerator	3	7	AIX, Linux
EC42		PCIe2 3D Graphics Adapter x1	6	10	Linux

2.5.4 SAS adapters

Table 2-14 lists the SAS adapters that are available for Power S814 and Power S824 systems.

Table 2-14 Available SAS adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
5901	57B3	PCIe Dual-x4 SAS Adapter	12	24	AIX, IBM i ,Linux
5913	57B5	PCIe2 1.8 GB Cache RAID SAS Adapter Tri-port 6Gb	6	10	AIX, IBM i, Linux
EJ0J	57B4	PCIe3 RAID SAS Adapter Quad-port 6Gb x8	6	10	AIX, IBM i ,Linux
EJ0L	57CE	PCIe3 12 GB Cache RAID SAS Adapter Quad-port 6Gb x8	6	10	AIX, IBM i ,Linux
EJ10	57B4	PCIe3 SAS Tape/DVD Adapter Quad-port 6Gb x8	6	10	AIX, IBM i, Linux
EJ14	57B1	PCIe3 12 GB Cache RAID PLUS SAS Adapter Quad-port 6Gb x8	12	22	AIX IBM i, Linux
EJ1P		PCIe1 SAS Tape/DVD Dual-port 3Gb x8 Adapter	8	8	AIX IBM i, Linux

2.5.5 Fibre Channel adapters

The systems support direct or SAN connection to devices that use Fibre Channel adapters. Table 2-15 on page 68 summarizes the available Fibre Channel adapters, which all have LC connectors.

If you are attaching a device or switch with an SC type fiber connector, then an LC-SC 50 Micron Fiber Converter Cable (#2456) or an LC-SC 62.5 Micron Fiber Converter Cable (#2459) is required.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000083

Table 2-15 Available Fibre Channel adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
5729	5729	PCIe2 8 Gb 4-port Fibre Channel Adapter	6	10	AIX, IBM i, Linux
5735	577D	8 Gigabit PCI Express Dual Port Fibre Channel Adapter	6	10	AIX, IBM i, Linux
5774	5774	4 Gigabit PCI Express Dual Port Fibre Channel Adapter	6	10	AIX, IBM i, Linux
EN0A	577F	PCIe2 16 Gb 2-port Fibre Channel Adapter	6	10	Linux
EN0G		PCIe2 8Gb 2-Port Fibre Channel Adapter	16	30	AIX, Linux, IBM i with VIOS
EN12		PCIe2 8Gb 4-port Fibre Channel Adapter	6	10	AIX, Linux, IBM i with VIOS

NPIV: The usage of N_Port ID Virtualization (NPIV) through the Virtual I/O Server (VIOS) requires an NPIV-capable Fibre Channel adapter, such as the #5729, #5735, and #EN0A.

2.5.6 Fibre Channel over Ethernet

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

Figure 2-14 compares existing Fibre Channel and network connections and FCoE connections.

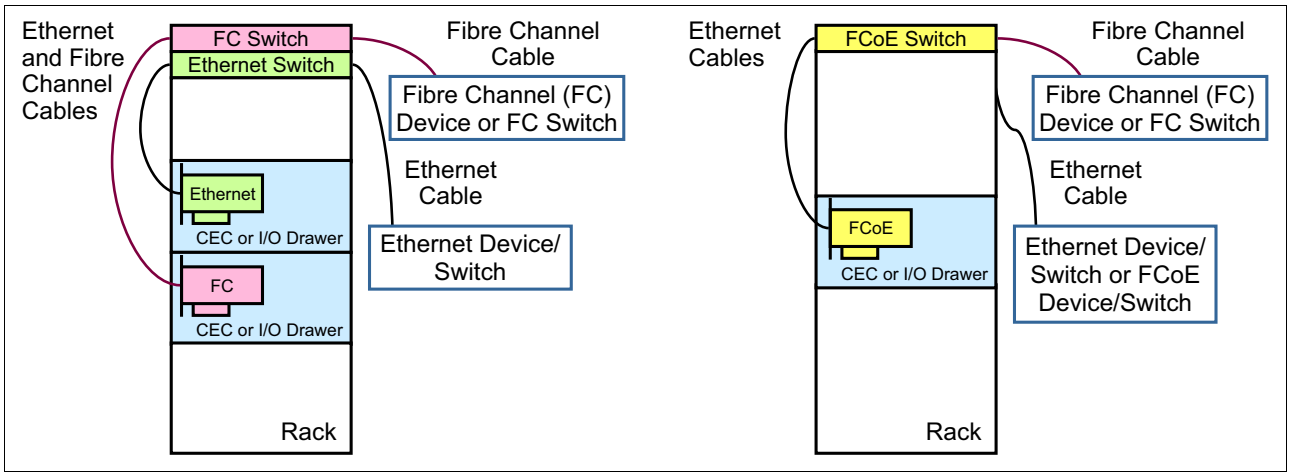

Figure 2-14 Comparison between an existing Fibre Channel and network connection and FCoE connection

Table 2-16 lists the available Fibre Channel over Ethernet adapters. They are high-performance, Converged Network Adapters (CNAs) using SR optics. Each port can simultaneously provide network interface card (NIC) traffic and Fibre Channel functions.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000084

Table 2-16 Available FCoE adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
EN0H	2B93	PCIe2 4-port (10 Gb FCoE & 1 GbE) SR&RJ45	6	10	AIX, IBM i through VIOS, Linux
EN0K		PCIe2 4-port (10 Gb FCoE & 1 GbE) SFP+Copper&RJ45	6	10	AIX, IBM i through VIOS, Linux

Note: Adapters #EN0H, #EN0K, #EN0M support SR-IOV when minimum firmware and software levels are met. See 3.4, “Single Root I/O Virtualization” on page 113 for more information.

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

2.5.7 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 that are necessary for present and future server systems with levels better than can be achieved by using bus-oriented I/O structures.

InfiniBand (IB) is an open set of interconnect standards and specifications. The main IB specification is published by the IBTA and is available at the following website:

http://www.infinibandta.org/

IB 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), which are used primarily in servers, or target channel adapters (TCAs), which are used primarily in storage subsystems.

The IB physical connection consists of multiple byte lanes. Each individual byte lane is a four-wire, 2.5, 5.0, or 10.0 Gbps bidirectional connection. Combinations of link width and byte lane speed allow for overall link speeds of 2.5 - 120 Gbps. The architecture defines a layered hardware protocol and also 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.

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

A connection to supported IB switches is accomplished by using the QDR optical cables #3290 and #3293.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000085

Table 2-17 lists the available IB adapter.

Table 2-17 Available IB adapter

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
EC3F	2CEA	PCIe3 2-port 100Gb EDR IB Adapter x16	2	4	Linux
EC3U	2CEA	PCIe3 1-port 100Gb EDR IB Adapter x16	2	4	Linux

2.5.8 Asynchronous and USB adapters

Asynchronous PCIe adapters provide the 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 that are listed in Table 2-18.

Table 2-18 Available asynchronous and USB adapters

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
EC46		PCIe2 LP 4-Port USB 3.0 Adapter	16	30	AIX, Linux
EN29		2 Port Async EIA-232 PCIe Adapter	10		IBM i

2.5.9 Cryptographic coprocessor

The cryptographic coprocessor card that is supported for the Power S814 and Power S824 is shown in Table 2-19.

Table 2-19 Available cryptographic coprocessor

Feature code	CCIN	Description	Maximum with one processor installed	Maximum with two processors installed	OS support
EJ32		PCIe3 Crypto Coprocessor no BSC 4767	6	10	AIX, IBM i, Linux
EC33		PCIe3 Crypto Coprocessor BSC-Gen3 4767	10	19	AIX, IBM i, Linux

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000086

2.5.10 Field-programmable gate array adapters

The field-programmable gate array (FPGA) adapters are PCIe adapters that are based on a semiconductor device that can be programmed.

Unlike an ASIC, which is designed and programmed to perform a single function, an FPGA can be programmed to run different product functions, adapt to new standards, and reconfigure its hardware for specific applications even after the product is installed in the field. An FPGA can implement any logical function that an ASIC can perform, but can have its code updated to include more functions or perform a different role.

Today, FPGAs have logic elements that can be reprogrammed, and include SRAM memory, high-speed interconnects that can range up to 400 Gbps, logic blocks, and routing. FPGAs allow for versatile solutions because a single adapter can perform several distinct functions depending on the code that is deployed on it.

By having a highly optimized software stack, the FPGAs can act as coprocessor for the server CPUs, running repetitive and complex functions at a fraction of the time and power, while allowing for the server CPUs to perform other functions at the same time.

The FPGA adapter that is supported for the Power S814 and Power S824 is shown in Table 2-20.

Table 2-20 Available FPGA adapter

Feature code	CCIN	Description	One Processor	Two Processors	OS support
EJ12	59AB	PCIe3 FPGA Accelerator Adapter	1	1	AIX, Linux, VIOS is not supported.

The #EJ12 is an FPGA adapter that is based on Altera Stratix V 28 nm hardware. In addition to its logic components, it has 8 GB DDR3 RAM.

The initial implementation for the #EJ12 is an adapter that can perform gzip compression. The zlib API, a software library that is responsible for data compression and is used in several software packages, can move the tasks directly to the FPGA, which allows increased compression performance, increased compression rates, and decreased CPU usage. Java 7.1 is already enabled to take advantage of this acceleration.

Other applications, such as big data, can take advantage of this approach to compression after the compression rates are higher than the ones that are achieved through software. The compression processes complete in a shorter time by allowing for more data density, disk savings, and faster data analysis.

2.5.11 CAPI adapters

The available CAPI adapters are shown in Table 2-21.

Table 2-21 Available CAPI adapters

Feature code	CCIN	Description	One Processor	Two Processors	OS support
EJ17		PCIe3 CAPI FlashSystem Accelerator Adapter	1	1	AIX

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000087

#EJ17 is a PCIe adapter with accelerator FPGA for low latency connection using CAPI (Coherent Accelerator Processor Interface). The adapter has two 8Gb optical SR fiber connections for attachment to FlashSystem Drawer.

The adapter must be placed in a x16 slot in the system unit which is CAPI enabled. The server must have CAPI enablement feature.

2.5.12 Flash storage adapters

The available flash storage adapters are shown in Table 2-22.

Table 2-22 Available flash storage adapters

Feature code	CCIN	Description	One Processor	Two Processors	OS support
EC55		PCIe3 1.6TB NVMe Flash Adapter	6	8	Linux
EC57		PCIe3 3.2TB NVMe Flash Adapter	6	8	Linux

2.6 Internal storage

The internal storage on the Power S814 and Power S824 servers depends on the DASD/Media backplane that is used. The servers support two DASD/Media backplanes: #EJ0N and #EJ0P.

The #EJ0N storage backplane contains the following items:

- A storage backplane for twelve 2.5-inch SFF Gen3 HDDs or SSDs.
- One SAS disk controller that is capable of RAID 0, RAID 5, RAID 6, and RAID 10, placed in a dedicated SAS controller slot (P1-C14).
- The optional split backplane feature #EJ0S adds a secondary SAS disk controller and allows DASDs to be split into two groups of six (6+6). This secondary controller is placed in the second dedicated SAS controller slot (P1-C15).

The #EJ0P storage backplane contains the following items:

- A storage backplane for eighteen 2.5-inch SFF Gen3 HDDs or SSDs.
- Two active-active SAS disk controllers that are capable of RAID 0, RAID 5, RAID 6, RAID 10, RAID 5T2, RAID 6T2, and RAID 10T2, which can be placed in dedicated SAS controller slots P1-C14 and P1-C15.
- The #EJTM SSD Module Cage, which is automatically added by e-config when #EJ0P is ordered, adds a secondary disk cage with eight 1.8-inch SSD module bays.
- Two external SAS ports for DASD drawers connectivity (through slot P1-C11) support one EXP24S SFF Gen2-bay Drawer (#5887).
- The storage split backplane function is not supported.

Table 2-23 presents a summarized view of these features.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000088

Table 2-23 Backplane options and summary of features

Feature	#EJ0N backplane	#EJ0P backplane
Supported RAID types	JBOD, RAID 0, RAID 5, RAID 6, and RAID 10	JBOD, RAID 0, RAID 5, RAID 6, RAID 10, RAID 5T2, RAID 6T2, and RAID 10T2
Disk bays	12 SFF Gen3 (HDDs/SSDs)	18 SFF Gen3 (HDDs/SSDs)
SAS controllers	Single	Dual active-active
Easy Tier capable controllers	No	Yes
External SAS ports	No	Yes, two SAS ports
Split backplane	Optional (#EJ0S) includes secondary SAS controller	No
SSD 1.8-inch module cage	No	Required (#EJTM)
PCIe slot P1-C11	Available for PCIe x8 cards	Used by external SAS ports

The 2.5-inch or small form factor (SFF) SAS bays can contain SAS drives (HDD or SSD) that are mounted on a Gen3 tray or carrier (also knows as SFF-3). SFF-1 or SFF-2 drives do not fit in an SFF-3 bay. All SFF-3 bays support concurrent maintenance or hot-plug capability.

Additionally, as an option for the #EJ0P backplane, the feature #EJTM adds an 8-bay 1.8-inch SSD Cage behind the server bezel. All eight bays are accessed by both of the SAS controllers and the bays support concurrent maintenance (hot-plug). At the time of writing, the supported disks for these bays are feature codes #ES0J and #ES0K, 1.8-inch SSD with 387 GB capacity.

The SSF-3 disk drives connect to the DASD backplane and are hot-swap and front-accessible. Optionally, the 1.8-inch SSD driver is housed behind the server bezel, but also is hot-swap after you remove the front bezel.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000089

The server front views (with both backplane options) are shown in Figure 2-15 and Figure 2-16.

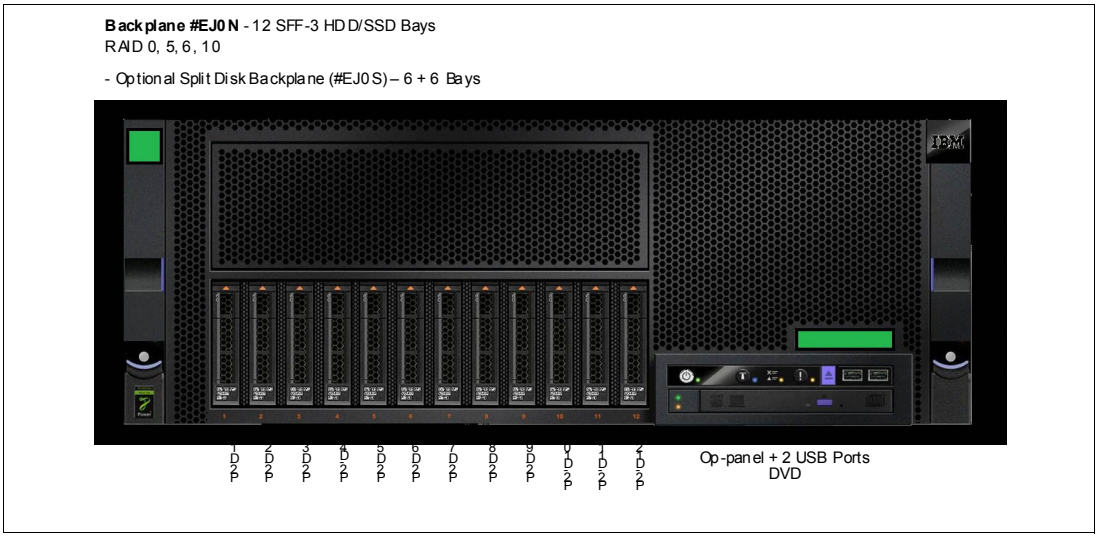

Figure 2-15 Server front view with #EJ0N backplane option

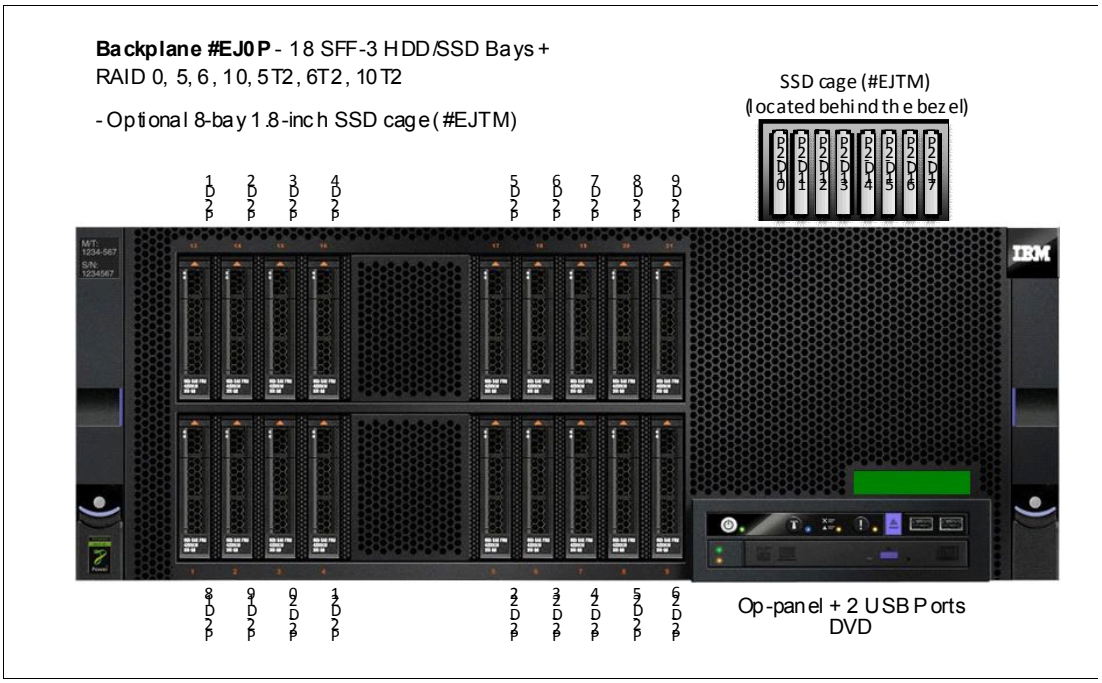

Figure 2-16 Server front view with #EJ0P backplane option

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000090

The internal connections to the physical disks are shown in Figure 2-17, Figure 2-18 on page 76, and Figure 2-19 on page 77.

Figure 2-17 shows the internal connections when using #EJ0N.

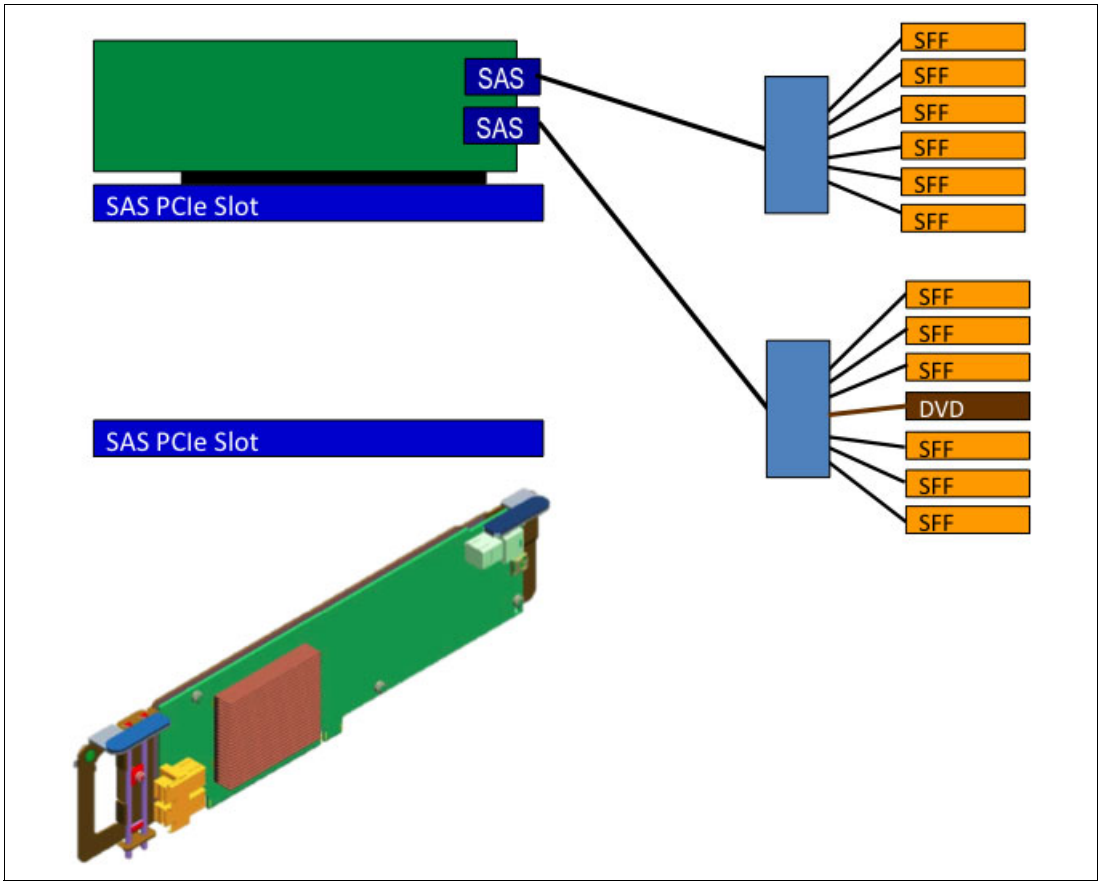

Figure 2-17 Internal topology overview for the #EJ0N DASD backplane

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000091

Figure 2-18 shows the internal topology overview for the #EJ0N backplane in a split backplane configuration (optional #EJ0S).

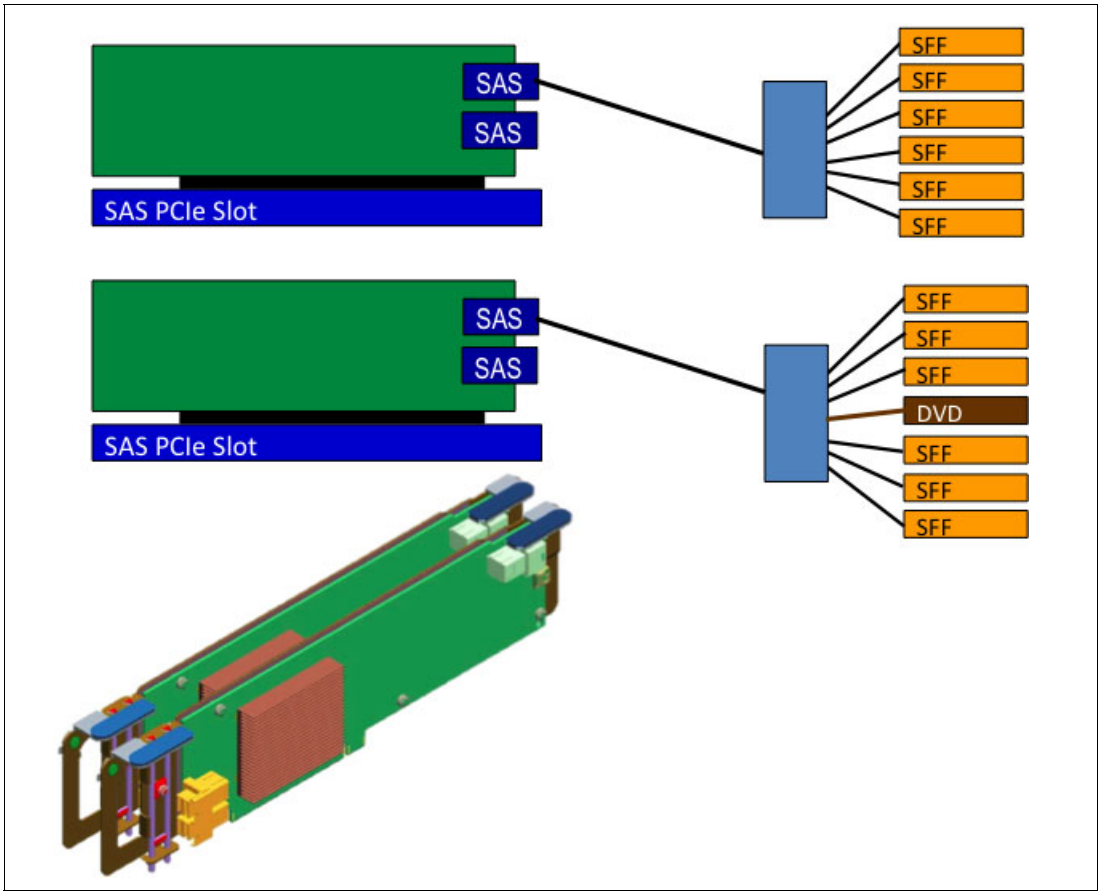

Figure 2-18 Internal topology overview for the #EJ0N DASD backplane with the #EJ0S split backplane feature

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000092

Figure 2-19 shows the details of the internal topology overview for the #EJ0P DASD backplane with the optional #EJTM 1.8-inch SSD Cage.

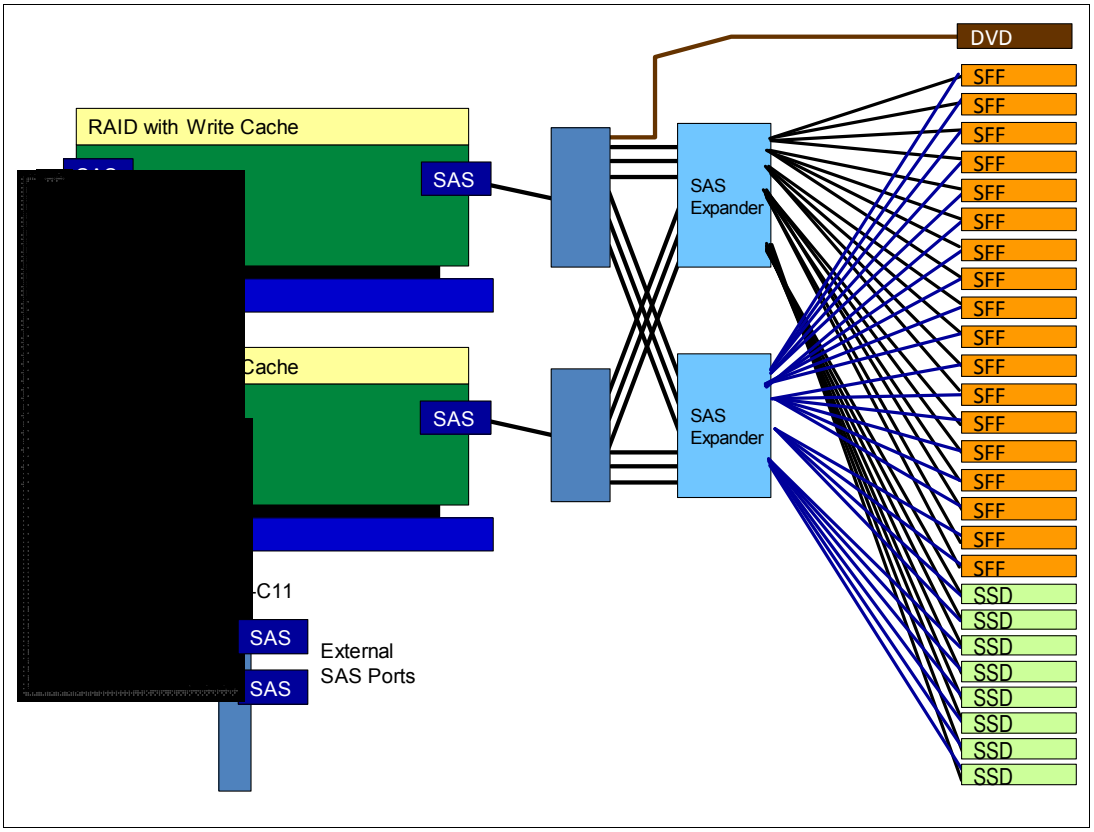

Figure 2-19 Internal topology overview for the #EJ0P DASD backplane

The external SAS ports that are provided by the #EJ0P support the attachment of a single EXP24S SFF Gen2-bay Drawer (#5887). Additional disk drawers can be attached by using supported PCIe SAS adapters.

2.6.1 RAID support

There are multiple protection options for HDD/SSD drives in the Power S814 and Power S824 systems, whether they are contained in the SAS SFF bays in the system unit or drives in disk-only I/O drawers. Although protecting drives is always recommended, AIX/Linux users can choose to leave a few or all drives unprotected at their own risk, and IBM supports these configurations.

Drive protection

HDD/SSD drive protection can be provided by AIX, IBM i, and Linux, or by the HDD/SSD hardware controllers.

The default storage backplane #EJ0N contains one SAS HDD/SSD controller and provides support for JBOD and RAID 0, 5, 6, and 10 for AIX or Linux. A secondary non-redundant controller is added when using split backplane feature #EJ0S, so each of the six disks have a separated disk controller.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000093

When you choose the optional #EJ0P storage backplane, the controller is replaced by a pair of High Performance RAID controllers with dual integrated SAS controllers with 1.8 GB of physical write cache. High performance controllers run 18 SFF-3 SAS bays, 1.8-inch SSD cage bays (only available on the Power S824), and a DVD bay in the system unit. Dual controllers (also called dual I/O adapters or paired controllers) and their write cache are placed in integrated slots and do not use PCIe slots. However, cache power protection hardware covers one PCIe x8 slot (P1-C11). Patented active/active configurations with at least two arrays are supported.

The write cache, which is responsible for increasing write performance by caching data before it is written to the physical disks, can have its data compression capabilities activated, providing up to 7.2 GB effective cache capacity. The write cache contents are protected against power loss by flash memory and super capacitors, which removes the need for battery maintenance.

The high performance SAS controllers provide RAID 0, RAID 5, RAID 6 and RAID 10 support, and its EasyTier variants (RAID 5T2, RAID 6T2, and RAID 10T2) if the server has both HDDs and SSDs installed.

The Easy Tier function is supported, so the dual controllers can automatically move hot data to an attached SSD and cold data to an attached HDD for AIX, Linux, and VIOS environments. If an EXP 24S SFF Gen2-bay Drawer (#5887) is attached to the adapters, the Easy Tier function is extended to the disks on this drawer. To learn more about Easy Tier, see 2.6.2, “Easy Tier” on page 79.

Table 2-24 lists the RAID support configurations by the storage backplane options.

Table 2-24 RAID support configurations

Storage backplane	JBOD	RAID 0, 5, 6, and 10	RAID 0, 5, 6, and 10 and Easy Tier (RAID 5T2, 6T2, and 10T2)	Split back-plane	External SAS port
#EJ0N	Yes	Yes	No	Optional	No
#EJ0P	Yes	Yes	Yes	No	Yes

AIX and Linux can use disk drives that are formatted with 512-byte blocks when they are 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 error-correcting code (ECC) and bad block detection, is gained in this reformatting. For example, a 300 GB disk drive, when reformatted, provides approximately 283 GB. IBM i always uses drives that are formatted to 528 bytes. SSDs are always formatted with 528 byte sectors.

Supported RAID functions

The base hardware supports RAID 0, 5, 6, and 10. When additional features are configured, the server supports hardware RAID 0, 5, 6, 10, 5T2, 6T2, and 10T2:

- 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 version of RAID increases I/O bandwidth by simultaneously accessing multiple data paths.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000094

- 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 that is separate from the ones that are used to store the data itself. This version of RAID provides data resiliency if a single drive fails in a RAID 5 array.

- 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. A RAID 6 configuration requires N+2 drives to accommodate the additional parity data, making it less cost-effective than RAID 5 for equivalent storage capacity. This version of RAID provides data resiliency if one or two drives fail in a RAID 6 array. When you work with large capacity disks, RAID 6 allows you to sustain data parity during the rebuild process.

- RAID 10 is 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 two-disk array for performance benefits. A duplicate of the first stripe set is then mirrored on another two-disk array for fault tolerance. This version of RAID provides data resiliency if a single drive fails, and it can provide resiliency for multiple drive failures.

RAID 5T2, RAID 6T2, and RAID 10T2 are RAID levels with EasyTier enabled. It requires that both types of disks exist on the system under the same controller (HDDs and SSDs) and that both are configured under the same RAID type.

2.6.2 Easy Tier

With the standard backplane (#EJ0N), the server can handle both HDDs and SSDs that are attached to its storage backplane if they are on separate arrays.

The high function backplane (#EJ0P) can handle both types of storage in two different ways:

- Separate arrays: SSDs and HDDs coexist on separate arrays, just like the Standard SAS adapter can.
- Easy Tier: SSDs and HDDs coexist under the same array.

When the SDDs and HDDS are under the same array, the adapter can automatically move the most accessed data to faster storage (SSDs) and less accessed data to slower storage (HDDs). This is called Easy Tier.

There is no need for coding or software intervention after the RAID is configured correctly. Statistics on block accesses are gathered every minute, and after the adapter realizes that some portion of the data is being frequently requested, it moves this data to faster devices. The data is moved in chunks of 1 MB or 2 MB called *bands*.

From the operating system point-of-view, there is just a regular array disk. From the SAS controller point-of-view, there are two arrays with parts of the data being serviced by one tier of disks and parts by another tier of disks.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000095

Figure 2-20 shows an Easy Tier array.

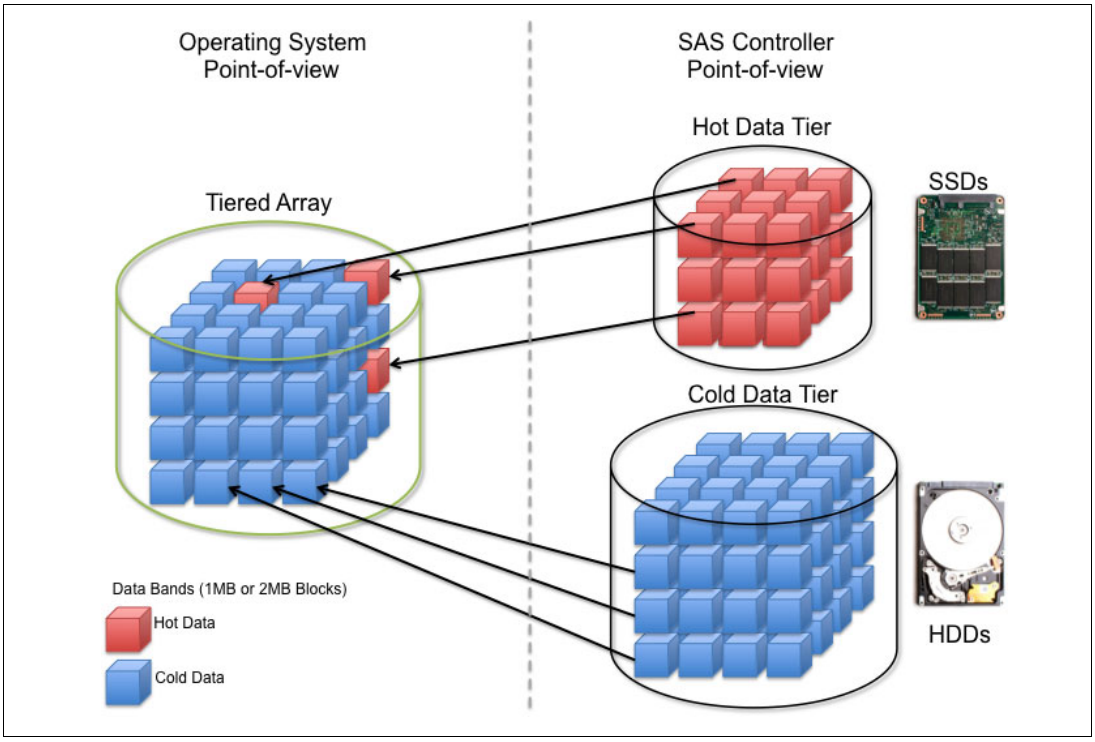

Figure 2-20 Easy Tier

The Easy Tier configuration is accomplished through a standard operating system SAS adapter configuration utility. On Figure 2-21 and Figure 2-22 on page 81, there are two examples of tiered array creation for AIX.

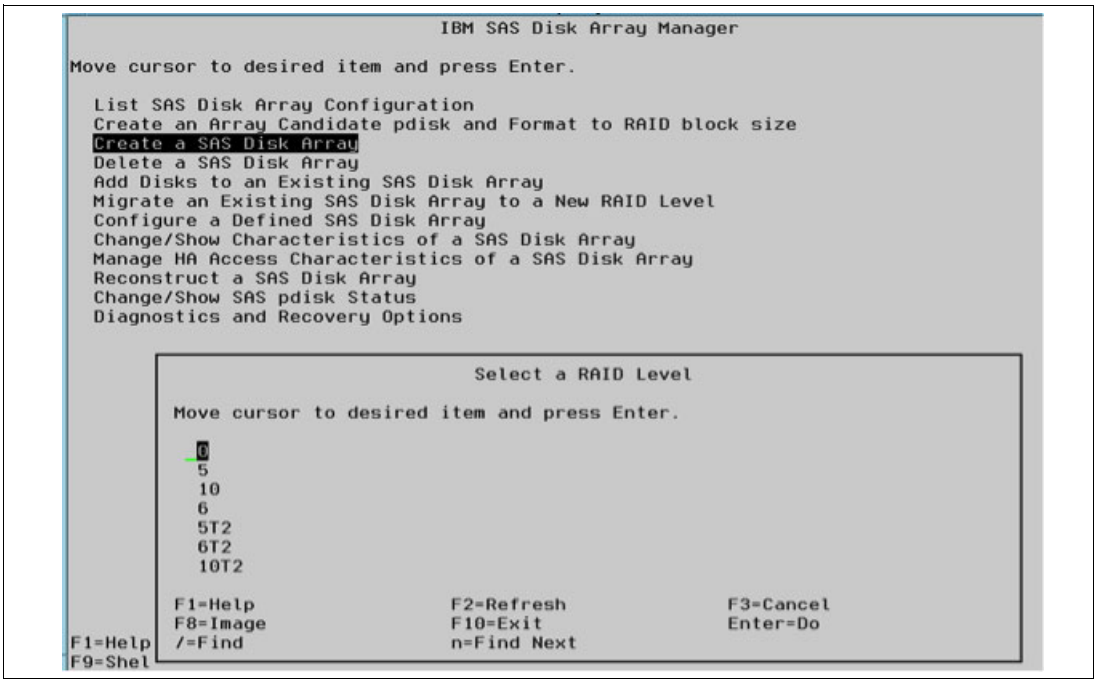

Figure 2-21 Array type selection panel on AIX RAID Manager

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000096

```
-----------------------------------------------------------------------
Name      Resource  State       Description              Size
-----------------------------------------------------------------------
sissas1   FEFFFFFF  Primary     PCIe3 12GB Cache RAID SAS Adapter Quad-port 6Gb x8
sissas0   FEFFFFFF  HA Linked   Remote adapter SN  00325001

hdisk1    FC0000FF  Optimal     RAID 5T2 Array (N/N)   773.5GB  <-- RAID 5T2
 pdisk0   000400FF  Active      Array Member           139.6GB
 pdisk1   000401FF  Active      Array Member           139.6GB
 pdisk2   000402FF  Active      Array Member           139.6GB
 pdisk3   000403FF  Active      Array Member           139.6GB
 pdisk7   000407FF  Active      SSD Array Member       177.8GB
 pdisk6   000406FF  Active      SSD Array Member       177.8GB
 pdisk8   000408FF  Active      SSD Array Member       177.8GB

hdisk2    FC0100FF  Optimal     RAID 6T2 Array (N/N)    1090GB  <-- RAID 6T2
 pdisk10  00040AFF  Active      SSD Array Member       387.9GB
 pdisk11  00040BFF  Active      SSD Array Member       387.9GB
 pdisk4   000404FF  Active      Array Member           139.6GB
 pdisk20  000414FF  Active      SSD Array Member       387.9GB
 pdisk21  000415FF  Active      SSD Array Member       387.9GB
 pdisk9   000409FF  Active      SSD Array Member       177.8GB
 pdisk5   000405FF  Active      Array Member           139.6GB
 pdisk12  00040CFF  Active      Array Member           139.6GB
 pdisk13  00040DFF  Active      Array Member           139.6GB
 pdisk14  00040EFF  Active      Array Member           139.6GB
 pdisk15  00040FFF  Active      Array Member           139.6GB

hdisk3    FC0200FF  Optimal     RAID 10T2 Array (O/O)  666.6GB  <-- RAID 10T2
 pdisk22  000416FF  Active      SSD Array Member       387.9GB
 pdisk23  000417FF  Active      SSD Array Member       387.9GB
 pdisk16  000410FF  Active      Array Member           139.6GB
 pdisk17  000411FF  Active      Array Member           139.6GB
 pdisk18  000412FF  Active      Array Member           139.6GB
 pdisk19  000413FF  Active      Array Member           139.6GB
```

RAID 5 SSD
- 2 +1 x 177.8 GB
RAID 5 HDD
- 3 + 1 x 139.6 GB

RAID 6 SSD
- 3 + 2 x 387.9 GB
RAID 6 HDD
- 4 + 2 x 139.6 GB

RAID 10 SSD
- 1 + 1 x 387.9 GB
RAID 10 HDD
- 2 + 2 x 139.6 GB

Figure 2-22 Tiered arrays (RAID 5T2, RAID 6T2 and RAID 10T2) example on AIX RAID Manager

To support Easy Tier, make sure that the server is running at least the following minimum levels:

- VIOS 2.2.3.3 with interim fix IV56366 or later
- AIX 7.1 TL3 SP3 or later
- AIX 6.1 TL9 SP3 or later
- RHEL 6.5 or later
- SLES 11 SP3 or later

IBM i supports Easy Tier only through VIOS.

2.6.3 External SAS port

The Power S814 and Power S824 DASD backplane (#EJ0P) offers a connection to an external SAS port:

- The SAS port connector is on slot P1-C11.
- The external SAS port is used for expansion to one external SAS EXP24S SFF Gen2-bay Drawer (#5887)

Additional drawers and the IBM System Storage 7226 Tape and DVD Enclosure Express (Model 1U3) can be attached by installing additional SAS adapters.

Note: Only one SAS drawer is supported from the external SAS port. Additional SAS drawers can be supported through SAS adapters.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000097

2.6.4 Media bays

Included in the EJ0N or EJ0P backplanes is a slimline media bay that can optionally house a SATA DVD-RAM (#5771). Direct dock and hot-plug of the DVD media device is supported.

The DVD drive and media device do not have an independent SAS adapter and cannot be assigned to an LPAR independently of the HDD/SSDs in the system.

2.7 External I/O subsystems

This section describes the PCIe Gen3 I/O expansion drawer that can be attached to the Power S814 and Power S824.

2.7.1 PCIe Gen3 I/O expansion drawer

The PCIe Gen3 I/O expansion (EMX0) drawer is a 4U high, PCI Gen3-based and rack mountable I/O drawer. It offers two PCIe Fan Out Modules (#EMXF) each of them providing six PCIe slots.

The physical dimensions of the drawer are 444.5 mm (17.5 in.) wide by 177.8 mm (7.0 in.) high by 736.6 mm (29.0 in.) deep for use in a 19-inch rack.

A PCIe x16 to Optical CXP converter adapter (#EJ07) and 3.0 m (#ECC7), 10.0 m (#ECC8) CXP 16X Active Optical cables (AOC) connect the system node to a PCIe Fan Out module in the I/O expansion drawer. One feature #ECC7, one #ECC8 ships two AOC cables.

Concurrent repair and add/removal of PCIe adapter cards is done by HMC guided menus or by operating system support utilities.

A blind swap cassette (BSC) is used to house the full high adapters which go into these slots. The BSC is the same BSC as used with the previous generation server's #5802/5803/5877/5873 12X attached I/O drawers.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000098

Figure 2-23 shows the back view of the PCIe Gen3 I/O expansion drawer.

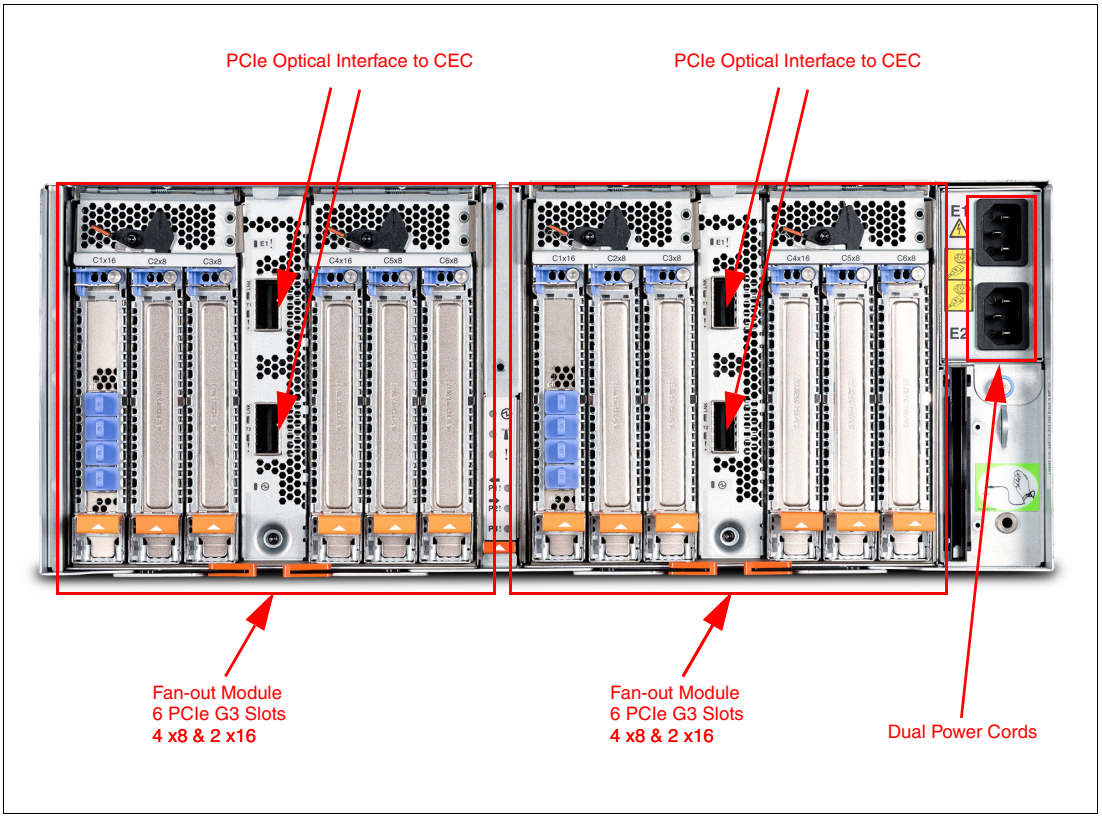

Figure 2-23 Rear view of the PCIe Gen3 I/O expansion drawer

2.7.2 PCIe Gen3 I/O expansion drawer optical cabling

I/O drawers are connected to the adapters in the system node with data transfer cables:

- 3.0 m Optical Cable Pair for PCIe3 Expansion Drawer (#ECC7)
- 10.0 m Optical Cable Pair for PCIe3 Expansion Drawer (#ECC8)

Cable lengths: Use the 3.0 m cables for intra-rack installations. Use the 10.0 m cables for inter-rack installations.

A minimum of one PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer (#EJ07) is required to connect to the PCIe3 6-slot Fan Out module in the I/O expansion drawer. The top port of the fan out module must be cabled to the top port of the #EJ07 port. Likewise, the bottom two ports must be cabled together.

1. Connect an active optical cable to connector T1 on the PCIe3 optical cable adapter in your server.
2. Connect the other end of the optical cable to connector T1 on one of the PCIe3 6-slot Fan Out modules in your expansion drawer.
3. Connect another cable to connector T2 on the PCIe3 optical cable adapter in your server.
4. Connect the other end of the cable to connector T2 on the PCIe3 6-slot Fan Out module in your expansion drawer.
5. Repeat the four steps above for the other PCIe3 6-slot Fan Out module in the expansion drawer, if required.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000099

Drawer connections: Each Fan Out module in a PCIe3 Expansion Drawer can only be connected to a single PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer (#EJ07). However the two Fan Out modules in a single I/O expansion drawer can be connected to different system nodes in the same server.

Figure 2-24 shows the connector locations for the PCIe Gen3 I/O expansion drawer.

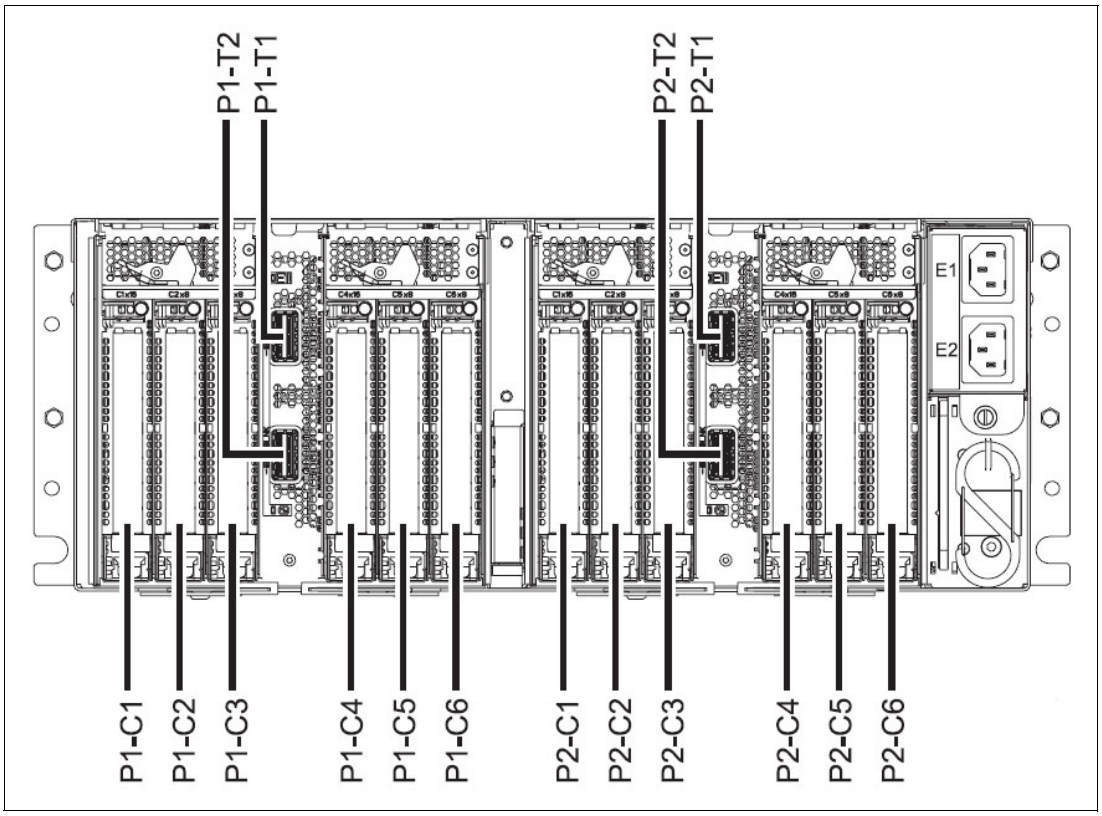

Figure 2-24 Connector locations for the PCIe Gen3 I/O expansion drawer

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000100

Figure 2-25 shows typical optical cable connections.

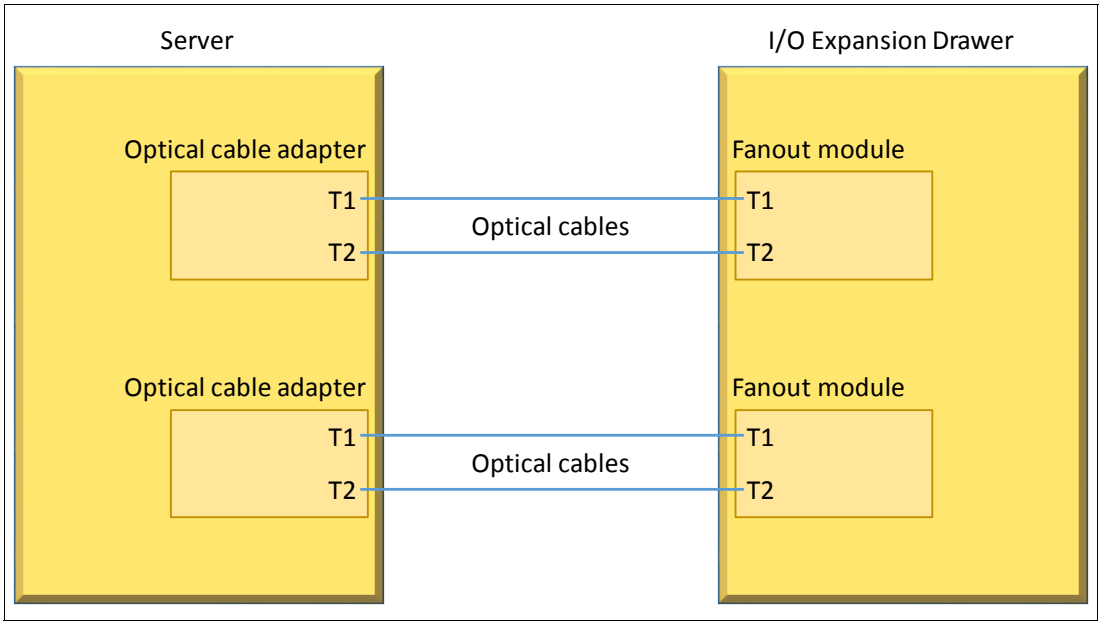

Figure 2-25 Typical optical cable connection

General rules for the PCI Gen3 I/O expansion drawer configuration

The PCIe3 optical cable adapter can be in any of the PCIe adapter slots in the Power S814 and Power S824 servers. However, we advise that you use the PCIe adapter slot priority information while selecting slots for installing PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer (#EJ07).

Table 2-25 shows PCIe adapter slot priorities in the Power S814 and Power S824 system.

Table 2-25 PCIe adapter slot priorities

Feature code	Description	Slot priorities
EJ07	PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer	2, 8, 4, 6, 1, 7, 3, 5

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000101

The following figure shows an example of a supported configuration. For simplification we have not shown every possible combination of the I/O expansion drawer to server attachments.

Figure 2-26 shows an example of a two socket Power S824 and two PCI Gen3 I/O expansion drawers.

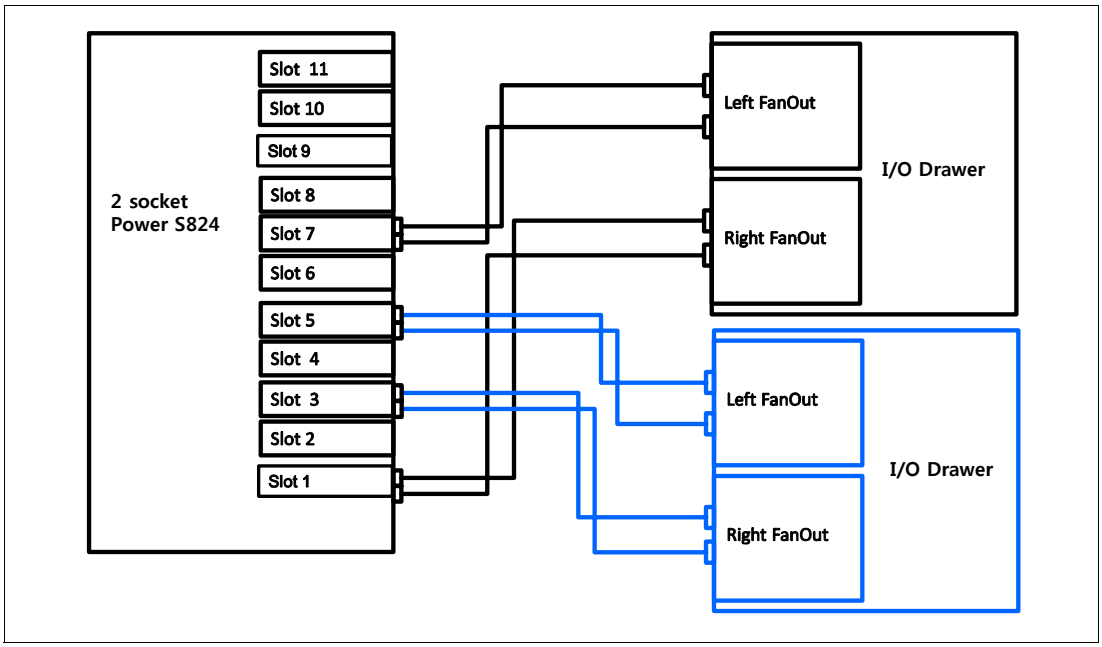

Figure 2-26 Example of a two socket Power S824 and two I/O drawers

2.7.3 PCIe Gen3 I/O expansion drawer SPCN cabling

There is no system power control network (SPCN) used to control and monitor the status of power and cooling within the I/O drawer. SPCN capabilities are integrated in the optical cables.

2.8 External disk subsystems

This section describes the following external disk subsystems that can be attached to the Power S814 and Power S824:

- EXP24S SFF Gen2-bay drawer for high-density storage (#5887)
- IBM System Storage

Note: The EXP30 Ultra SSD Drawer (#EDR1 or #5888), the EXP12S SAS Disk Drawer (#5886), and the EXP24 SCSI Disk Drawer (#5786) are not supported on the Power S814 and Power S824.

2.8.1 EXP24S SFF Gen2-bay drawer

The EXP24S SFF Gen2-bay drawer (#5887) is an expansion drawer that supports up to 24 hot-swap 2.5-inch SFF SAS HDDs on POWER6, POWER6+, POWER7, POWER7+, or

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000102

POWER8 servers in 2U of 19-inch rack space. The EXP24S drawer includes redundant AC power supplies and two power cords.

The EXP24S uses Gen2 or SFF-2 SAS drives that physically do not fit in the SFF-3 bays of the Power S814 and Power S824 system unit.

The EXP24S drawer is attached to SAS ports on either a PCIe SAS adapter in the server or to the SAS ports on the rear of the server. Two SAS ports on the rear of the server are enabled with the expanded-function storage backplane with dual IOA support (#EJ0P).

The SAS controller and the EXP24S SAS ports are attached by using the appropriate SAS Y or X cables.

The following internal SAS adapters support the EXP24S:

- PCIe 380 MB Cache Dual - x4 3 Gb SAS RAID Adapter (#5805, CCIN 574E)
- PCIe Dual port x4 SAS adapter (#5901, CCIN 57B3)
- PCIe 380 MB Cache Dual - x4 3 Gb SAS RAID adapter (#5903, CCIN 574E)
- PCIe2 1.8 GB Cache RAID SAS adapter Tri-Port 6 Gb (#5913, CCIN 57B5)
- PCIe2 RAID SAS Adapter Dual-port 6 Gb (#ESA1, CCIN 57B4)
- PCIe2 1.8 GB Cache RAID SAS Adapter Tri-port 6 Gb (#ESA3, CCIN 57BB)
- PCIe3 RAID SAS Adapter Quad-port 6 Gb (#EJ0J, CCIN 57B4)
- PCIe3 12 GB Cache RAID SAS Adapter Quad-port 6 Gb (#EJ0L, CCIN 57CE)
- PCIe3 12 GB Cache RAID PLUS SAS Adapter Quad-port 6Gb x8 (#EJ14, CCIN 57B1)

The SAS disk drives that are contained in the EXP24S SFF Gen2-bay Drawer are controlled by one or two PCIe SAS adapters that are connected to the EXP24S through SAS cables. The SAS cable varies, depending on the adapter being used, the operating system being used, and the protection you want.

In addition to the existing SAS disks options, IBM has the following available disk models:

- 1.1 TB 10K RPM SAS HDD in Gen-2 Carrier for IBM i (#ESD2)
- 1.2 TB 10K RPM SAS HDD in Gen-2 Carrier for AIX and Linux (#ESD3)

The EXP24S SFF Gen2-bay drawer can be ordered in one of three possible mode settings, which are configured by manufacturing (not by customer setup), of one, two, or four sets of disk bays.

With IBM AIX and Linux, the EXP24S can be ordered with four sets of six bays (mode 4), 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).

There are six SAS connectors on the rear of the EXP24S drawer to which to SAS adapters or controllers are attached. They are labeled T1, T2, and T3; there are two T1, two T2, and two T3 connectors. Figure 2-27 on page 88 shows the rear connectors of the EXP24S drawer.

- In mode 1, two or four of the six ports are used. Two T2 ports are used for a single SAS adapter, and two T2 and two T3 ports are used with a paired set of two adapters or dual adapters configuration.
- In mode 2 or mode 4, four ports are used, two T2s and two T3s, to access all SAS bays.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000103

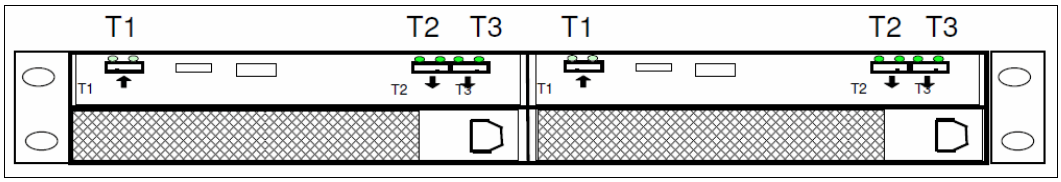

Figure 2-27 EXP24S SFF Gen2-bay drawer rear connectors

An EXP24S drawer in mode 4 can be attached to two or four SAS controllers and provide high configuration flexibility. An EXP24S in mode 2 has similar flexibility. Up to 24 HDDs can be supported by any of the supported SAS adapters or controllers.

Any EXP24S order includes the EXP24S drawer no-charge specify codes to indicate to IBM Manufacturing the mode to which the drawer should be set. The drawer is delivered with this configuration. If multiple EXP24S drawers are ordered, mixing modes should not be within that order. There is no externally visible indicator regarding the drawer's mode.

Notes:

- The modes for the EXP24S drawer are set by IBM Manufacturing. There is no option to reset after the drawer is shipped.
- One #5887 EXP24S drawer in mode 1 can be attached to the two SAS ports on the rear of the Power S814 and Power S824 server by using two SAS YO cables, such as the #ECBT, #ECBU, #ECBV, or #ECBW cable options.
- Either SSDs or HDDs can be placed in this drawer, but SSDs and HDDs cannot be mixed in this drawer when attached using #EJ0P.
- Up to 14 EXP24S drawers can be attached to the Power S814 and Power S824 servers.
- Longer distance SAS cables are thicker and can fill the cable management arm more quickly.

For more information about SAS cabling, see the “SAS cabling for the 5887 drawer” topic in the information center that is found at this website:

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

2.8.2 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 to high-end storage systems. For more information about the various offerings, see the following website:

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

The following section highlights a few of the offerings.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000104

IBM Network-Attached Storage

IBM Network-Attached Storage (NAS) products provide a wide range of network attachment capabilities to a broad range of host and client systems, such as IBM Scale Out Network Attached Storage and the IBM System Storage Nxxx series. For more information about the hardware and software, see the following website:

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

IBM Storwize family

The IBM Storwize® family is the ideal solution to optimize the data architecture for business flexibility and data storage efficiency. Different models, such as the Storwize V3700, V5000, and V7000, offer storage virtualization, IBM Real-time Compression™, Easy Tier, and many other functions. For more information, see the following website:

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

IBM Flash Storage

IBM Flash Storage delivers extreme performance to derive measurable economic value across the data architecture (servers, software, applications, and storage). IBM offers a comprehensive flash portfolio with the IBM FlashSystem™ family. For more information, see the following website:

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

IBM XIV Storage System

IBM XIV® is a high-end disk storage system, helping thousands of enterprises meet the challenge of data growth with hotspot-free performance and ease of use. Simple scaling, high service levels for dynamic, heterogeneous workloads, and tight integration with hypervisors and the OpenStack platform enable optimal storage agility for cloud environments.

XIV extends ease of use with integrated management for large and multi-site XIV deployments, reducing operational complexity and enhancing capacity planning. For more information, see the following website:

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

IBM System Storage DS8000

The IBM System Storage DS8800 is a high-performance, high-capacity, and secure storage system that is designed to deliver the highest levels of performance, flexibility, scalability, resiliency, and total overall value for the most demanding, heterogeneous storage environments. The system is designed to manage a broad scope of storage workloads that exist in today's complex data center, doing it effectively and efficiently.

Additionally, the IBM System Storage DS8000 includes a range of features that automate performance optimization and application quality of service, and also provide the highest levels of reliability and system uptime. For more information, see the following website:

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

2.9 Hardware Management Console (optional)

The Hardware Management Console (HMC) is a dedicated appliance that allows administrators to configure and manage system resources on IBM Power Systems servers

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000105

that use IBM POWER6, POWER6+ POWER7, POWER7+, and POWER8 processors and the PowerVM hypervisor. The HMC provides basic virtualization management support for configuring logical partitions (LPARs) and dynamic resource allocation, including processor and memory settings for selected Power Systems servers. The HMC also supports advanced service functions, including guided repair and verification, concurrent firmware updates for managed systems, and around-the-clock error reporting through IBM Electronic Service Agent™ for faster support.

The HMC management features help improve server usage, simplify systems management, and accelerate provisioning of server resources by using the PowerVM virtualization technology.

Requirements: When using the HMC with the Power S814 and Power S824 servers, the HMC code must be running at V8R8.1.0 level, or later.

The Power S814 and Power S824 platforms support two main service environments:

- Attachment to one or more HMCs. This environment is the common configuration for servers supporting logical partitions with dedicated or virtual I/O. In this case, all servers have at least one logical partition.
- No HMC attachment. Systems that are not attached to an HMC fall under one of two service strategies:
  - Full system partition: A single partition owns all the server resources and only one operating system may be installed.
  - Partitioned system: The system has more than one partition, and the partitions can be running different operating systems. In this environment, partitions are managed by the Integrated Virtualization Manager (IVM), which includes some of the functions that are offered by the HMC.

Hardware support for customer-replaceable units comes standard along with the HMC. In addition, users can upgrade this support level to IBM onsite support to be consistent with other Power Systems servers.

2.9.1 HMC code level

If you are attaching an HMC to a new server or adding a function to an existing server that requires a firmware update, the HMC machine code might need to be updated to support the firmware level of the server. In a dual HMC configuration, both HMCs must be at the same version and release of the HMC code.

To determine the HMC machine code level that is required for the firmware level on any server, go to the following website to access the Fix Level Recommendation Tool (FLRT) on or after the planned availability date for this product:

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

FLRT identifies the correct HMC machine code for the selected system firmware level.

Note: Access to firmware and machine code updates is conditional on entitlement and license validation in accordance with IBM policy and practice. IBM may verify entitlement through customer number, serial number electronic restrictions, or any other means or methods that are employed by IBM at its discretion.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000106

2.9.2 HMC RAID 1 support

HMCs now offer a high-availability feature. The 7042-CR8, and CR9 by default, includes two HDDs with RAID 1 configured. RAID 1 is also offered on the 7042-CR6, 7042-CR7, and 7042-CR8 (if the feature was removed from the initial order) as a miscellaneous equipment specification (MES) upgrade option.

RAID 1 uses data mirroring. Two physical drives are combined into an array, and the same data is written to both drives. This makes the drives a *mirror image* of each other. If one of the drives experiences a failure, it is taken offline and the HMC continues operating with the other drive.

To use an existing HMC to manage any POWER8 processor-based server, the HMC must be a model CR5, or later, rack-mounted HMC, or model C08, or later, deskside HMC. The latest HMC model is the 7042-CR9. For your reference, Table 2-26 lists a comparison between the 7042-CR8 and the 7042-CR9 HMC models.

Note: The 7042-CR9 ships with 16 GB of memory, and is expandable to 192 GB with an upgrade feature. 16 GB is advised for large environments or where external utilities, such as PowerVC and other third party monitors, are to be implemented.

Table 2-26 Comparison between 7042-CR8 and 7042-CR9 models

Feature	CR8	CR9
IBM System x model	x3550 M4 7914 PCH	x3550 M5 5463 AC1
HMC model	7042-CR8	7042-CR9
Processor	Intel 8-Core Xeon v2 2.00 GHz	Intel 18-core Xeon v3 2.4 GHz
Memory max:	16 GB (when featured)	16 GB DDR4 expandable to 192 GB
DASD	500 GB	500 GB
RAID 1	Default	Default
USB ports	Two front, four back	Two front, four rear
Integrated network	Four 1 Gb Ethernet	Four 1 Gb Ethernet
I/O slots	One PCI Express 3.0 slot	One PCI Express 3.0 slot

2.9.3 HMC connectivity to the POWER8 processor-based systems

POWER8 processor-based servers, and their predecessor systems, 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000107

For the HMC to communicate properly with the managed server, eth0 of the HMC must be connected to either the HMC1 or HMC2 ports of the managed server, although other network configurations are possible. You may attach a second HMC to the remaining HMC port of the server for redundancy. The two HMC ports must be addressed by two separate subnets.

Figure 2-28 shows a simple network configuration to enable the connection from the HMC to the server and to allow for dynamic LPAR operations. For more information about HMC and the possible network connections, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491.

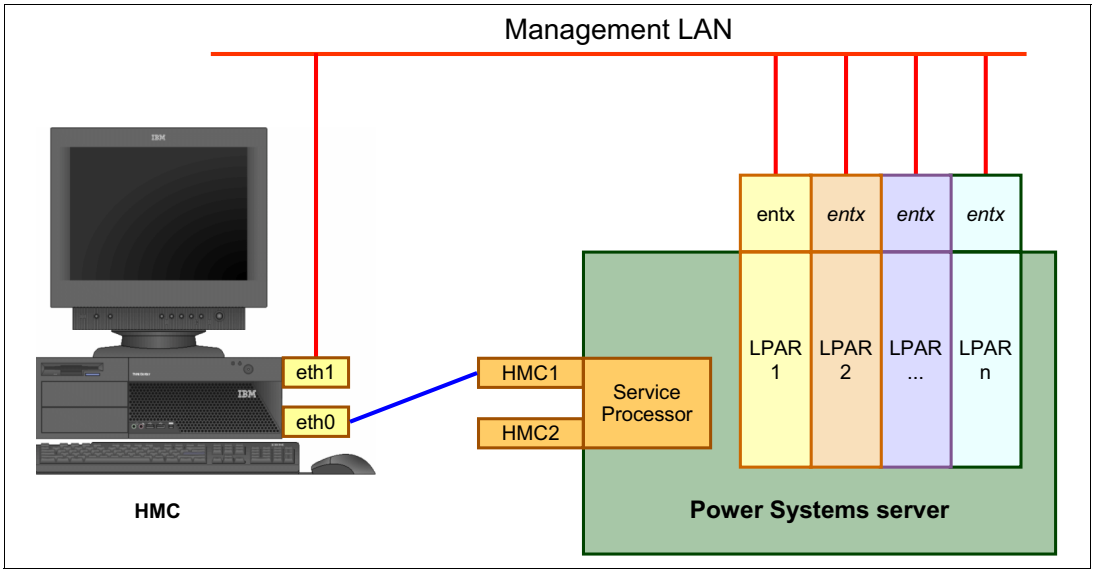

Figure 2-28 Network connections from the HMC to service processor and LPARs

By default, the service processor HMC ports are configured for dynamic IP address allocation. The HMC can be configured as a DHCP server, providing an IP address at the time that the managed server is powered on. In this case, the flexible service processor (FSP) is allocated an IP address from a set of address ranges that are predefined in the HMC software.

If the service processor of the managed server does not receive a DHCP reply before timeout, predefined IP addresses are set up on both ports. Static IP address allocation is also an option and can be configured using the ASMI menus.

Notes: The two service processor HMC ports have the following features:

- Run at a speed of 1 Gbps
- 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 directly from a client web browser
- Use the following network configuration if no IP addresses are set:
  - Service processor eth0 (HMC1 port): 169.254.2.147 with netmask 255.255.255.0
  - Service processor eth1 (HMC2 port): 169.254.3.147 with netmask 255.255.255.0

For more information about the service processor, see “Service processor” on page 158.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000108

2.9.4 High availability HMC configuration

The HMC is an important hardware component. Although Power Systems servers and their hosted partitions can continue to operate when the managing HMC becomes unavailable, certain operations, such as dynamic LPAR, partition migration using PowerVM Live Partition Mobility, or the creation of a new partition, cannot be performed without the HMC. To avoid such situations, consider installing a second HMC, in a redundant configuration, to be available when the other is not (during maintenance, for example).

To achieve HMC redundancy for a POWER8 processor-based server, the server must be connected to two HMCs:

- The HMCs must be running the same level of HMC code.
- The HMCs must use different subnets to connect to the service processor.
- The HMCs must be able to communicate with the server's partitions over a public network to allow for full synchronization and functionality.

Figure 2-29 shows one possible highly available HMC configuration that is managing two servers. Each HMC is connected to one FSP port of each managed server.

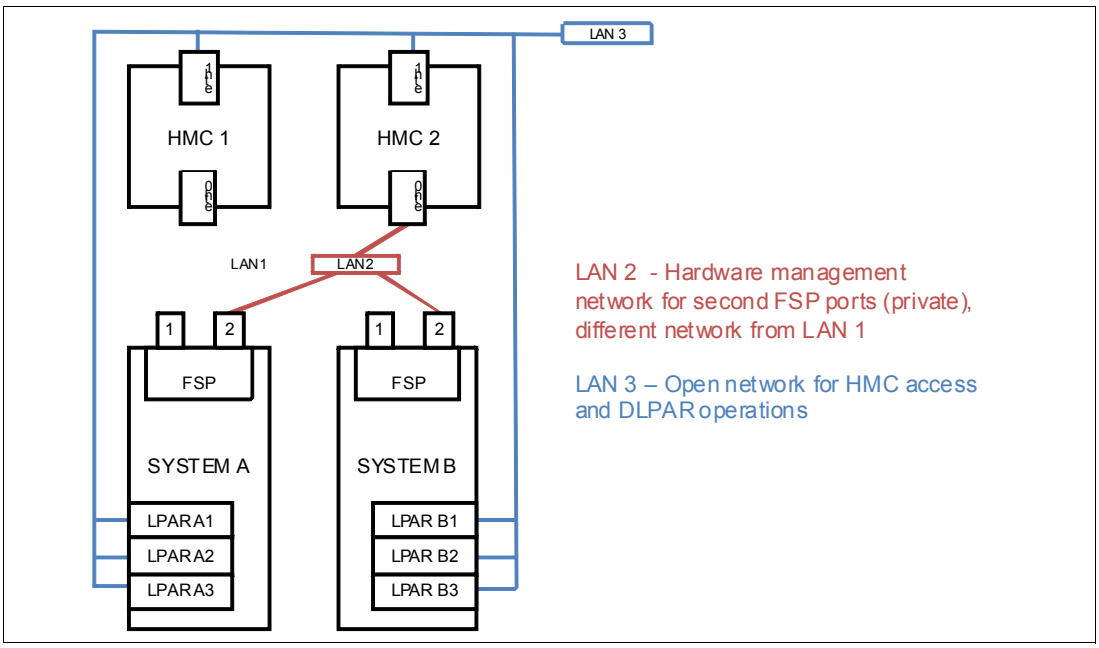

Figure 2-29 Highly available HMC networking example

For simplicity, only the hardware management networks (LAN1 and LAN2) are highly available (Figure 2-29). However, the open network (LAN3) can be made highly available by using a similar concept and adding a second network between the partitions and HMCs.

For more information about redundant HMCs, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000109

2.10 Operating system support

The Power S814 and Power S824 servers support the following operating systems:

- AIX
- IBM i
- Linux

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

For more information about the software that is available on IBM Power Systems, visit the IBM Power Systems Software™ website:

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

2.10.1 AIX operating system

The following sections describe 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, downloading, and obtaining the CD-ROM is at the Fix Central website:

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

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

The Service Update Management Assistant (SUMA), 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

AIX 6.1

The following minimum level of AIX 6.1 supports the Power S814 and Power S824:

AIX 6.1 with the 6100-09 Technology Level and Service Pack 3, with APAR IV56366 or later

These additional AIX levels are supported in an LPAR using virtualized I/O only:

- AIX 6.1 with the 6100-09 Technology Level and Service Pack 1, or later
- AIX 6.1 with the 6100-08 Technology Level and Service Pack 1, or later
- AIX 6.1 with the 6100-07 Technology Level and Service Pack 6, or later

AIX 7.1

The following minimum level of AIX 7.1 supports the Power S814 and Power S824:

AIX 7.1 with the 7100-03 Technology Level and Service Pack 3, with APAR IV56367 or later

These additional AIX levels are supported in an LPAR using virtualized I/O only:

- AIX 7.1 with the 7100-03 Technology Level and Service Pack 1, or later
- AIX 7.1 with the 7100-02 Technology Level and Service Pack 1, or later
- AIX 7.1 with the 7100-01 Technology Level and Service Pack 6, or later

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000110

Note:

- The POWER8 compatibility mode is supported on AIX 7.1 with the 7100-03 Technology Level and Service Pack 3 and later.
- All other prior AIX 7.1 levels and AIX 6.1 can run in POWER6, POWER6+, and POWER7 compatibility mode.
- The Easy Tier function that comes with the 18 bay storage backplane (#EJ0P) requires:
  - AIX 7.1 with the 7100-03 Technology Level and Service Pack 3 with APAR IV56367 or later
  - AIX 6.1 with the 6100-09 Technology Level and Service Pack 3 with APAR IV56366 or later

2.10.2 IBM i

IBM i is supported on the Power S814 and Power S824 with the following minimum required levels:

- IBM i 7.1 with Technology Refresh 8, or later
- IBM i 7.2, or later

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

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

For compatibility information for hardware features and the corresponding AIX and IBM i Technology Levels, visit the IBM Prerequisite website:

http://www-912.ibm.com/e_dir/eserverprereq.nsf

2.10.3 Linux operating system

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

The supported versions of Linux on the Power S814 and Power S824 servers are as follows:

- Red Hat Enterprise Linux 6.5, or later
- SUSE Linux Enterprise Server 11 Service Pack 3, or later

Linux supports almost all of the Power System I/O and the configurator verifies support on order.

Be sure to connect your systems to the IBM Service and Productivity Tools for PowerLinux™ repository and keep up to date with the latest Linux service and productivity tools, available from IBM at the following website:

http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/yum.html

For information about the PowerLinux Community, see the following website:

https://www.ibm.com/developerworks/group/tpl

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000111

For information about the features and external devices that are supported by Linux, see the following website:

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

For more information about SUSE Linux Enterprise Server, see the following website:

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

For more information about Red Hat Enterprise Linux Advanced Server, see the following website:

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

2.10.4 Virtual I/O Server

The minimum required level of VIOS for both the Power S814 and Power S824 is VIOS 2.2.3.51.

However some of the new adapters, such as SR-IOV adapters, require VIOS 2.2.3.51. For more detailed information about the different requirements, see the announcement letter on the IBM Offering Information website at:

http://www-01.ibm.com/common/ssi/index.wss?request_locale=en

Note: The Easy Tier function that comes with the 18-bay storage backplane (#EJ0P) requires VIOS 2.2.3.3 with interim fix IV56366 or later.

IBM regularly updates the VIOS code. To find information about the latest updates, visit the Fix Central website:

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

2.10.5 Java

There are unique considerations when running Java on POWER8 servers. For the best use of the performance capabilities and most recent improvements of POWER8 technology, upgrade your JVM (or JDK) to IBM Java7 Release1 when possible, although IBM Java7, Java6, or Java5 still can run on POWER8. For more information, see the AIX Download and service information website:

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

2.11 Energy management

The Power S814 and Power S824 systems have features to help clients become more energy efficient. EnergyScale technology enables advanced energy management features to conserve power dramatically and dynamically and further improve energy efficiency. Intelligent Energy optimization capabilities enable the POWER8 processor to operate at a higher frequency for increased performance and performance per watt, or dramatically reduce frequency to save energy.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000112

2.11.1 IBM EnergyScale technology

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

EnergyScale uses power and thermal information that is collected from the system to implement policies that can lead to better performance or better energy usage. IBM EnergyScale has the following features:

- Power trending

  EnergyScale provides continuous collection of real-time server energy consumption. It 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 data center energy consumption at various times of the day, week, or month.

- 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 almost 50% down from nominal frequency (the actual value depends on the server type and configuration).

  Power saver mode is not supported during system start, although it is a persistent condition that is sustained after the boot when the system starts running instructions.

- Dynamic power saver mode

  Dynamic power saver mode varies processor frequency and voltage based on the usage of the POWER8 processors. Processor frequency and usage are inversely proportional for most workloads, implying that as the frequency of a processor increases, its usage 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 usage.

  When a system is idle, the system firmware lowers the frequency and voltage to power energy saver mode values. When fully used, the maximum frequency varies, depending on whether the user favors power savings or system performance. If an administrator prefers energy savings and a system is fully used, the system is designed to reduce the maximum frequency to about 95% of nominal values. If performance is favored over energy consumption, the maximum frequency can be increased to up to 111.3% 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 extra power that was 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 data center. 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000113

- Soft power capping

  There are two power ranges into which the power cap can be set: power capping, as described previously, and soft power capping. Soft power capping extends the allowed energy capping range further, beyond a region that can be ensured 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

  IBM POWER8 processor uses a low-power mode that is called *nap* that stops processor execution when there is no work to do on that processor core. The latency of exiting nap mode is small, typically not generating any impact on applications running. Therefore, the IBM 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 off the clock on most of the circuits in 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 and causes a cumulative effect. Nap mode saves 10 - 15% of power consumption in the processor core.

- Processor core sleep mode

  To save even more energy, the POWER8 processor has an even lower power mode referred to as *sleep*. Before a core and its associated private L2 cache enter sleep mode, the cache is flushed, transition lookaside buffers (TLB) are invalidated, and the hardware clock is turned off in the core and in the cache. Voltage is reduced to minimize leakage current. Processor cores that are inactive in the system (such as capacity on demand (CoD) processor cores) are kept in sleep mode. Sleep mode saves about 80% power consumption in the processor core and its associated private L2 cache.

- Processor chip winkle mode

  The most amount of energy can be saved when a whole POWER8 chiplet enters the *winkle* mode. In this mode, the entire chiplet is turned off, including the L3 cache. This mode can save more than 95% power consumption.

- Fan control and altitude input

  System firmware dynamically adjusts 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), the fan speed varies based on power consumption and ambient temperature.

- Processor folding

  Processor folding is a consolidation technique that dynamically adjusts, over the short term, the number of processors that are available for dispatch to match the number of processors that are demanded by the workload. As the workload increases, the number of processors made available increases. As the workload decreases, the number of processors that are 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 POWER8 processor-based systems automatically power off hot-pluggable PCI adapter slots that are empty or not being used. System firmware automatically scans all

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000114

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 POWER8 processor-based servers and the expansion units that they support.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000115

- Server power down

  If overall data center processor usage is low, workloads can be consolidated on fewer numbers of servers so that some servers can be turned off completely. Consolidation makes sense when there are long periods of low usage, such as weekends. Live Partition Mobility can be used to move workloads to consolidate partitions onto fewer systems, reducing the number of servers that are powered on and therefore reducing the power usage.

  On POWER8 processor-based systems, several EnergyScale technologies are embedded in the hardware and do not require an operating system or external management component. Fan control, environmental monitoring, and system energy management are controlled by the On Chip Controller (OCC) and associated components. The power mode can also be set up without external tools, by using the ASMI interface, as shown in Figure 2-30.

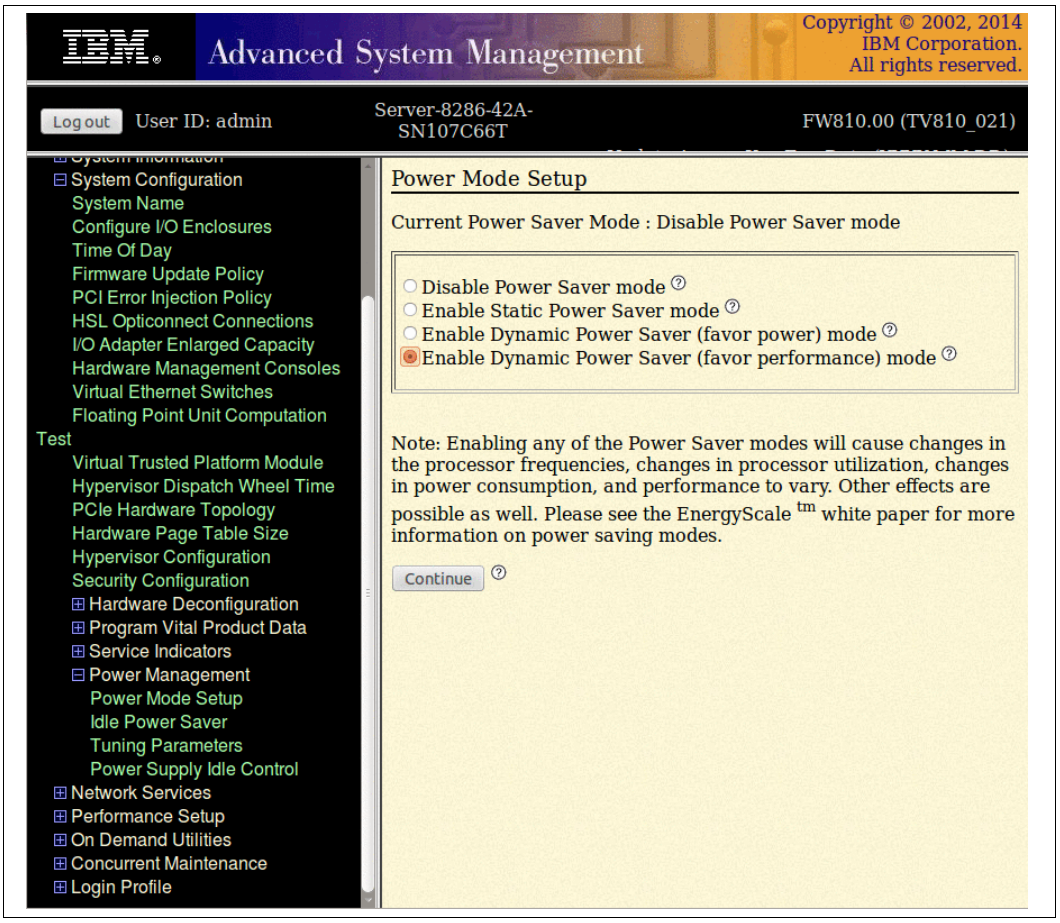

Figure 2-30 Setting the power mode in ASMI

2.11.2 On Chip Controller

To maintain the power dissipation of POWER7+ with its large increase in performance and bandwidth, POWER8 invested significantly in power management innovations. A new OCC using an embedded IBM PowerPC® core with 512 KB of SRAM runs real-time control firmware to respond to workload variations by adjusting the per-core frequency and voltage based on activity, thermal, voltage, and current sensors.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000116

The on-die nature of the OCC allows for approximately 100× speedup in response to workload changes over POWER7+, enabling reaction under the timescale of a typical OS time slice and allowing for multi-socket, scalable systems to be supported. It also enables more granularity in controlling the energy parameters in the processor, and increases reliability in energy management by having one controller in each processor that can perform certain functions independently of the others.

POWER8 also includes an internal voltage regulation capability that enables each core to run at a different voltage. Optimizing both voltage and frequency for workload variation enables better increase in power savings versus optimizing frequency only.

2.11.3 Energy consumption estimation

Often, for Power Systems, various energy-related values are important:

- Maximum power consumption and power source loading values

  These values are important for site planning and are described in the hardware information center that is found at the following website:

  http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp

  Search for type and model number and “server specifications”. For example, for the Power S814 and Power S824 system, search for “8286-41A” or “8286-42A server specifications”.

- An estimation of the energy consumption for a certain configuration

  The calculation of the energy consumption for a certain configuration can be done in the IBM Systems Energy Estimator, found at the following website:

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

  In that tool, select the type and model for the system, and enter some details about the configuration and wanted CPU usage. As a result, the tool shows the estimated energy consumption and the waste heat at the wanted usage and also at full usage.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000117

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 offer key capabilities that can help you consolidate and simplify your IT environment:

- Improve server usage and share I/O resources to reduce total cost of ownership (TCO) and make better usage of IT assets.
- Improve business responsiveness and operational speed by dynamically reallocating 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, so you can make business-driven policies to deliver resources based on time, cost, and service-level requirements.

Single Root I/O Virtualization (SR-IOV) is now supported on the Power S814 and Power S824. For more information about SR-IOV see chapter 3.4, “Single Root I/O Virtualization” on page 113.

Note: PowerKVM is not supported on the Power S814 and Power S824 server.

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000119

3.1 IBM POWER Hypervisor

Combined with features in the POWER8 processors, the IBM 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 120).
- 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 reduce the need for physical Ethernet adapters for inter-partition communication.
- Monitors the service processor and performs 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 that is required by the POWER Hypervisor firmware varies according to several factors:

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

The minimum amount of physical memory that is required to create a partition is the size of the system's logical memory block (LMB). The default LMB size varies according to the amount of memory that is configured in the Central Electronics Complex (see Table 3-1).

Table 3-1 Configured Central Electronics Complex memory-to-default logical memory block size

Configurable Central Electronics Complex memory	Default logical memory block
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 greater than 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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000120

3.1.1 Virtual SCSI

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

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

A Virtual I/O Server (VIOS) partition or an IBM i partition can define virtual SCSI server adapters. Other partitions are *client* partitions. The VIOS partition is a special logical partition, which is described in 3.5.4, “Virtual I/O Server” on page 122. The VIOS software is included on all PowerVM editions. When using the PowerVM Standard Edition and PowerVM Enterprise Edition, dual VIOS can be deployed to provide maximum availability for client partitions when performing VIOS maintenance.

3.1.2 Virtual Ethernet

The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions on the same server to use fast and secure communication without any need for physical interconnection. The virtual Ethernet allows a transmission speed up to 20 Gbps, depending on the maximum transmission unit (MTU) size, type of communication, and CPU entitlement. Virtual Ethernet support began with IBM AIX Version 5.3, Red Hat Enterprise Linux 4, and SUSE Linux Enterprise Server, 9, and it is supported on all later versions. For more information, see 3.5.8, “Operating system support for PowerVM” on page 128.

Virtual Ethernet is part of the base system configuration and has the following 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 (or 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 4096. 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 VIOS partition, also known as Shared Ethernet Adapter. For more information about shared Ethernet, see 3.5.4, “Virtual I/O Server” on page 122.

Adapter and access: 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000121

3.1.3 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 VIOS logical partition. The VIOS logical partition provides the connection between the virtual Fibre Channel adapters on the VIOS logical partition and the physical Fibre Channel adapters on the managed system. Figure 3-1 shows the connections between the client partition virtual Fibre Channel adapters and the external storage. For more information, see 3.5.8, “Operating system support for PowerVM” on page 128.

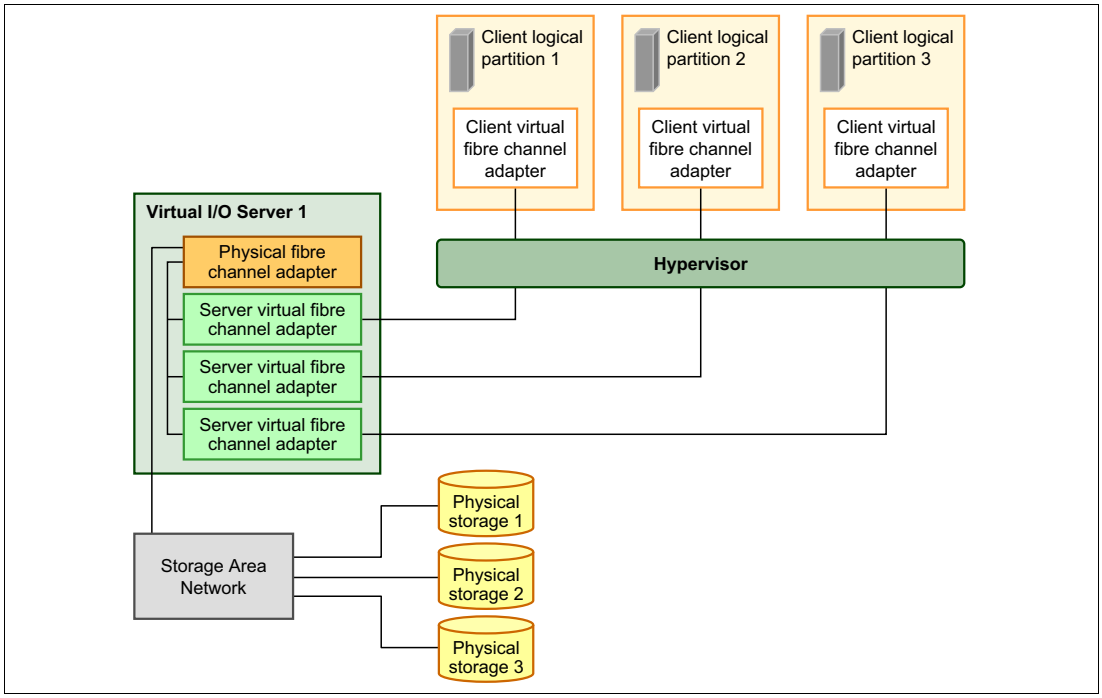

Figure 3-1 Connectivity between virtual Fibre Channels adapters and external SAN devices

3.1.4 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 (HMC) virtual TTY, IVM virtual TTY, or from a terminal emulator that is connected to a system port.

3.2 POWER processor modes

Although they are not virtualization features, the POWER processor modes are described here because they affect various virtualization features.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000122

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

- 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 website:

  http://power.org/wp-content/uploads/2012/07/PowerISA_V2.05.pdf

- POWER6+ compatibility mode

  This mode is similar to POWER6, with eight more storage protection keys.

- POWER7 compatibility mode

  This is the mode for POWER7+ and POWER7 processors, implementing Version 2.06 of the Power Instruction Set Architecture. For more information, visit the following website:

  http://power.org/wp-content/uploads/2012/07/PowerISA_V2.06B_V2_PUBLIC.pdf

- POWER8 compatibility mode

  This is the native mode for POWER8 processors implementing Version 2.07 of the Power Instruction Set Architecture. For more information, visit the following address:

  https://www.power.org/documentation/power-isa-version-2-07/

The selection of the mode is made on a per-partition basis, from the managed console, by editing the partition profile.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000123

Figure 3-2 shows the compatibility modes within the LPAR profile.

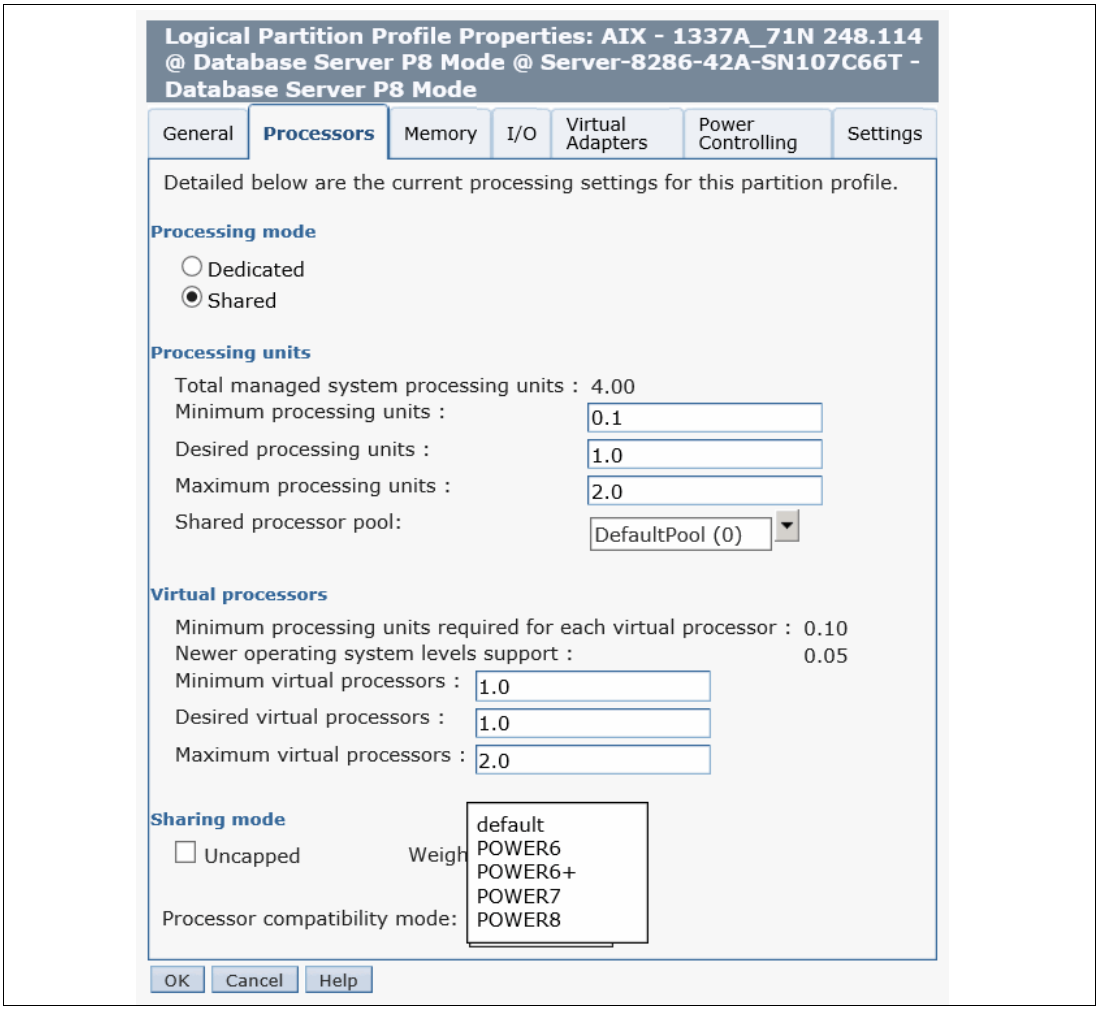

Figure 3-2 Configuring partition profile compatibility mode using the HMC

Table 3-2 lists the differences between processors modes.

Table 3-2 Differences between POWER6, POWER7, and POWER8 compatibility modes

POWER6 and POWER6+ mode	POWER7 mode	POWER8 mode	Customer value
2-thread SMT	4-thread SMT	8-thread SMT	Throughput performance, and processor core usage
Vector Multimedia Extension/ AltiVec (VMX)	Vector scalar extension (VSX)	VSX2 In-Core Encryption Acceleration	High-performance computing
Affinity off by default	3-tier memory, micropartition affinity, and dynamic platform optimizer	► HW memory affinity tracking assists ► Micropartition prefetch ► Concurrent LPARs per core	Improved system performance for system images spanning sockets and nodes

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000124

POWER6 and POWER6+ mode	POWER7 mode	POWER8 mode	Customer value
64-core and 128-thread scaling	► 32-core and 128-thread scaling ► 64-core and 256-thread scaling ► 128-core and 512-thread scaling ► 256-core and 1024-thread scaling	► 1024-thread Scaling ► Hybrid threads ► Transactional memory ► Active system optimization hardware assists	Performance and scalability for large scale-up single system image workloads (such as OLTP, ERP scale-up, and WPAR consolidation)
EnergyScale CPU Idle	EnergyScale CPU Idle and Folding with NAP and SLEEP	WINKLE, NAP, SLEEP, and Idle power saver	Improved energy efficiency

3.3 Active Memory Expansion

Active Memory Expansion is an optional feature for the Power S814 and Power S824 when you select feature code #4794 in the e-Config tool.

This feature enables memory expansion on the system. By using compression and decompression of memory, content can effectively expand the maximum memory capacity, providing additional server workload capacity and performance.

Active Memory Expansion is a technology that allows the effective maximum memory capacity to be much larger than the true physical memory maximum. Compression and decompression of memory content can allow memory expansion up to 125% for AIX partitions, which in turn enables a partition to perform 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.

Note: The Active Memory Expansion feature is not supported by IBM i and the Linux operating system.

Active Memory Expansion uses the CPU resource of a partition to compress and 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 and decompression.

The POWER8 processor includes Active Memory Expansion on the processor chip to provide dramatic improvement in performance and greater processor efficiency. To take advantage of the hardware compression offload, AIX 6.1 Technology Level 8, or later is required.

Tests in IBM laboratories, using sample work loads, showed excellent results for many workloads in terms of memory expansion per additional CPU used. Other test workloads had more modest results. The ideal scenario is when there are many cold pages, that is, infrequently referenced pages. However, if many memory pages are referenced frequently, Active Memory Expansion might not be a preferred choice.

Tip: If the workload is based on Java, the garbage collector must be tuned so that it does not access the memory pages so often, that is, turning cold pages to hot.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000125

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 you want in each partition to help control the amount of CPU that is used by the Active Memory Expansion function. An initial program load (IPL) is required for the specific partition that is turning on or off memory expansion. After it is turned on, monitoring capabilities are available in standard AIX performance tools, such as **lparstat**, **vmstat**, **topas**, and **svmon**. For specific POWER8 hardware compression, the **amepat** tool is used to configure the offload details.

Figure 3-3 represents the percentage of CPU that is used to compress memory for two partitions with separate profiles. Curve 1 corresponds to a partition that has spare processing power capacity. Curve 2 corresponds to a partition that is constrained in processing power.

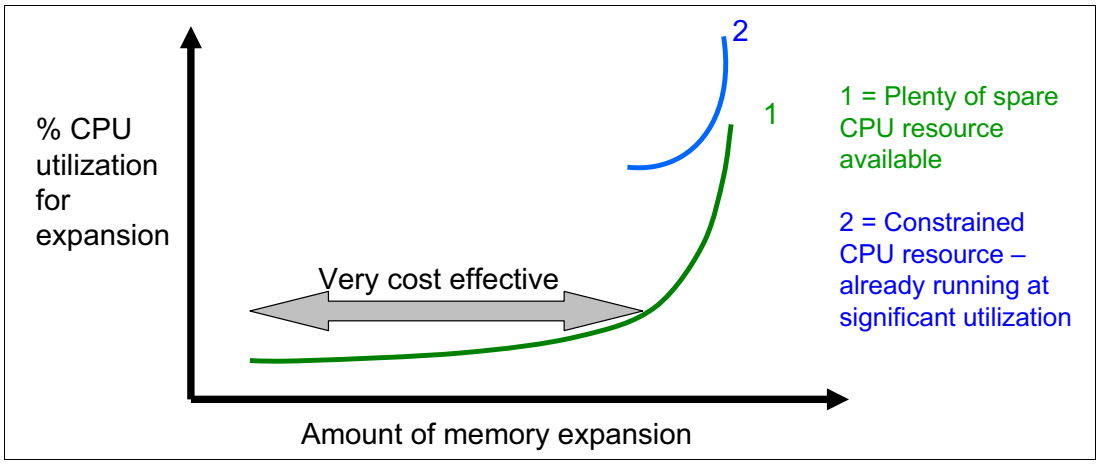

Figure 3-3 CPU usage versus memory expansion effectiveness

Both cases show that there is a “knee-of-curve” relationship for the CPU resource that is required for memory expansion:

- Busy processor cores do not have resources to spare for expansion.
- The more memory expansion that is done, the more CPU resource is 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 (ROI).

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000126

To help you do this study, use the **amepat** planning tool (introduced in AIX 6.1 Technology Level 4 SP2), which you can use to sample actual workloads, and estimate how expandable the partition's memory is and how much CPU resource is needed. Any Power System can run the planning tool.

Figure 3-4 shows an example of the output that is returned by this planning tool. The tool outputs various real memory and CPU resource combinations to achieve the wanted effective memory. It also recommends one particular combination. Figure 3-4 shows an example of how the tool recommends that you allocate 13% of processing power to benefit from 119% extra memory capacity.

```
Active Memory Expansion Modeled Statistics:
-------------------------------------------
Modeled Expanded Memory Size : 52.00 GB
Achievable Compression ratio : 4.51

Expansion    Modeled True      Modeled              CPU Usage
Factor       Memory Size       Memory Gain          Estimate
---------    -------------     ------------------   -----------
    1.40         37.25 GB        14.75 GB [ 40%]   0.00 [  0%]
    1.80         29.00 GB        23.00 GB [ 79%]   0.87 [  5%]
    2.19         23.75 GB        28.25 GB [119%]   2.13 [ 13%]
    2.57         20.25 GB        31.75 GB [157%]   2.96 [ 18%]
    2.98         17.50 GB        34.50 GB [197%]   3.61 [ 23%]
    3.36         15.50 GB        36.50 GB [235%]   4.09 [ 26%]

Active Memory Expansion Recommendation:
---------------------------------------
The recommended AME configuration for this workload is to configure the LPAR
with a memory size of 23.75 GB and to configure a memory expansion factor
of 2.19.  This will result in a memory gain of 119%. With this
configuration, the estimated CPU usage due to AME is approximately 2.13
physical processors, and the estimated overall peak CPU resource required for
the LPAR is 11.65 physical processors.
```

Figure 3-4 Output from the Active Memory Expansion planning tool

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000127

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 managed console.

Figure 3-5 shows the activation of AME for each LPAR.

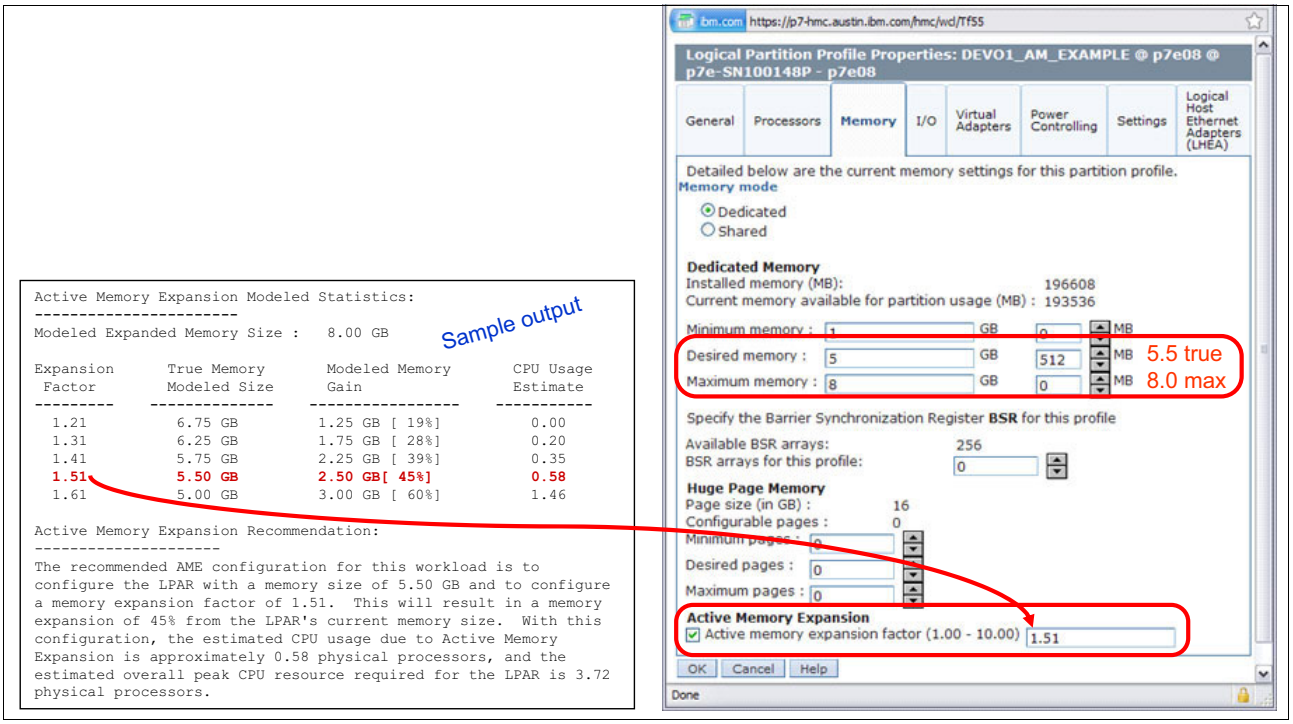

Figure 3-5 Using the planning tool result to configure the partition

On the HMC menu that describes the partition, select the **Active Memory Expansion** check box and enter the true and maximum memory, and the memory expansion factor. To turn off expansion, clear the check box. In both cases, reboot the partition 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 by using the Power Systems Capacity on Demand website:

http://www.ibm.com/systems/power/hardware/cod/

Active Memory Expansion can be ordered with the initial order of the server or as a miscellaneous equipment specification (MES) order. A software key is provided when the enablement feature is ordered, which 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, independent of the number of partitions using the memory expansion.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000128

From the HMC, you can view whether the Active Memory Expansion feature was activated for the server or not.

Figure 3-6 shows the available server capabilities.

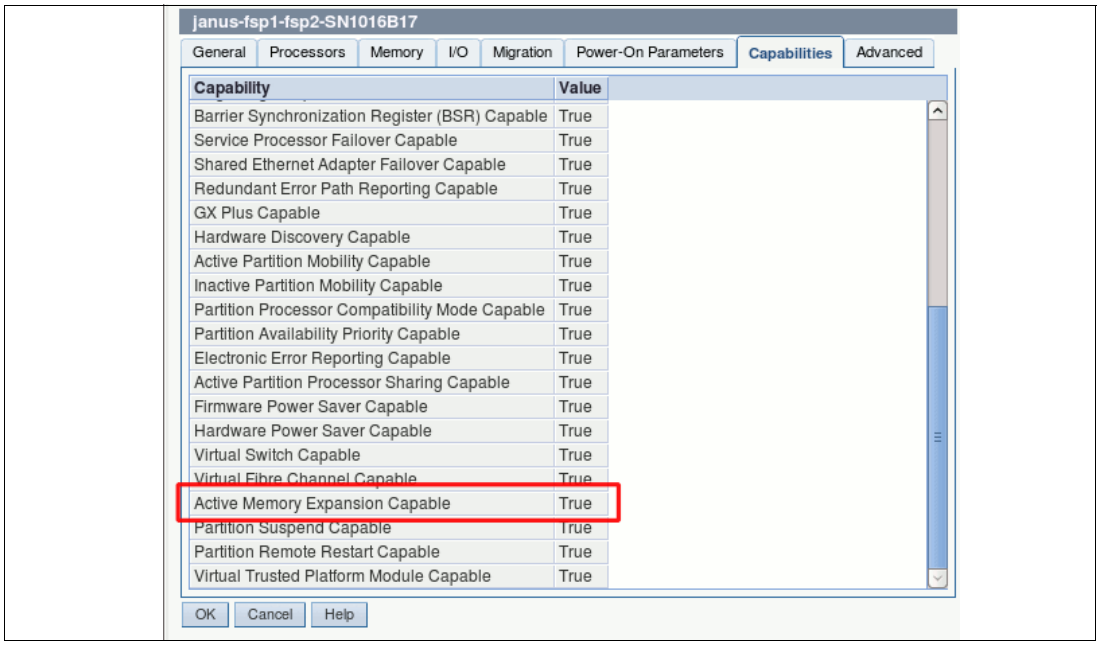

Figure 3-6 Server capabilities that are listed from the HMC

Moving an LPAR: If you want to move an LPAR that has Active Memory Expansion enabled to another physical server using Live Partition Mobility, the target server must have Active Memory Expansion activated with the software key. If the target system does not have Active Memory Expansion activated, the mobility operation fails during the premobility check phase, and an appropriate error message is displayed.

For more information about Active Memory Expansion, download the document *Active Memory Expansion: Overview and Usage Guide*, found at:

http://public.dhe.ibm.com/common/ssi/ecm/en/pow03037usen/POW03037USEN.PDF

3.4 Single Root I/O Virtualization

Single root I/O virtualization (SR-IOV) is an extension to the PCI Express (PCIe) specification that allows multiple operating systems to simultaneously share a PCIe adapter with little or no runtime involvement from a hypervisor or other virtualization intermediary.

SR-IOV is PCI standard architecture that enables PCIe adapters to become self-virtualizing. It enables adapter consolidation, through sharing, much like logical partitioning enables server consolidation. With an adapter capable of SR-IOV, you can assign virtual *slices* of a single physical adapter to multiple partitions through logical ports; all of this is done without the need for a Virtual I/O Server (VIOS).

Initial SR-IOV deployment supports up to 48 logical ports per adapter, depending on the adapter. You can provide additional fan-out for more partitions by assigning a logical port to a

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000129

VIOS, and then using that logical port as the physical device for a Shared Ethernet Adapter (SEA). VIOS clients can then use that SEA through a traditional virtual Ethernet configuration.

Overall, SR-IOV provides integrated virtualization without VIOS and with greater server efficiency as more of the virtualization work is done in the hardware and less in the software.

The following are the hardware requirements to enable SR-IOV:

- One of the following pluggable PCIe adapters:
  - PCIe2 4-port (10Gb FCoE and 1GbE) SR&RJ45 Adapter (#EN0H)
  - PCIe2 4-port (10Gb FCoE and 1GbE) SFP+Copper and RJ4 Adapter (#EN0K)
  - PCIe2 4-port (10Gb FCoE and 1GbE) LR&RJ45 Adapter (#EN0M)
  - PCIe3 4-port (10Gb) SR Adapter (#EN15)
  - PCIe3 4-port (10Gb) SFP+Copper Adapter (#EN17)

The minimum operating system requirements, related to SR-IOV functions, are as follows:

- VIOS
  - Virtual I/O Server Version 2.2.3.51
- AIX
  - AIX 6.1 Technology Level 9 with Service Pack 5 and APAR IV68443 or later
  - AIX 7.1 Technology Level 3 with Service Pack 5 and APAR IV68444 or later
- Linux
  - SUSE Linux Enterprise Server 11 SP3, or later
  - SUSE Linux Enterprise Server 12, or later
  - Red Hat Enterprise Linux 6.5, or later
  - Red Hat Enterprise Linux 7, or later
- IBM i
  - IBM i 7.1 TR10 or later
  - IBM i 7.2 TR2 or later

Firmware level: SR-IOV requires firmware level 8.3, or later for the Power S814 and Power S824 servers. Check the Fix Central portal to verify the specific firmware levels for your type of the machine at:

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

The entire adapter (all four ports) is configured for SR-IOV or none of the ports is. (FCoE not supported when using SR-IOV).

SR-IOV provides significant performance and usability benefits, as described in the following sections.

For more information about SR-IOV see the *IBM Power Systems SR-IOV: Technical Overview and Introduction*, REDP-5065 at:

http://www.redbooks.ibm.com/abstracts/redp5065.html?Open

3.4.1 Direct access I/O and performance

The primary benefit of allocating adapter functions directly to a partition, as opposed to using a virtual intermediary (VI) like VIOS, is performance. The processing overhead involved in passing client instructions through a VI, to the adapter and back, are substantial.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000130

With direct access I/O, SR-IOV capable adapters running in shared mode allow the operating system to directly access the slice of the adapter that has been assigned to its partition, so there is no control or data flow through the hypervisor. From the partition perspective, the adapter appears to be physical I/O. With respect to CPU and latency, it exhibits the characteristics of physical I/O; and because the operating system is directly accessing the adapter, if the adapter has special features, like multiple queue support or receive side scaling (RSS), the partition can leverage those also, if the operating system has the capability in its device drivers.

3.4.2 Adapter sharing

The current trend of consolidating servers to reduce cost and improve efficiency is increasing the number of partitions per system, driving a requirement for more I/O adapters per system to accommodate them. SR-IOV addresses and simplifies that requirement by enabling the sharing of SR-IOV capable adapters. Because each adapter can be shared and directly accessed by up to 48 partitions, depending on the adapter, the partition to PCI slot ratio can be significantly improved without adding the overhead of a virtual intermediary.

3.4.3 Adapter resource provisioning (QoS)

Power Systems SR-IOV provides QoS controls to specify a capacity value for each logical port, improving the ability to share adapter ports effectively and efficiently. The capacity value determines the desired minimum percentage of the physical port's resources that should be applied to the logical port.

The exact resource represented by the capacity value can vary based on the physical port type and protocol. In the case of Ethernet physical ports, capacity determines the minimum percentage of the physical port's transmit bandwidth that the user desires for the logical port.

For example, consider Partitions A, B, and C, with logical ports on the same physical port. If Partition A is assigned an Ethernet logical port with a capacity value of 20%, Partitions B and C cannot use more than 80% of the physical port's transmission bandwidth unless Partition A is using less than 20%. Partition A can use more than 20% if bandwidth is available. This ensures that, although the adapter is being shared, the partitions maintain their portion of the physical port resources when needed.

3.4.4 Flexible deployment

Power Systems SR-IOV enables flexible deployment configurations, ranging from a simple, single-partition deployment, to a complex, multi-partition deployment involving VIOS partitions and VIOS clients running different operating systems.

In a single-partition deployment, the SR-IOV capable adapter in shared mode is wholly owned by a single partition, and no adapter sharing takes place. This scenario offers no practical benefit over traditional I/O adapter configuration, but the option is available.

In a more complex deployment scenario, an SR-IOV capable adapter could be shared by both VIOS and non-VIOS partitions, and the VIOS partitions could further virtualize the logical ports as shared Ethernet adapters for VIOS client partitions. This scenario leverages the benefits of direct access I/O, adapter sharing, and QoS that SR-IOV provides, and also the benefits of higher-level virtualization functions, such as Live Partition Mobility (for the VIOS clients), that VIOS can offer.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000131

3.4.5 Reduced costs

SR-IOV facilitates server consolidation by reducing the number of physical adapters, cables, switch ports, and I/O slots required per system. This translates to reduced cost in terms of physical hardware required, and also reduced associated energy costs for power consumption, cooling, and floor space. You may save additional cost on CPU and memory resources, relative to a VIOS adapter sharing solution, because SR-IOV does not have the resource overhead inherent in using a virtualization intermediary to interface with the adapters.

3.5 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that delivers industry-leading virtualization on the IBM Power Systems. It is the umbrella branding term for Power Systems virtualization (logical partitioning, IBM Micro-Partitioning®, POWER Hypervisor, VIOS, Live Partition Mobility, and more). As with Advanced Power Virtualization in the past, PowerVM is a combination of hardware enablement and software. The licensed features of each of the two separate editions of PowerVM are described in 3.5.1, “PowerVM editions” on page 116.

3.5.1 PowerVM editions

The two editions of PowerVM are suited for various purposes:

- 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 infrastructures. It includes unique features, such as Active Memory Sharing and Live Partition Mobility.

Table 3-3 lists the editions of PowerVM that are available on the Power S814 and Power S824 servers.

Table 3-3 Available PowerVM editions

Servers	Standard	Enterprise
Power S814	#5227	#5228
Power S824	#5227	#5228

3.5.2 Logical partitions

Logical partitions (LPARs) and virtualization increase the usage of system resources and add a level of configuration possibilities.

Logical partitioning

Logical partitioning was introduced with the POWER4 processor-based product line and the AIX Version 5.1, Red Hat Enterprise Linux 3.0, and SUSE Linux Enterprise Server 9.0 operating systems. This technology was able to divide an IBM eServer™ pSeries (now IBM

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000132

System p®) 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 running. AIX Version 5.2, with all the necessary enhancements to enable dynamic LPAR, was introduced in 2002. At the same time, Red Hat Enterprise Linux 5 and SUSE Linux Enterprise 9.0 were also able to support dynamic logical partitioning. 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000133

Micro-Partitioning

When you use the Micro-Partitioning technology, you can 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 *micropartition*. Micropartitions run over a set of processors that are called a *shared processor pool*, and virtual processors are used to let the operating system manage the fractions of processing power that are assigned to the logical partition. From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system is enhanced to determine 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*.

When defining a shared processor partition, several options must be defined:

- The minimum, wanted, 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

  Select a pool from the list with the names of each configured shared processor pool. This list also shows, in parentheses, the pool ID of each configured shared processor pool. If the name of the wanted shared processor pool is not available here, you must first configure the shared processor pool by using the shared processor pool Management window. Shared processor partitions use the default shared processor pool, called DefaultPool by default. For more information about multiple shared processor pools, see 3.5.3, “Multiple shared processor pools” on page 121.

- Whether the partition can access extra processing power to “fill up” its virtual processors above its capacity entitlement (you select either to cap or uncap your partition).

  If spare processing power is 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) if there is an uncapped partition.

- The minimum, wanted, and maximum number of virtual processors.

The POWER Hypervisor calculates partition processing power based on minimum, wanted, and maximum values, processing mode, and the requirements of other active partitions. The actual entitlement is never smaller than the processing unit’s wanted value, but can exceed that value if it is an uncapped partition and up to the number of virtual processors that are allocated.

On the POWER8 processors, a partition can be defined with a processor capacity as small as 0.05processing units. This number represents 0.05 of a physical core. Each physical core can be shared by up to 20 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 HMC.

The Power S814 supports up to eight cores in a single system. Here are the maximum numbers:

- Eight dedicated partitions
- 160 micropartitions (maximum of 20 micropartitions per physical active core)

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000134

The Power S824 supports up to 24 cores in a single system. Here are the maximum numbers:

- 24 dedicated partitions
- 480 micropartitions (maximum of 20 micropartitions per physical active core)

An important point is that the maximum amounts are supported by the hardware, but the practical limits depend on application workload demands.

Consider the following additional information about virtual processors:

- A virtual processor can be running (dispatched) either on a physical core or as standby waiting for a physical core to became 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 a 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000135

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 that are described in this section.

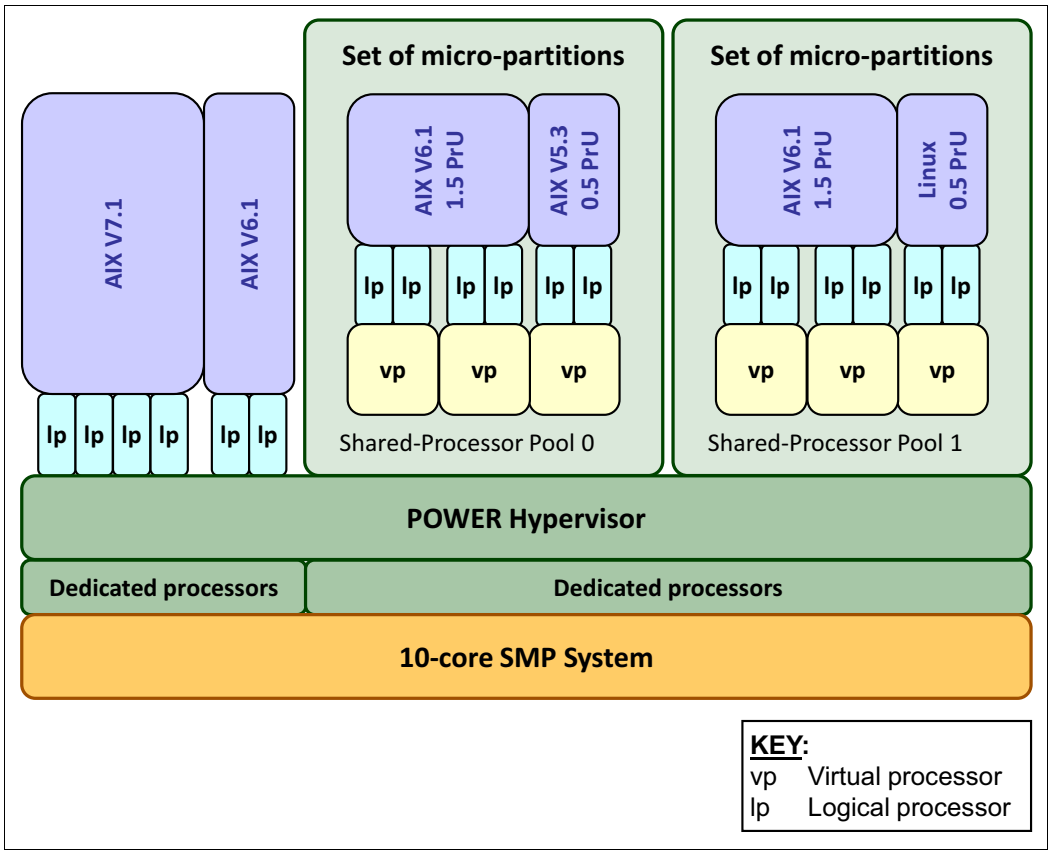

Figure 3-7 Logical partitioning concepts

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to partitions. The simultaneous multithreading feature in the POWER8 processor core allows the core to run instructions from two, four, or eight independent software threads simultaneously.

To support this feature, consider the concept of *logical processors*. The operating system (AIX or Linux) sees one physical core as two, four, or eight logical processors if the simultaneous multithreading feature is on. It can be turned off and on dynamically while the operating system is running (for AIX, run **`smtctl`**, for Linux, run **`ppc64_cpu --smt`**, and for IBM i, use **`QPRCMLTTSK`** system value). If simultaneous multithreading is off, each physical core is presented as one logical processor in AIX or Linux, and thus only one thread.

Shared dedicated mode

On POWER8 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 can help increase system usage without compromising the computing power for critical workloads in a dedicated processor.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000136

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 20 times the number of processing units that are assigned to the partition). From the POWER Hypervisor perspective, 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 dispatch cycle (10 ms), all partitions 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 simultaneous multithreading is turned on or off.

3.5.3 Multiple shared processor pools

Multiple shared processor pools (MSPPs) are supported on POWER8 processor-based servers. This capability allows a system administrator to create a set of micropartitions with the purpose of controlling the processor capacity that can be consumed from the physical shared processor pool.

Implementing MSPPs depends on a set of underlying techniques and technologies. Figure 3-8 shows an overview of the architecture of multiple shared processor pools.

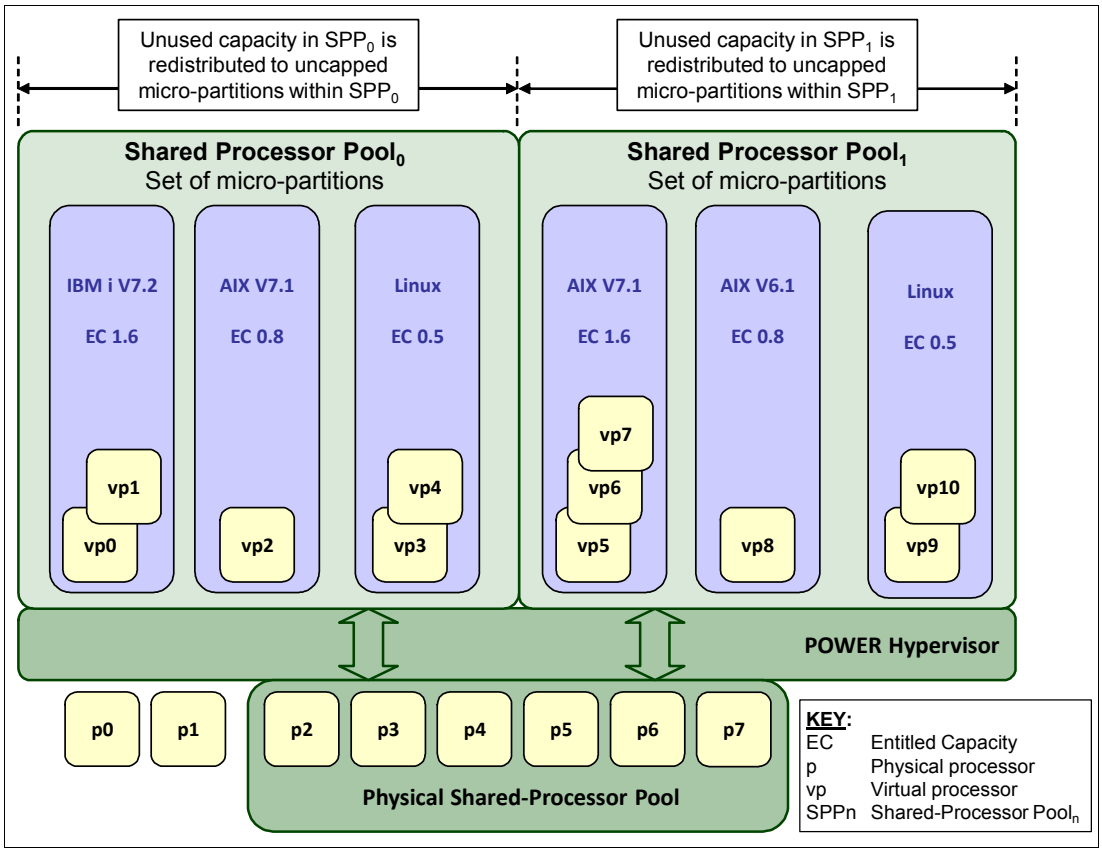

Figure 3-8 Overview of the architecture of multiple shared processor pools

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000137

Micropartitions 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 micropartitions 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 micropartition within a shared processor pool is ensured its processor entitlement plus any capacity that it might be allocated from the reserved pool capacity if the micropartition is uncapped.

If certain micropartitions in a shared processor pool do not use their capacity entitlement, the unused capacity is ceded and other uncapped micropartitions 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 micropartitions within that shared processor pool.

All Power Systems servers that support the multiple shared processor pools capability have a minimum of one (the default) shared processor pool and up to a maximum of 64 shared processor pools.

For more information and requirements about the shared storage pool, see *TotalStorage Productivity Center V3.3 Update Guide*, SG24-7490.

3.5.4 Virtual I/O Server

The VIOS 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 usage (for example, consolidation). In this case, the VIOS 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 VIOS eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The VIOS supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses. Figure 3-9 shows an overview of a VIOS configuration.

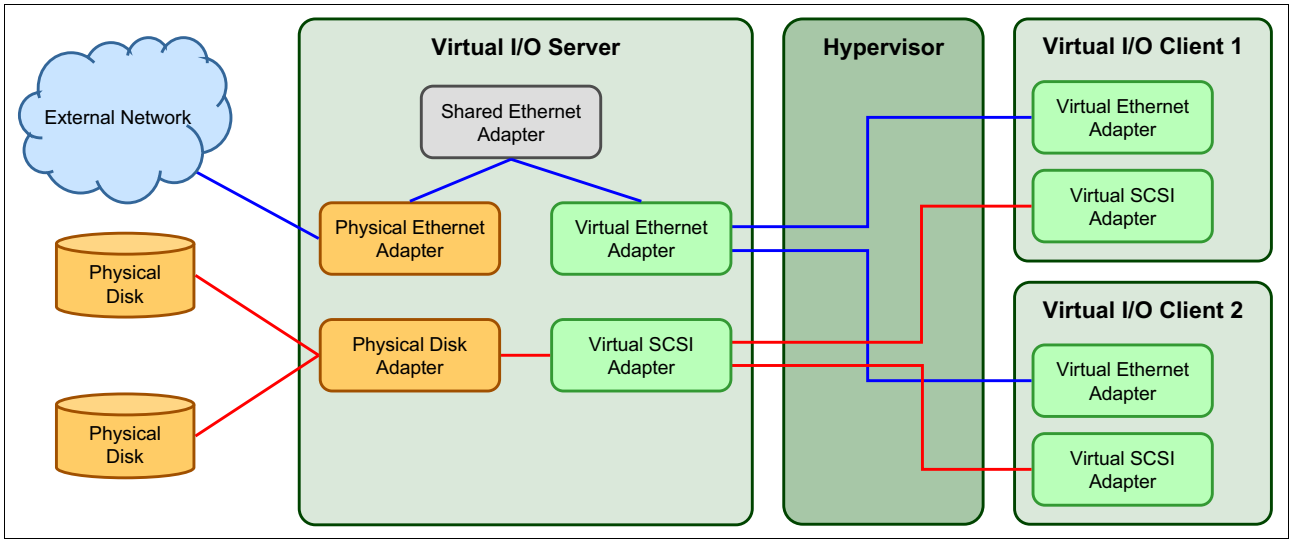

Figure 3-9 Architectural view of the VIOS

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000138

Because the VIOS is an operating system-based appliance server, redundancy for physical devices that are attached to the VIOS can be provided by using capabilities such as Multipath I/O and IEEE 802.3ad Link Aggregation.

Installation of the VIOS 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 VIOS, and is supported only in special VIOS partitions. Three major virtual devices are supported by the VIOS:

- Shared Ethernet Adapter
- Virtual SCSI
- Virtual Fibre Channel adapter

The Virtual Fibre Channel adapter is used with the NPIV feature, as described in 3.5.8, “Operating system support for PowerVM” on page 128.

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 POWER 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 by using a physical adapter. The Shared Ethernet Adapter service can be hosted only in the VIOS, 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 VIOS.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000139

Figure 3-10 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 VIOS that shares the physical access.

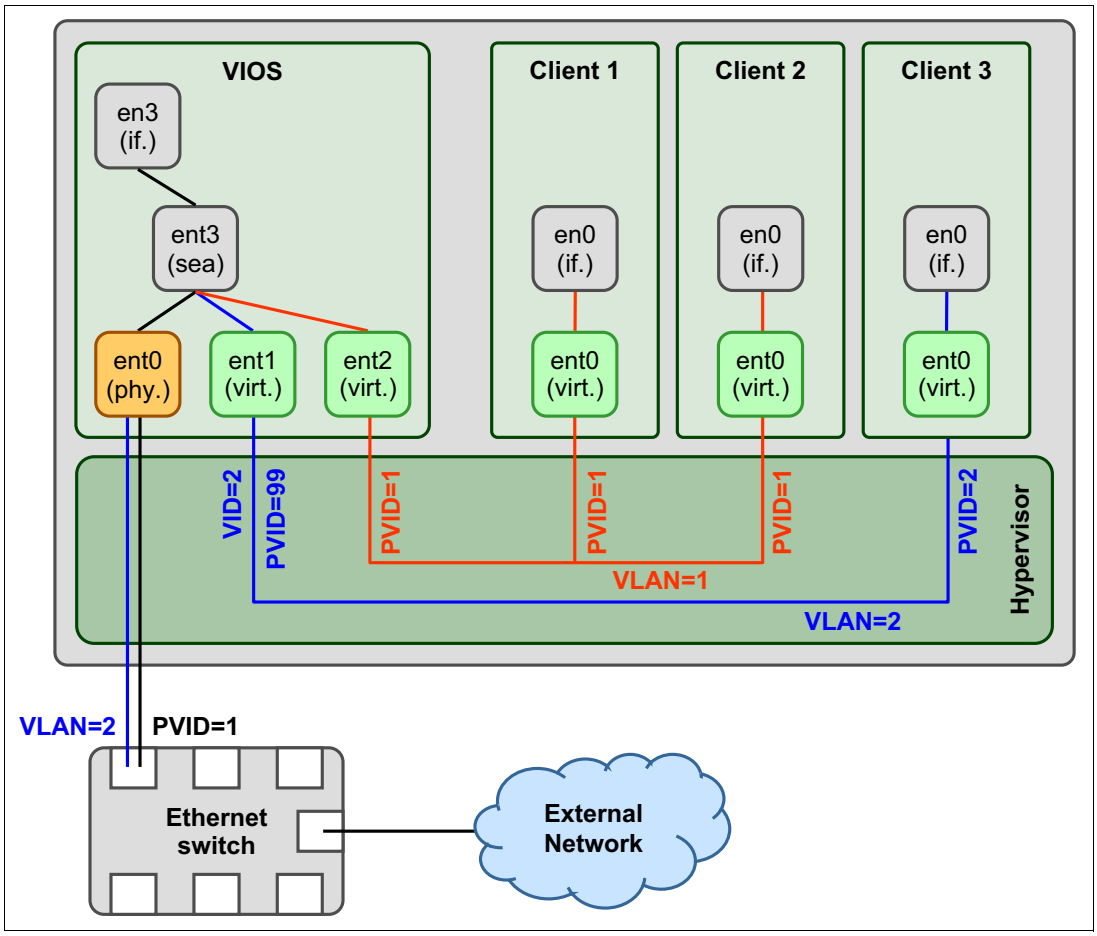

Figure 3-10 Architectural view of a Shared Ethernet Adapter

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 VIOS 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.

Virtual SCSI

Virtual SCSI is used to see a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a client/server relationship. The VIOS 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 that are provided by the VIOS as clients.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000140

The virtual I/O adapters (virtual SCSI server adapter and a virtual SCSI client adapter) are configured by using a managed console or through the Integrated Virtualization Manager on smaller systems. The virtual SCSI server (target) adapter is responsible for running any SCSI commands that it receives. It is owned by the VIOS 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 VIOS.

Physical disks that are presented to the Virtual/O Server can be exported and assigned to a client partition in various 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 VIOS 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.

For more information about specific storage devices that are supported for VIOS, see the following website:

http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html

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 VIOS partition that acts only as a pass-through, managing the data transfer through the POWER Hypervisor.

Each partition that uses 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 more information and requirements for NPIV, see the following resources:

- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

Virtual I/O Server functions

The VIOS has many features, including monitoring solutions and the following items:

- Support for Live Partition Mobility starting on POWER6 processor-based systems with the PowerVM Enterprise Edition. For more information about Live Partition Mobility, see 3.5.5, “PowerVM Live Partition Mobility” on page 126.
- Support for virtual SCSI devices that are 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 function.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000141

- System Planning Tool (SPT) and Workload Estimator, which are designed to ease the deployment of a virtualized infrastructure. For more information about the System Planning Tool, see 3.6, “System Planning Tool” on page 132.
- IBM Systems Director agent and several preinstalled IBM Tivoli® agents, such as the following examples:
  - IBM Tivoli Identity Manager, which allows easy integration into an existing Tivoli Systems Management infrastructure
  - IBM Tivoli Application Dependency Discovery Manager (ADDM), which creates and automatically maintains application infrastructure maps, including dependencies, change histories, and deep configuration values
- vSCSI enterprise reliability, availability, and serviceability (eRAS).
- Additional CLI statistics in **`svmon`**, **`vmstat`**, **`fcstat`**, and **`topas`**.
- The VIOS Performance Advisor tool provides advisory reports based on key performance metrics for various partition resources that are collected from the VIOS environment.
- Monitoring solutions to help manage and monitor the VIOS and shared resources. Commands and views provide additional metrics for memory, paging, processes, Fibre Channel HBA statistics, and virtualization.

For more information about the VIOS and its implementation, see *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940

3.5.5 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 or 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.

Live Partition Mobility provides systems management flexibility and improves system availability through the following functions:

- 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 that run on several small, underused 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000142

The PowerVM Server Evacuation function allows you to perform a server evacuation operation. Server Evacuation is used to move all migration-capable LPARs from one system to another if there are no active migrations in progress on the source or the target servers. Multiple migrations can occur based on the concurrency setting of the HMC. Migrations are performed as sets, with the next set of migrations starting when the previous set completes. Any upgrade or maintenance operations can be performed after all the partitions are migrated and the source system is powered off.

You can migrate all the migration-capable AIX, and Linux partitions from the source server to the destination server by running the following command from the HMC command line:

```
migrlpar -o m -m source_server -t target_server --all
```

Hardware and operating system requirements for Live Partition Mobility

PowerVM Live Partition Mobility requires a license for PowerVM Enterprise Edition, and it is supported in compliance with all operating systems that are compatible with POWER8 technology.

The VIOS partition itself cannot be migrated.

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility (Obsolete - See Abstract for Information)*, SG24-7460.

3.5.6 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 available only with the Enterprise version of PowerVM.

The physical memory of an IBM Power System can be assigned to multiple partitions in either dedicated or shared mode. The system administrator can 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 used to increase memory usage 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 also.

For more information regarding Active Memory Sharing, see *IBM PowerVM Virtualization Active Memory Sharing*, REDP-4470.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000143

3.5.7 Active Memory Deduplication

In a virtualized environment, the systems might have a considerable amount of duplicated information that is stored on RAM after each partition has its own operating system, and some of them might even share the same kinds of applications. On heavily loaded systems, this behavior might lead to a shortage of the available memory resources, forcing paging by the Active Memory Sharing partition operating systems, the Active Memory Deduplication pool, or both, which might decrease overall system performance

Active Memory Deduplication allows the POWER Hypervisor to map dynamically identical partition memory pages to a single physical memory page within a shared memory pool. This way enables a better usage of the Active Memory Sharing 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.

Active Memory Deduplication depends on the Active Memory Sharing feature being available, and it consumes CPU cycles that are donated by the Active Memory Sharing pool's VIOS partitions to identify deduplicated pages. The operating systems that are running on the Active Memory Sharing partitions can "hint" to the POWER Hypervisor that some pages (such as frequently referenced read-only code pages) are good for deduplication.

To perform deduplication, the hypervisor cannot compare every memory page in the Active Memory Sharing 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, a look-up of its signature is done in 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 a duplicate is found, the hypervisor remaps the partition memory to the existing memory page and returns the duplicate page to the Active Memory Sharing pool.

From the LPAR perspective, the Active Memory Deduplication feature is not apparent. If an LPAR attempts to modify a deduplicated page, the hypervisor grabs a free page from the Active Memory Sharing pool, copies the duplicate page contents into the new page, and maps the LPAR's reference to the new page so that the LPAR can modify its own unique page.

For more information regarding Active Memory Deduplication, see *Power Systems Memory Deduplication*, REDP-4827

3.5.8 Operating system support for PowerVM

At the time of writing, all PowerVM features are supported by the operating systems that are compatible with the POWER8 servers, except the Active Memory Expansion, which is not supported by IBM i and Linux.

PowerVM supports Big Endian (BE) and Little Endian (LE) mode. Table 3-4 lists the operating system levels for BE and LE modes:

Table 3-4 PowerVM BE and LE mode operating system requirements

Big Endian (BE)	Little Endian (LE)
► AIX Version 6 ► AIX Version 7	n/a
► IBM i 7	n/a

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000144

Big Endian (BE)	Little Endian (LE)
► Red Hat Enterprise Linux 6.5, or later ► Red Hat Enterprise Linux 7.0, or later	► Red Hat Enterprise Linux 7.1, or later
► SUSE Linux Enterprise Server 11 Service Pack 3, or later	► SUSE Linux Enterprise Server 12, or later
n/a	► Ubuntu 15.04, or later

Table 3-5 summarizes the PowerVM features that are supported by the operating systems that are compatible with the POWER8 processor-based servers.

Table 3-5 Virtualization features supported by AIX, IBM i, and Linux

Feature	AIX 6.1 TL9 SP1	AIX 7.1 TL03 SP1	IBM i 7.1 TR 9	IBM i 7.2 TR 1	RHEL 6.6	RHEL 7.1	SLES 11 SP3	SLES 12	Ubuntu 15.04
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 processor add/remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR memory add	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DLPAR memory remove	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	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
VIOS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Integrated Virtualization Manager	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Suspend/resume and hibernation[a]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Shared Storage Pools	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Thin provisioning	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Active Memory Sharing[b]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Active Memory Deduplication	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Live Partition Mobility	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simultaneous multithreading (SMT)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Active Memory Expansion	Yes	Yes	No	No	No	No	No	No	No

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000145

a. At the time of writing, Suspend/Resume is not available. Check with your IBM System Services Representative (SSR) for availability on POWER8 platforms.
b. At the time of writing, Active Memory Sharing when used with Live Partition Mobility is not supported. Check with your IBM SSR for availability on POWER8 platforms.

For more information about specific features for Linux, see the following website:

http://pic.dhe.ibm.com/infocenter/lnxinfo/v3r0m0/index.jsp?topic=%2Fliaam%2Fsupportedfeaturesforlinuxonpowersystemsservers.htm

3.5.9 Linux support

The IBM Linux Technology Center (LTC) contributes to the development of Linux by providing support for IBM hardware in Linux distributions. In particular, the LTC has available tools and code for the Linux communities to take advantage of the POWER8 technology and develop POWER8 optimized software.

For more information about specific Linux distributions, see the following website:

http://pic.dhe.ibm.com/infocenter/lnxinfo/v3r0m0/index.jsp?topic=%2Fliaam%2Fliaamdistros.htm

3.5.10 PowerVM simplification

With the availability of HMC V8R8.2.0 PowerVM simplification was introduced. Figure 3-11 shows new options in HMC V8R8.2.0 that are available for PowerVM simplification. Clicking **Manage PowerVM** opens a new window, where an user can view and manage all the aspects of a PowerVM configuration, such as SEA, virtual networks, and virtual storage using the graphical interface only.

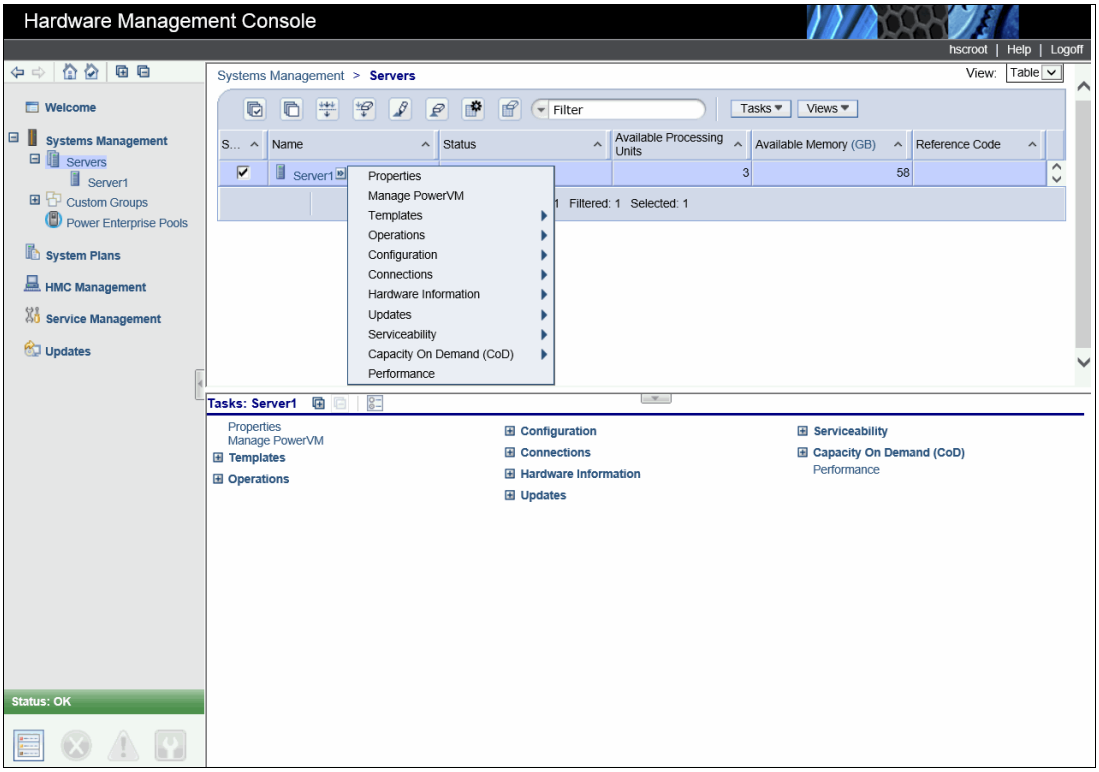

Figure 3-11 PowerVM new level tasks

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000146

Templates allow you to specify the configuration details for the system I/O, memory, storage, network, processor, and other partition resources. A user can use the predefined or captured templates that are available in the template library to deploy a system. Two types of templates are available in the template library:

- The System Template contains configuration details for system resources, such as system I/O, memory, storage, processors, network, and physical I/O adapters. You can use these system templates to deploy the system.
- The Partition Template contains configuration details for the partition resources. A user can configure logical partitions with the predefined templates or by creating a custom templates.

Using these options, a user can deploy a system, select a System Template, and click **Deploy**. After deploying a system, a user can choose a Partition Template from the library.

The Performance function opens the Performance and Capacity Monitoring window, as shown at the top of Figure 3-12.

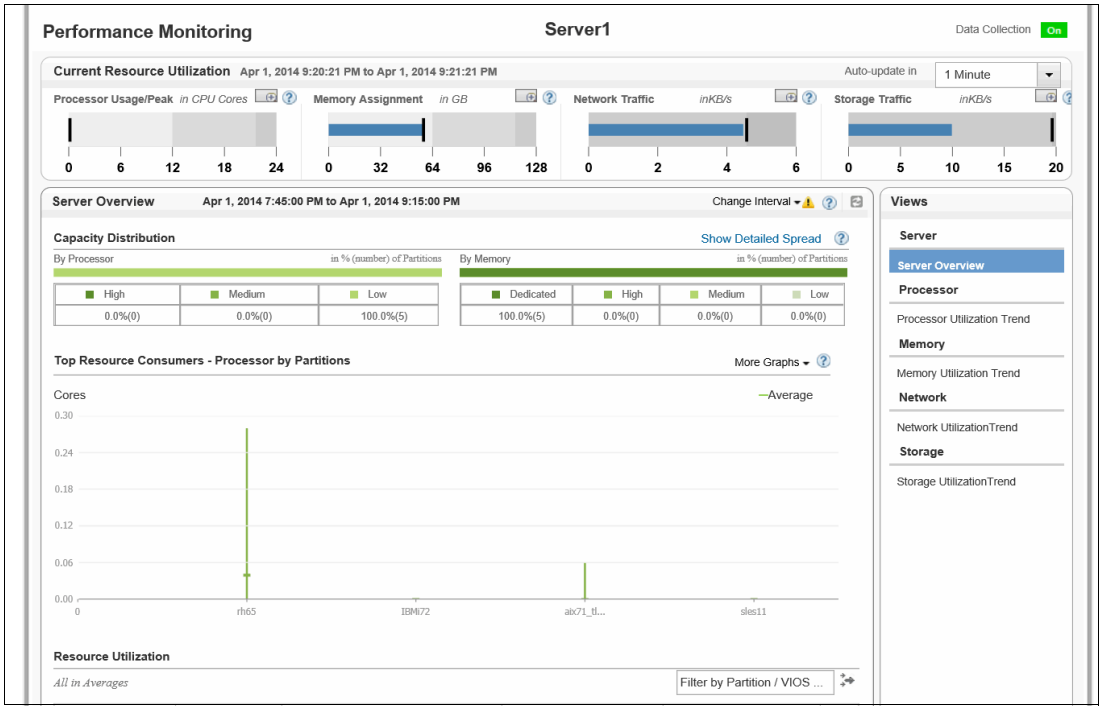

Figure 3-12 HMC performance monitoring CPU Memory assignment (top of the window)

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000147

Figure 3-13 shows the bottom of the window where performance and capacity data are presented in a graphical format.

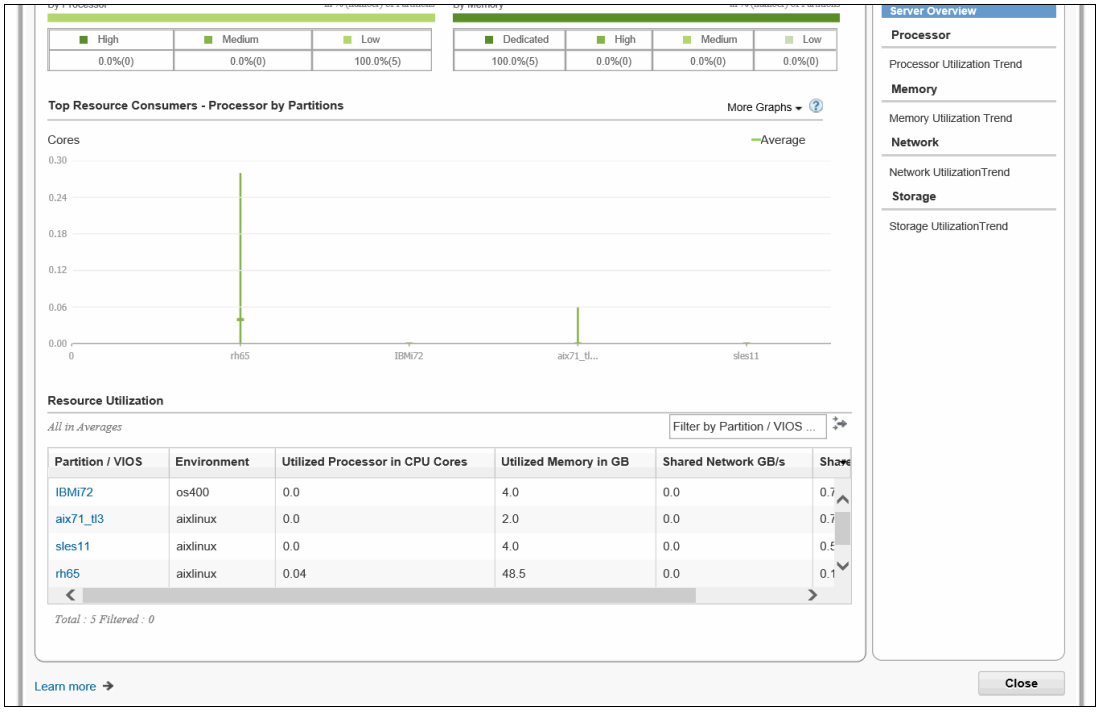

Figure 3-13 HMC performance monitoring CPU Memory assignment (bottom of the window)

3.6 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 the following 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 HMC or Integrated Virtualization Manager (IVM).

> **Automatically deploy:** Ask your IBM SSR or IBM 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 that is the same as what is specified in your `.sysplan` file.

You can create a new system configuration, or you can create a system configuration that is based on any of the following items:

- Performance data from an existing system that the new system replaces
- Performance estimates that anticipate future workloads that you must support
- Sample systems that you can customize to fit your needs

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000148

Integration between the System Planning Tool and both the Workload Estimator and IBM Performance Management 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 have.

Using the SPT is an effective way of documenting and backing up key system settings and partition definitions. With it, the user can create records of systems and export them to their personal workstation or backup system of choice. These same backups can then be imported back on to the same managed console when needed. This step 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 at the IBM System Planning Tool website:

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

3.7 IBM Power Virtualization Center

IBM Power Virtualization Center (IBM PowerVC) is designed to simplify the management of virtual resources in your Power Systems environment.

After the product code is loaded, the IBM PowerVC no-menus interface guides you through three simple configuration steps to register physical hosts, storage providers, and network resources, and start capturing and intelligently deploying your virtual machines (VMs), among the other tasks that are shown in the following list:

- Create virtual machines and then resize and attach volumes to them.
- Import existing virtual machines and volumes so they can be managed by IBM PowerVC.
- Monitor the usage of the resources that are in your environment.
- Migrate virtual machines while they are running (hot migration).
- Deploy images quickly to create virtual machines that meet the demands of your ever-changing business needs.

IBM PowerVC is built on OpenStack. OpenStack is an open source software that controls large pools of server, storage, and networking resources throughout a data center. Power VC can manage AIX, IBM i, and Linux VMs running under PowerVM virtualization and Linux VMs running under PowerKVM virtualization.

IBM PowerVC is available as IBM Power Virtualization Center Standard Edition.

In April, 2015 PowerVC V1.2.3 was announced. This new release supports all Power Systems servers that are built on IBM POWER8 technology.

PowerVC includes the following features and benefits:

- Virtual machine image capture, deployment, resizing, and management
- Policy-based VM placement to help improve usage and reduce complexity
- VM Mobility with placement policies to help reduce the burden on IT staff in a simplified GUI
- A management system that manages existing virtualization deployments
- Integrated management of storage, network, and compute, which simplifies administration

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000149

For more information about IBM PowerVC, see *IBM PowerVC Version 1.2 Introduction and Configuration*, SG24-8199.

3.8 IBM Power Virtualization Performance

IBM Power Virtualization Performance (IBM PowerVP™) for Power Systems is a new product that offers a performance view in to an IBM PowerVM virtualized environment running on the latest firmware of IBM Power Systems. It can show which virtual workloads are using specific physical resources on an IBM Power Systems server.

IBM PowerVP helps reduce time and complexity to find and display performance bottlenecks through a simple dashboard that shows the performance health of the system. It can help simplify both prevention and troubleshooting and thus reduce the cost of performance management.

It assists you in the following ways:

- Shows workloads in real time, which highlights possible problems or bottlenecks (overcommitted resources)
- Helps better use virtualized IBM Power System servers by showing the distribution of workloads
- Can replay saved historical data
- Helps with the resolution of performance-related issues
- Helps to proactively address future issues that can affect performance

IBM PowerVP is integrated with the POWER Hypervisor and collects performance data directly from PowerVM Hypervisor, which offers the most accurate performance information about virtual machines running on IBM Power Systems. This performance information is displayed on a real-time, continuous GUI dashboard and it is also available for historical review.

Features of IBM PowerVP include the following ones:

- Real-time, continuous graphical monitor (dashboard) that delivers an easy-to-read display showing the overall performance health of the Power Systems server.
- Customizable performance thresholds that enable you to customize the dashboard to match your monitoring requirements.
- Historical statistics that enable you to go back in time and replay performance data sequences to discover performance bottlenecks.
- System-level performance views that show all VMs and how they are using real system resources.
- Virtual machine drill down, which gives you more performance details for each VM, displaying detailed information about various resources, such as CPU, memory, and disk activity.
- Support for all virtual machine types, including AIX, IBM i, and Linux.
- Background data collection, which enables performance data to be collected when the GUI is not active.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000150

IBM PowerVP even allows an administrator to drill down and view specific adapter, bus, or CPU usage. An administrator can see the hardware adapters and how much workload is placed on them. IBM PowerVP provides both an overall and detailed view of IBM Power System server hardware, so it is easy to see how virtual machines are consuming resources. For more information about this topic, see the following website:

http://www.ibm.com/systems/power/software/performance/

The latest Power VP 1.1.3 release has been enhanced with these other new features and support:

- A new capability to export PowerVP performance data to an external repository
- Integration with the VIOS performance advisor
- New thresholds and alert
- The ability to run the PowerVM user interface in a browser
- Support for monitoring RHEL 7.1, SLES 12, and Ubuntu 15.04 guests running under PowerVM Little Endian (LE) mode

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000151

4

Reliability, availability, and serviceability

This chapter provides information about IBM Power Systems reliability, availability, and serviceability (RAS) design and features.

The elements of RAS can be described as follows:

Reliability	Indicates how infrequently a defect or fault in a server occurs
Availability	Indicates how infrequently the functioning of a system or application is impacted by a fault or defect
Serviceability	Indicates how well faults and their effects are communicated to system managers and how efficiently and nondisruptively the faults are repaired

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000153

4.1 Introduction

The POWER8 processor modules support an enterprise level of reliability and availability. The processor design has extensive error detection and fault isolation (ED/FI) capabilities to allow for a precise analysis of faults, whether they are hard or soft. They use advanced technology, including stacked latches and Silicon-on-Insulator (SOI) technology, to reduce susceptibility to soft errors, and advanced design features within the processor for correction or try again after soft error events. In addition, the design incorporates spare capacity that is integrated into many elements to tolerate certain faults without requiring an outage or parts replacement. Advanced availability techniques are used to mitigate the impact of other faults that are not directly correctable in the hardware.

Features within the processor and throughout systems are incorporated to support design verification. During the design and development process, subsystems go through rigorous verification and integration testing processes by using these features. During system manufacturing, systems go through a thorough testing process to help ensure high product quality levels, again taking advantage of the designed ED/FI capabilities.

Fault isolation and recovery of the POWER8 processor and memory subsystems are designed to use a dedicated service processor and are meant to be largely independent of any operating system or application deployed.

The Power S814 and Power S824 POWER8 processor-based servers are designed to support a “scale-out” system environment consisting of multiple systems working in concert. In such environments, application availability is enhanced by the superior availability characteristics of each system.

4.1.1 RAS enhancements of POWER8 processor-based scale-out servers

The Power S814 and Power S824 servers, in addition to being built on advanced RAS characteristics of the POWER8 processor, offer reliability and availability features that often are not seen in such scale-out servers.

Some of these features are improvements for POWER8 or features that were found previously only in higher-end Power Systems.

Here is a brief summary of these features:

- Processor Enhancements Integration

  POWER8 processor chips are implemented using 22 nm technology and integrated on to SOI modules.

  The processor design now supports a spare data lane on each fabric bus, which is used to communicate between processor modules. A spare data lane can be substituted for a failing one dynamically during system operation.

  A POWER8 processor module has improved performance compared to POWER7+, including support of a maximum of 12 cores compared to a maximum of eight cores in POWER7+. This is because doing more work with less hardware in a system supports greater reliability.

  The processor module integrates a new On Chip Controller (OCC). This OCC is used to handle Power Management and Thermal Monitoring without the need for a separate controller, which was required in POWER7+. In addition, the OCC can also be programmed to run other RAS-related functions independent of any host processor.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000154

The memory controller within the processor is redesigned. From a RAS standpoint, the ability to use a replay buffer to recover from soft errors is added.

- I/O Subsystem

  The POWER8 processor now integrates PCIe controllers. PCIe slots that are directly driven by PCIe controllers can be used to support I/O adapters directly in the systems or, as a statement of direction, be used to attach external I/O drawers. For greater I/O capacity, the POWER8 processor-based scale-out servers also support a PCIe switch to provide additional integrated I/O capacity.

  These integrated I/O adapters can be repaired in these servers concurrently, which is an improvement over comparable POWER7/7+ systems that did not allow for adapter "hot-plug."

- Memory Subsystem

  Custom DIMMs (CDIMMS) are used, which, in addition to the ability to correct a single DRAM fault within an error-correcting code (ECC) word (and then an additional bit fault) to avoid unplanned outages, also contain a spare DRAM module per port (per nine DRAMs for x8 DIMMs), which can be used to avoid replacing memory.

4.2 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, and architecture-based approach is designed to improve overall system reliability with each successive generation of system offerings. Reliability can be improved in primarily three ways:

- Reducing the number of components
- Using higher reliability grade parts
- Reducing the stress on the components

In the POWER8 systems, elements of all three are used to improve system reliability.

4.2.1 Designed for reliability

Systems that are 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 reduce the opportunity for system failures. The POWER8 chip has more cores per processor module, and the I/O Hub Controller function is integrated in the processor module, which generates a PCIe BUS directly from the Processor module. Parts selection also plays a critical role in overall system reliability.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000155

IBM uses stringent design criteria to select server grade components that are extensively tested and qualified to meet and exceed a minimum design life of seven years. By selecting higher reliability grade components, the frequency of all failures is lowered, and wear-out is not expected within the operating system life. Component failure rates can be further improved by burning in select components or running the system before shipping it to the client. This period of high stress removes the weaker components with higher failure rates, that is, it cuts off the front end of the traditional failure rate bathtub curve (see Figure 4-1).

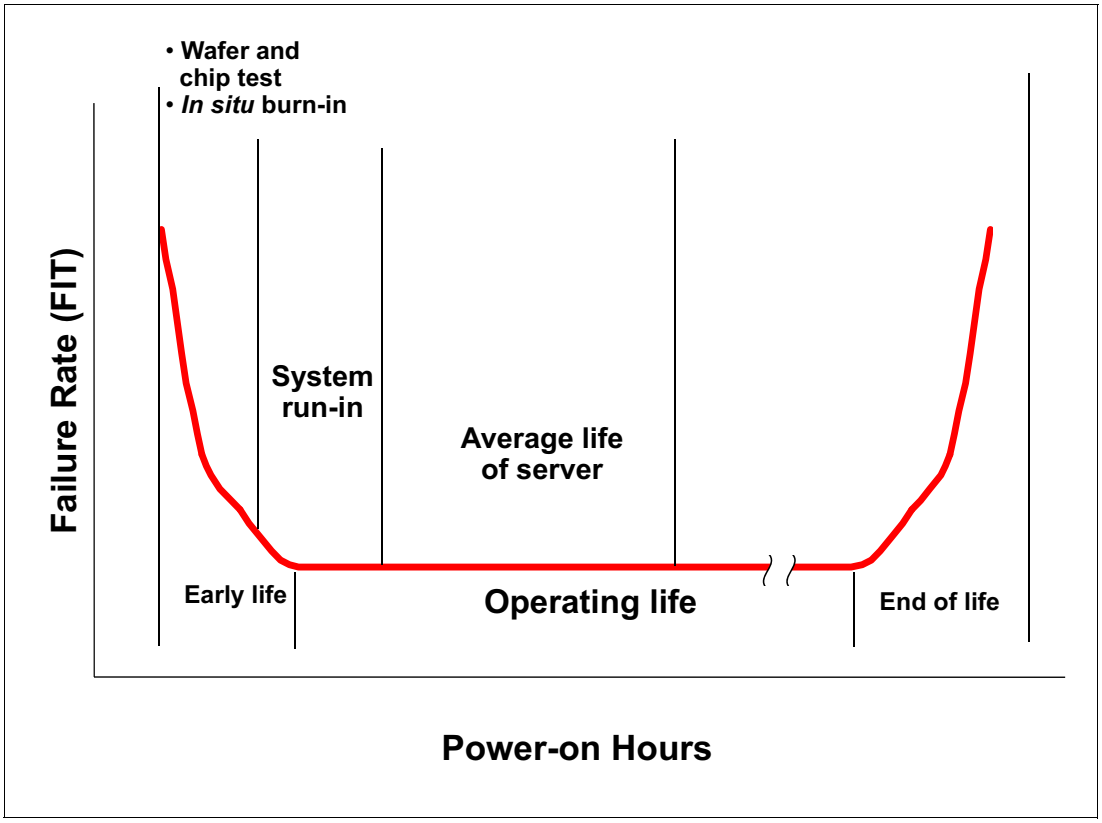

Figure 4-1 Failure rate bathtub curve

4.2.2 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. Large decreases in component reliability are directly correlated to relatively small increases in temperature. All POWER processor-based systems are packaged to ensure adequate cooling. Critical system components, such as the POWER8 processor chips, are positioned on the system board so that they receive clear air flow during operation. POWER8 systems use a premium fan with an extended life to further reduce overall system failure rate and provide adequate cooling for the critical system components.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000156

4.3 Processor/Memory availability details

The more reliable a system or subsystem is, the more available it should be. Nevertheless, considerable effort is made to design systems that can detect faults that do occur and take steps to minimize or eliminate the outages that are associated with them. These design capabilities extend availability beyond what can be obtained through the underlying reliability of the hardware.

This design for availability begins with implementing an architecture for ED/FI.

First-Failure Data Capture (FFDC) is the capability of IBM hardware and microcode to continuously monitor hardware functions. Within the processor and memory subsystem, detailed monitoring is done by circuits within the hardware components themselves. Fault information is gathered into fault isolation registers (FIRs) and reported to the appropriate components for handling.

Processor and memory errors that are recoverable in nature are typically reported to the dedicated service processor built into each system. The dedicated service processor then works with the hardware to determine the course of action to be taken for each fault.

4.3.1 Correctable error introduction

Intermittent or soft errors are typically tolerated within the hardware design by using error correction code or advanced techniques to try operations again after a fault.

Tolerating a correctable solid fault runs the risk that the fault aligns with a soft error and causes an uncorrectable error situation. There is also the risk that a correctable error is predictive of a fault that continues to worsen over time, resulting in an uncorrectable error condition.

You can predictively deallocate a component to prevent correctable errors from aligning with soft errors or other hardware faults and causing uncorrectable errors to avoid such situations. However, unconfiguring components, such as processor cores or entire caches in memory, can reduce the performance or capacity of a system. This in turn typically requires that the failing hardware is replaced in the system. The resulting service action can also temporarily impact system availability.

To avoid such situations in solid faults in POWER8, processors or memory might be candidates for correction by using the “self-healing” features built into the hardware, such as taking advantage of a spare DRAM module within a memory DIMM, a spare data lane on a processor or memory bus, or spare capacity within a cache module.

When such self-healing is successful, the need to replace any hardware for a solid correctable fault is avoided. The ability to predictively unconfigure a processor core is still available for faults that cannot be repaired by self-healing techniques or because the sparing or self-healing capacity is exhausted.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000157

4.3.2 Uncorrectable error introduction

An uncorrectable error can be defined as a fault that can cause incorrect instruction execution within logic functions, or an uncorrectable error in data that is stored in caches, registers, or other data structures. In less sophisticated designs, a detected uncorrectable error nearly always results in the termination of an entire system. More advanced system designs in some cases might be able to terminate just the application by using the hardware that failed. Such designs may require that uncorrectable errors are detected by the hardware and reported to software layers, and the software layers must then be responsible for determining how to minimize the impact of faults.

The advanced RAS features that are built in to POWER8 processor-based systems handle certain “uncorrectable” errors in ways that minimize the impact of the faults, even keeping an entire system up and running after experiencing such a failure.

Depending on the fault, such recovery may use the virtualization capabilities of PowerVM in such a way that the operating system or any applications that are running in the system are not impacted or must participate in the recovery.

4.3.3 Processor Core/Cache correctable error handling

Layer 2 (L2) and Layer 3 (L3) caches and directories can correct single bit errors and detect double bit errors (SEC/DEC ECC). Soft errors that are detected in the level 1 caches are also correctable by a try again operation that is handled by the hardware. Internal and external processor “fabric” busses have SEC/DEC ECC protection as well.

SEC/DEC capabilities are also included in other data arrays that are not directly visible to customers.

Beyond soft error correction, the intent of the POWER8 design is to manage a solid correctable error in an L2 or L3 cache by using techniques to delete a cache line with a persistent issue, or to repair a column of an L3 cache dynamically by using spare capability.

Information about column and row repair operations is stored persistently for processors, so that more permanent repairs can be made during processor reinitialization (during system reboot, or individual Core Power on Reset using the Power On Reset Engine.)

4.3.4 Processor Instruction Retry and other try again techniques

Within the processor core, soft error events might occur that interfere with the various computation units. When such an event can be detected before a failing instruction is completed, the processor hardware might be able to try again the operation by using the advanced RAS feature that is known as *Processor Instruction Retry.*

Processor Instruction Retry allows the system to recover from soft faults that otherwise result in outages of applications or the entire server.

Try again techniques are used in other parts of the system as well. Faults that are detected on the memory bus that connects processor memory controllers to DIMMs can be tried again. In POWER8 systems, the memory controller is designed with a replay buffer that allows memory transactions to be tried again after certain faults internal to the memory controller faults are detected. This complements the try again abilities of the memory buffer module.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000158

4.3.5 Alternative processor recovery and Partition Availability Priority

If Processor Instruction Retry for a fault within a core occurs multiple times without success, the fault is considered to be a solid failure. In some instances, PowerVM can work with the processor hardware to migrate a workload running on the failing processor to a spare or alternative processor. This migration is accomplished by migrating the pertinent processor core state from one core to another with the new core taking over at the instruction that failed on the faulty core. Successful migration keeps the application running during the migration without needing to terminate the failing application.

Successful migration requires that there is sufficient spare capacity that is available to reduce the overall processing capacity within the system by one processor core. Typically, in highly virtualized environments, the requirements of partitions can be reduced to accomplish this task without any further impact to running applications.

In systems without sufficient reserve capacity, it might be necessary to terminate at least one partition to free resources for the migration. In advance, PowerVM users can identify which partitions have the highest priority and which do not. When you use this Partition Priority feature of PowerVM, if a partition must be terminated for alternative processor recovery to complete, the system can terminate lower priority partitions to keep the higher priority partitions up and running, even when an unrecoverable error occurred on a core running the highest priority workload.

Partition Availability Priority is assigned to partitions by using a weight value or integer rating. The lowest priority partition is rated at 0 (zero) and the highest priority partition is rated at 255. The default value is set to 127 for standard partitions and 192 for Virtual I/O Server (VIOS) partitions. Priorities can be modified through the Hardware Management Console (HMC).

4.3.6 Core Contained Checkstops and other PowerVM error recovery

PowerVM can handle certain other hardware faults without terminating applications, such as an error in certain data structures (faults in translation tables or lookaside buffers).

Other core hardware faults that alternative processor recovery or Processor Instruction Retry cannot contain might be handled in PowerVM by a technique called Core Contained Checkstops. This technique allows PowerVM to be signaled when such faults occur and terminate code in use by the failing processor core (typically just a partition, although potentially PowerVM itself if the failing instruction were in a critical area of PowerVM code).

Processor designs without Processor Instruction Retry typically must resort to such techniques for all faults that can be contained to an instruction in a processor core.

4.3.7 Cache uncorrectable error handling

If a fault within a cache occurs that cannot be corrected with SEC/DED ECC, the faulty cache element is unconfigured from the system. Typically, this is done by purging and deleting a single cache line. Such purge and delete operations are contained within the hardware itself, and prevent a faulty cache line from being reused and causing multiple errors.

During the cache purge operation, the data that is stored in the cache line is corrected where possible. If correction is not possible, the associated cache line is marked with a special ECC code that indicates that the cache line itself has bad data.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000159

Nothing within the system terminates just because such an event is encountered. Rather, the hardware monitors the usage of pages with marks. If such data is never used, hardware replacement is requested, but nothing terminates as a result of the operation. Software layers are not required to handle such faults.

Only when data is loaded to be processed by a processor core, or sent out to an I/O adapter, is any further action needed. In such cases, if data is used as owned by a partition, then the partition operating system might be responsible for terminating itself or just the program using the marked page. If data is owned by the hypervisor, then the hypervisor might choose to terminate, resulting in a system-wide outage.

However, the exposure to such events is minimized because cache-lines can be deleted, which eliminates repetition of an uncorrectable fault that is in a particular cache-line.

4.3.8 Other processor chip functions

Within a processor chip, there are other functions besides just processor cores.

POWER8 processors have built-in accelerators that can be used as application resources to handle such functions as random number generation. POWER8 also introduces a controller for attaching cache-coherent adapters that are external to the processor module. The POWER8 design contains a function to “freeze” the function that is associated with some of these elements, without taking a system-wide checkstop. Depending on the code using these features, a “freeze” event might be handled without an application or partition outage.

As indicated elsewhere, single bit errors, even solid faults, within internal or external processor “fabric busses”, are corrected by the error correction code that is used. POWER8 processor-to-processor module fabric busses also use a spare data-lane so that a single failure can be repaired without calling for the replacement of hardware.

4.3.9 Other fault error handling

Not all processor module faults can be corrected by these techniques. Therefore, a provision is still made for some faults that cause a system-wide outage. In such a “platform” checkstop event, the ED/FI capabilities that are built in to the hardware and dedicated service processor work to isolate the root cause of the checkstop and unconfigure the faulty element were possible so that the system can reboot with the failed component unconfigured from the system.

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.

The auto-restart (reboot) option must be enabled from the Advanced System Management Interface (ASMI) or from the Control (Operator) Panel.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000160

4.3.10 Memory protection

POWER8 processor-based systems have a three-part memory subsystem design. This design consists of two memory controllers in each processor module, which communicate to buffer modules on memory DIMMS through memory channels and access the DRAM memory modules on DIMMs, as shown in Figure 4-2.

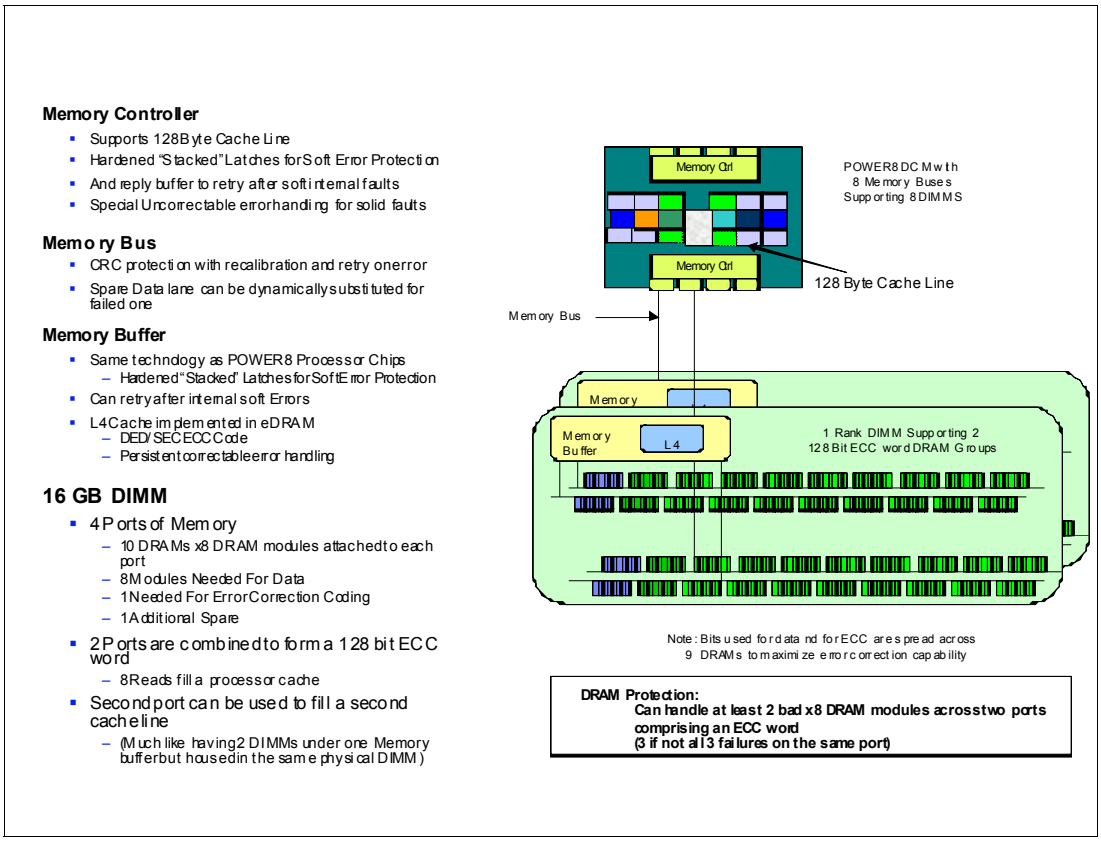

Figure 4-2 Memory protection features

The memory buffer chip is made by the same 22 nm technology that is used to make the POWER8 processor chip, and the memory buffer chip incorporates the same features in the technology to avoid soft errors. It implements a try again for many internally detected faults. This function complements a replay buffer in the memory controller in the processor, which also handles internally detected soft errors.

The bus between a processor memory controller and a DIMM uses CRC error detection that is coupled with the ability to try soft errors again. The bus features dynamic recalibration capabilities plus a spare data lane that can be substituted for a failing bus lane through the recalibration process.

The buffer module implements an integrated L4 cache using eDRAM technology (with soft error hardening) and persistent error handling features.

The memory buffer on each DIMM has four ports for communicating with DRAM modules. The 16 GB DIMM, for example, has one rank that is composed of four ports of x8 DRAM modules, each port containing 10 DRAM modules.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000161

For each such port, there are eight DRAM modules worth of data (64 bits) plus another DRAM module's worth of error correction and other such data. There is also a spare DRAM module for each port that can be substituted for a failing port.

Two ports are combined into an ECC word and supply 128 bits of data. The ECC that is deployed can correct the result of an entire DRAM module that is faulty. (This is also known as Chipkill correction). Then, it can correct at least an additional bit within the ECC word.

The additional spare DRAM modules are used so that when a DIMM experiences a Chipkill event within the DRAM modules under a port, the spare DRAM module can be substituted for a failing module, avoiding the need to replace the DIMM for a single Chipkill event.

Depending on how DRAM modules fail, it might be possible to tolerate up to four DRAM modules failing on a single DIMM without needing to replace the DIMM, and then still correct an additional DRAM module that is failing within the DIMM.

There are other DIMMs offered with these systems. A 32 GB DIMM has two ranks, where each rank is similar to the 16 GB DIMM with DRAM modules on four ports, and each port has ten x8 DRAM modules.

In addition, there is a 64 GB DIMM that is offered through x4 DRAM modules that are organized in four ranks.

In addition to the protection that is provided by the ECC and sparing capabilities, the memory subsystem also implements scrubbing of memory to identify and correct single bit soft-errors. Hypervisors are informed of incidents of single-cell persistent (hard) faults for deallocation of associated pages. However, because of the ECC and sparing capabilities that are used, such memory page deallocation is not relied upon for repair of faulty hardware,

Finally, should an uncorrectable error in data be encountered, the memory that is impacted is marked with a special uncorrectable error code and handled as described for cache uncorrectable errors.

4.3.11 I/O subsystem availability and Enhanced Error Handling

Usage of multi-path I/O and VIOS for I/O adapters and RAID for storage devices should be used to prevent application outages when I/O adapter faults occur.

To permit soft or intermittent faults to be recovered without failover to an alternative device or I/O path, Power Systems hardware supports *Enhanced Error Handling* (EEH) for I/O adapters and PCIe bus faults.

EEH allows EEH-aware device drivers to try again after certain non-fatal I/O events to avoid failover, especially in cases where a soft error is encountered. EEH also allows device drivers to terminate if there is an intermittent hard error or other unrecoverable errors, while protecting against reliance on data that cannot be corrected. This action typically is done by "freezing" access to the I/O subsystem with the fault. Freezing prevents data from flowing to and from an I/O adapter and causes the hardware/firmware to respond with a defined error signature whenever an attempt is made to access the device. If necessary, a special uncorrectable error code may be used to mark a section of data as bad when the freeze is first initiated.

In POWER8 processor-based systems, the external I/O hub and bridge adapters were eliminated in favor of a topology that integrates PCIe Host Bridges into the processor module itself. PCIe busses that are generated directly from a host bridge may drive individual I/O slots

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000162

or a PCIe switch. The integrated PCIe controller supports try again (end-point error recovery) and freezing.

IBM device drivers under AIX are fully EEH-capable. For Linux under PowerVM, EEH support extends to many frequently used devices. There might be various third-party PCI devices that do not provide native EEH support.

4.4 Serviceability

The purpose of serviceability is to repair the system while attempting to minimize or eliminate service cost (within budget objectives) and maintaining application availability and high customer satisfaction. Serviceability includes system installation, miscellaneous equipment specification (MES) (system upgrades/downgrades), and system maintenance/repair. Depending on the system and warranty contract, service might be performed by the customer, an IBM System Services Representative (SSR), or an authorized warranty service provider.

The serviceability features that are delivered in this system provide a highly efficient service environment by incorporating the following attributes:

- Design for customer setup (CSU), customer installed features (CIF), and customer-replaceable units (CRU)
- ED/FI incorporating FFDC
- Converged service approach across multiple IBM server platforms
- Concurrent Firmware Maintenance (CFM)

This section provides an overview of how these attributes contribute to efficient service in the progressive steps of error detection, analysis, reporting, notification, and repair found in all POWER processor-based systems.

4.4.1 Detecting introduction

The first and most crucial component of a solid serviceability strategy is the ability to detect accurately and effectively 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 has no opportunity to evaluate and act if necessary. POWER processor-based systems employ IBM System z® server-inspired error detection mechanisms, extending from processor cores and memory to power supplies and hard disk drives (HDDs).

4.4.2 Error checkers, fault isolation registers, and First-Failure Data Capture

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 that is coupled with Processor Instruction Retry and bus try again, to ECC correction on caches and system buses.

Within the processor/memory subsystem error-checker, error-checker signals are captured and stored in hardware 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 diagnostic tests can be deterministic so that for every check station, the unique error domain for that checker

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000163

is defined and mapped to field-replaceable units (FRUs) that can be repaired when necessary.

Integral to the Power Systems design is the concept of FFDC. FFDC is a technique that involves sufficient error checking stations and co-ordination of faults so that faults are detected and the root cause of the fault is isolated. FFDC also expects that necessary fault information can be collected at the time of failure without needing re-create 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 isolated at the time of the failure without intervention by an IBM SSR. For all faults, good FFDC design still makes failure information available to the IBM SSR, and this information can be used to confirm the automatic diagnosis. More detailed information can be collected by an IBM SSR for rare cases where the automatic diagnosis is not adequate for fault isolation.

4.4.3 Service processor

In POWER8 processor-based systems with a dedicated service processor, the dedicated service processor is primarily responsible for fault analysis of processor/memory errors.

The service processor is a microprocessor that is powered separately from the main instruction processing complex.

In addition to FFDC functions, the service processor performs many serviceability functions:

- Several remote power control options
- Reset and boot features
- Environmental monitoring

  The service processor interfaces with the OCC function, which monitors the server's built-in temperature sensors and sends instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range. By using a designed 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 the following circumstances:

  - 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 shut down a system in the following circumstances:
    - The 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.
- POWER Hypervisor (system firmware) and HMC connection surveillance.

  The service processor monitors the operation of the firmware during the boot process, and also monitors the hypervisor for termination. The hypervisor monitors the service processor and can perform a reset and reload if it detects the loss of the service processor. If the reset/reload operation does not correct the problem with the service

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000164

processor, the hypervisor notifies the operating system, and then the operating system can then take appropriate action, including calling for service. The FSP also monitors the connection to the HMC and can report loss of connectivity to the operating system partitions for system administrator notification.

- Uncorrectable error recovery

  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.

  The auto-restart (reboot) option must be enabled from the ASMI or from the Control (Operator) Panel.

- Concurrent access to the service processors menus of the ASMI

  This access allows nondisruptive abilities to change system default parameters, interrogate service processor progress and error logs, and set and reset service indicators (Light Path for low-end servers), and access all service processor functions without having to power down the system to the standby state. The administrator or IBM SSR dynamically can access the menus from any web browser-enabled console that is attached to the Ethernet service network, concurrently with normal system operation. Some options, such as changing the hypervisor type, do not take effect until the next boot.

- Management of the interfaces for connecting uninterruptible power source systems to the POWER processor-based systems and performing timed power-on (TPO) sequences.

4.4.4 Diagnosing

General diagnostic objectives are to detect and identify problems so that they can be resolved quickly. The IBM diagnostic strategy includes the following elements:

- Provide a common error code format that is equivalent to a system reference code, system reference number, checkpoint, or firmware error code.
- Provide fault detection and problem isolation procedures. Support a remote connection ability that is used by the IBM Remote Support Center or IBM Designated Service.
- Provide interactive intelligence within the diagnostic tests with detailed online failure information while connected to IBM back-end system.

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.

Because of the FFDC technology that is designed in to IBM servers, re-creating diagnostic tests for failures or requiring user intervention is not necessary. Solid and intermittent errors are designed to be correctly detected and isolated at the time that the failure occurs. Runtime and boot time diagnostic tests fall into this category.

Boot time

When an IBM Power Systems server powers up, the service processor initializes the system hardware. Boot-time diagnostic testing uses a multitier approach for system validation, starting with managed low-level diagnostic tests that are supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines.

To minimize boot time, the system determines which of the diagnostic tests are required to be started to ensure correct operation, which is based on the way that the system was powered off, or on the boot-time selection menu.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000165

Host Boot IPL

In POWER8, the initialization process during IPL changed. The Flexible Service Processor (FSP) is no longer the only instance that initializes and runs the boot process. With POWER8, the FSP initializes the boot processes, but on the POWER8 processor itself, one part of the firmware is running and performing the Central Electronics Complex chip initialization. A new component that is called the PNOR chip stores the Host Boot firmware and the Self Boot Engine (SBE) is an internal part of the POWER8 chip itself and is used to boot the chip.

With this Host Boot initialization, new progress codes are available. An example of an FSP progress code is C1009003. During the Host Boot IPL, progress codes, such as CC009344, appear.

If there is a failure during the Host Boot process, a new Host Boot System Dump is collected and stored. This type of memory dump includes Host Boot memory and is offloaded to the HMC when it is available.

Run time

All Power Systems servers can monitor critical system components during run time, and they can take corrective actions when recoverable faults occur. The IBM hardware error-check architecture can report non-critical errors in the Central Electronics Complex in an *out-of-band* communications path to the service processor without affecting system performance.

A significant part of IBM runtime diagnostic capabilities originate with the service processor. Extensive diagnostic and fault analysis routines were 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 by using logic that is 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 the following items:

- 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, diagnostic tests are best performed by operating system-specific drivers, most notably adapters or I/O devices that are owned directly by a logical partition. In these cases, the operating system device driver often works 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. In non-HMC managed servers, the OS can start the Call Home application to report the service event to IBM. For optional HMC managed servers, the event is reported to the HMC, which can initiate the Call Home request to IBM. I/O devices can also include specific exercisers that can be started by the diagnostic facilities for problem recreation (if required by service procedures).

4.4.5 Reporting

In the unlikely event that a system hardware or environmentally induced failure is diagnosed, IBM Power Systems servers report the error through various 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000166

Using the Call Home infrastructure, the system automatically can send an alert through a phone line to a pager, or call for service if there is a critical system failure. A hardware fault also illuminates the amber system fault LED, which is on the system unit, to alert the user of an internal hardware problem.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000167

On POWER8 processor-based servers, hardware and software failures are recorded in the system log. When a management console is attached, an ELA routine analyzes the error, forwards the event to the Service Focal Point™ (SFP) application running on the management console, and can notify the system administrator that it isolated a likely cause of the system problem. The service processor event log also records unrecoverable checkstop conditions, forwards them to the SFP application, and notifies the system administrator. After the information is logged in the SFP application, if the system is correctly 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 also contains the client contact information as defined in the IBM Electronic Service Agent (ESA) guided setup wizard. With the new HMC V8R8.1.0 a Serviceable Event Manager is available to block problems from being automatically transferred to IBM. For more information about this topic, see “Service Event Manager” on page 168.

Error logging and analysis

When the root cause of an error is identified by a fault isolation component, an error log entry is created with basic data, such as the following examples:

- An error code that uniquely describes 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 that contains information about the effect that the repair has 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, a call is placed with service and support, and a notification is sent to the contact that is defined in the ESA-guided setup wizard.

Remote support

The Remote Management and Control (RMC) subsystem is delivered as part of the base operating system, including the operating system that runs on the HMC. RMC provides a secure transport mechanism across the LAN interface between the operating system and the optional HMC and is used by the operating system diagnostic application for transmitting error information. It performs several other functions, but they are not used for the service infrastructure.

Service Focal Point application for partitioned systems

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 be reported only once, regardless of how many logical partitions experience the potential effect of the error. The SFP application on the management console or in the Integrated Virtualization Manager (IVM) is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management. The SFP application provides other service-related functions, such as controlling service indicators, setting up Call Home, and providing guided maintenance.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000168

When a local or globally reported service request is made to the operating system, the operating system diagnostic subsystem uses the RMC subsystem to relay error information to the optional HMC. For global events (platform unrecoverable errors, for example), the service processor also forwards error notification of these events to the HMC, providing a redundant error-reporting path in case there are errors in the RMC subsystem network.

The first occurrence of each failure type is recorded in the Manage Serviceable Events task on the management console. This task then filters and maintains a history of duplicate reports from other logical partitions or from the service processor. It then looks at all active service event requests within a predefined timespan, analyzes the failure to ascertain the root cause and, if enabled, initiates a Call Home for service. This methodology ensures that all platform errors are reported through at least one functional path, ultimately resulting in a single notification for a single problem. Similar service functionality is provided through the SFP application on the IVM for providing service functions and interfaces on non-HMC partitioned servers.

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 that is collected depends on the invocation method, but includes information such as 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 either to assist the service support organization with preparing a service action plan for the IBM SSR or for additional analysis.

System dump handling

In certain circumstances, an error might require a memory dump to be automatically or manually created. In this event, the memory dump may be offloaded to the optional HMC. Specific management console information is included as part of the information that optionally can be sent to IBM Support for analysis. If additional information that relates to the memory dump is required, or if viewing the memory dump remotely becomes necessary, the management console memory dump record notifies the IBM Support center regarding on which managements console the memory dump is located. If no management console is present, the memory dump might be either on the FSP or in the operating system, depending on the type of memory dump that was initiated and whether the operating system is operational.

4.4.6 Notification

After a Power Systems server detects, diagnoses, and reports 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 client notification might range from a simple notification to having field service personnel automatically dispatched to the client site with the correct replacement part.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000169

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 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 events include the following examples:

- 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 because they indicate that something happened that requires client awareness if the client wants to take further action. These events can be reported to IBM at the discretion of the client.

Call Home

Call Home refers to an automatic or manual call from a customer location to an IBM Support structure with error log data, server status, or other service-related information. The Call Home feature starts the service organization so that the appropriate service action can begin. Call Home can be done through HMC or most non-HMC managed systems. Although configuring a Call Home function is optional, clients are encouraged to implement this feature to obtain service enhancements, such as reduced problem determination and faster and potentially more accurate transmission of error information. In general, using the Call Home feature can result in increased system availability. The ESA application can be configured for automated Call Home. For more information, see 4.5.4, “Electronic Services and Electronic Service Agent” on page 166.

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, the manufacturing level of the parts, and so on. These records provide valuable information that can be used by remote support and IBM SSRs, enabling the IBM SSRs to assist in keeping the firmware and software current on the server.

IBM Service and Support 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 that are taken by the IBM SSR, is recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

4.4.7 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000170

Packaging for service

The following 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 may not be concurrently maintained (they might require that the system is 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 require basic tools such as flathead screw drivers, to service the hardware components.

- Positive retention

  Positive retention mechanisms help 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 risk become loose during shipping or installation, which prevents 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 function is for scale-out systems, including Power Systems such as models Power S814L and Power S824L, that can be repaired by clients. In the Light Path LED implementation, when a fault condition is detected on the POWER8 processor-based system, an amber FRU fault LED is illuminated (turned on solid), which is then rolled up to the system fault LED. The Light Path system pinpoints the exact part by lighting the amber FRU fault LED that is associated with the part that must be replaced.

The servicer can clearly identify components for replacement by using specific component level identify LEDs, and can also guide the IBM SSR directly to the component by signaling (flashing) the FRU component identify LED, and rolling up to the blue enclosure Locate LED.

After the repair, the LEDs shut off automatically when the problem is fixed. The Light Path LEDs are only visible while system is in standby power. There is no gold cap or battery implemented.

Service labels

Service providers use these labels to assist with maintenance actions. Service labels are in various formats and positions, and are intended to transmit readily available information to the IBM SSR during the repair process.

Several of these service labels and their purposes are described in the following list:

- *Location diagrams* are strategically positioned on the system hardware and relate information about 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, sockets, processor cards, fans, adapter, LEDs, and power supplies.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000171

- *Remove or replace procedure labels* contain procedures that are often found on a cover of the system or in other locations that are accessible to the IBM SSR. 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 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 an LCD display (two rows by 16 elements) 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 an IBM SSR or client to change various boot-time options and for other limited service functions.

Concurrent maintenance

The IBM POWER8 processor-based systems are designed with the understanding that certain components have higher intrinsic failure rates than others. These components can include fans, power supplies, and physical storage devices. Other devices, such as I/O adapters, can begin to wear from repeated plugging and unplugging. For these reasons, these devices are designed to be concurrently maintainable when properly configured. Concurrent maintenance is facilitated because of the redundant design for the power supplies, fans, and physical storage.

In addition to the previously mentioned components, the operator panel can be replaced concurrently by using service functions of the ASMI menu.

Repair and verify services

Repair and verify (R&V) services are automated service procedures that are used to guide a service provider, step-by-step, through the process of repairing a system and verifying that the problem was repaired. The steps are customized in the appropriate sequence for the particular repair for the specific system being serviced. The following scenarios are covered by R&V services:

- Replacing a defective FRU or a CRU
- 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

R&V procedures can be used by IBM SSR 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 R&V 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.

Clients can subscribe through the subscription services on the IBM Support Portal to obtain notifications about the latest updates that are available for service-related documentation.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000172

IBM Knowledge Center

IBM Knowledge Center provides you with a single information center where you can access product documentation for IBM systems hardware, operating systems, and server software.

The latest version of the documentation is accessible through the Internet; however, a CD-ROM based version is also available.

The purpose of Knowledge Center, in addition to providing client related product information, is to provide softcopy information to diagnose and fix any problems that might occur with the system. Because the information is electronically maintained, changes due to updates or addition of new capabilities can be used by service representatives immediately.

The Knowledge Center contains sections specific to each server model, and include detailed service procedures for a number of potential repair situations. The service procedure repository for a particular server model can be found in the “Troubleshooting, service and support” section.

The IBM Knowledge Center can be found online at:

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

QR code labels for servicing information

A label containing a QR code can be found on the top service cover of the Power S814 and Power S824 servers. This can be scanned with an appropriate app on a mobile device to link to a number of sources of information that simplify the servicing of the system.

From this quick access link you can find information on topics including:

- Installing and configuring the system
- Troubleshooting and problem analysis
- Reference code lookup tables
- Part location guides
- Removing and replacing field replaceable units
- Video guides for removal and installation of customer replaceable units
- Warranty and maintenance contracts
- Full product documentation

4.5 Manageability

Several functions and tools help manageability, so you can efficiently and effectively manage your system.

4.5.1 Service user interfaces

The service interface allows support personnel or the client to communicate with the service support applications in a server by 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 that are 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. Here are the primary service interfaces:

- Light Path (See “Light Path ” on page 155 and “Service labels ” on page 155.)
- Service processor and ASMI

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000173

- Operator panel
- Operating system service menu
- SFP on the HMC
- SFP Lite on IVM

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

The service processor is used to monitor and manage the system hardware resources and devices. The service processor checks the system for errors, ensuring that the connection to the management console for manageability purposes and accepting ASMI Secure Sockets Layer (SSL) network connections. The service processor can view and manage the machine-wide settings by using the ASMI, and enables complete system and partition management from the HMC.

Analyzing a system that does not boot: The FSP can analyze a system that does not boot. Reference codes and detailed data is available in the ASMI and are transferred to the HMC.

The service processor uses two Ethernet ports that run at 1 Gbps speed. Consider 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. The ASMI options can be accessed through an HTTP server that is integrated into the service processor operating environment.
- Both Ethernet ports support only auto-negotiation. Customer-selectable media speed and duplex settings are not available.
- Both Ethernet ports have a default IP address, as follows:
  - Service processor eth0 (HMC1 port) is configured as 169.254.2.147.
  - Service processor eth1 (HMC2 port) is configured as 169.254.3.147.

The following functions are available through the service processor:

- Call Home
- ASMI
- Error information (error code, part number, and location codes) menu
- View of guarded components
- Limited repair procedures
- Generate dump
- LED Management menu
- Remote view of ASMI menus
- Firmware update through a USB key

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000174

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 VPD. Various repair procedures require connection to the ASMI.

The ASMI is accessible through the management console. 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, but this is available only while the system is in the platform powered-off mode.

Use the ASMI to change the service processor IP addresses or to apply certain security policies and prevent access from unwanted 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 the following methods:

- Use a management console.

  If configured to do so, the management console connects directly to the ASMI for a selected system from this task.

  To connect to the ASMI from a management console, complete the following steps:

  a. Open **Systems Management** from the navigation pane.
  b. From the work window, select one of the managed systems.
  c. From the System Management tasks list, click **Operations** → **Launch Advanced System Management (ASM)**.

- Use a web browser.

  At the time of writing, supported web browsers are Microsoft Internet Explorer (Version 10.0.9200.16439), Mozilla Firefox ESR (Version 24), and Chrome (Version 30). Later versions of these browsers might work, but are not officially supported. The JavaScript language and cookies must be enabled and TLS 1.2 might need to be enabled.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000175

The web interface is available during all phases of system operation, including the initial program load (IPL) and run time. However, several 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 an SSL web connection to the service processor. To establish an SSL connection, open your browser by using the following address:

```
https://<ip_address_of_service_processor>
```

Note: To make the connection through Internet Explorer, click **Tools Internet Options**. Clear the **Use TLS 1.0** check box, and click **OK**.

- Use 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 powered-off mode. The ASMI on an ASCII console is not available during several phases of system operation, such as the IPL and run time.

- Command-line start of the ASMI

  Either on the HMC itself or when properly configured on a remote system, it is possible to start ASMI web interface from the HMC command line. Open a terminal window on the HMC or access the HMC with a terminal emulation and run the following command:

  ```
  asmmenu --ip <ip address>
  ```

  On the HMC itself, a browser window opens automatically with the ASMI window and, when configured properly, a browser window opens on a remote system when issued from there.

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 two ways:

- By using the normal operational front view
- By pulling it out to access the switches and viewing the LCD display

Here are several of the operator panel features:

- A 2 x 16 character LCD display
- Reset, enter, power On/Off, increment, and decrement buttons
- Amber System Information/Attention, and a green Power LED
- Blue Enclosure Identify LED on the Power S614 and Power S824
- Altitude sensor
- USB Port
- Speaker/Beeper

The following 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 diagnostic tests consist of IBM i service tools, stand-alone diagnostic tests that are loaded from the DVD drive, and online diagnostic tests (available in AIX).

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000176

Online diagnostic tests, 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 modes are as follows:

- Service mode

  This mode requires a service mode boot of the system and enables the checking of system devices and features. Service mode provides the most complete self-check of the system resources. All system resources, except the SCSI adapter and the disk drives that are used for paging, can be tested.

- Concurrent mode

  This mode 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 a diagnostic application before testing can be done.

- Maintenance mode

  This mode 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 started. Maintenance mode requires that all activity on the operating system is stopped. Run **`shutdown -m`** 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.

The service processor's error log can be accessed on the ASMI menus.

You can also access the system diagnostics from a Network Installation Management (NIM) server.

Alternative method: When you order a Power System, a DVD-ROM or DVD-RAM might be an option. An alternative method for maintaining and servicing the system must be available if you do not order the DVD-ROM or DVD-RAM.

IBM i and its associated machine code provide dedicated service tools (DSTs) as part of the IBM i licensed machine code (Licensed Internal Code) and System Service Tools (SSTs) as part of IBM i. DSTs can be run in dedicated mode (no operating system is loaded). DSTs and diagnostic tests are a superset of those available under SSTs.

The IBM i End Subsystem (**`ENDSBS *ALL`**) command can shut down all IBM and customer applications subsystems except for 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 keeps all applications running, by using the IBM i Start Service Tools (**`STRSST`**) command (when signed onto IBM i with the appropriately secured user ID).

With DSTs and SSTs, you can look at various logs, run various diagnostic tests, or take several kinds of system memory dumps or other options.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000177

Depending on the operating system, the following service-level functions are what you typically see when you use the operating system service menus:

- Product activity log
- Trace Licensed Internal Code
- Work with communications trace
- Display/Alter/Dump
- Licensed Internal Code log
- Main storage memory 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 management console can help streamline this process.

Each logical partition reports errors that it detects and forwards the event to the SFP application that is running on the management console, 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 management console, 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 memory dumps.

4.5.2 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 its associated I/O adapters.

Firmware entitlement

With the new HMC Version V8R8.1.0.0 and Power Systems servers, the firmware installations are restricted to entitled servers. The customer must be registered with IBM and entitled with a service contract. During the initial machine warranty period, the access key already is installed in the machine by manufacturing. The key is valid for the regular warranty period plus some additional time. The Power Systems Firmware is relocated from the public repository to the access control repository. The I/O firmware remains on the public repository, but the server must be entitled for installation. When the **`lslic`** command is run to display the firmware levels, a new value, **`update_access_key_exp_date`**, is added. The HMC GUI and the ASMI menu show the Update access key expiration date.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000178

When the system is no longer entitled, the firmware updates fail. Some new System Reference Code (SRC) packages are available:

- E302FA06: Acquisition entitlement check failed
- E302FA08: Installation entitlement check failed

Any firmware release that was made available during the entitled time frame can still be installed. For example, if the entitlement period ends on 31 December 2014, and a new firmware release is release before the end of that entitlement period, then it can still be installed. If that firmware is downloaded after 31 December 2014, but it was made available before the end of the entitlement period, it still can be installed. Any newer release requires a new update access key.

Note: The update access key expiration date requires a valid entitlement of the system to perform firmware updates.

You can find an update access key at the IBM CoD Home website:

http://www.ibm.com/pod/pod

To access the IBM entitled Software Support page for further details, go to the following website:

http://www.ibm.com/servers/eserver/ess

Firmware updates

System firmware is delivered as a release level or a service pack. Release levels support the general availability (GA) of new functions 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 introduce new functions. A *service pack* is an update to an existing release level.

If the system is managed by a management console, you use the management console for firmware updates. By using the management console, you can take advantage of the CFM option when concurrent service packs are available. CFM is the IBM Power Systems Firmware updates that can be partially or wholly concurrent or nondisruptive. With the introduction of CFM, IBM is increasing its clients' opportunity to stay on a given release level for longer periods. Clients that want maximum stability can defer until there is a compelling reason to upgrade, such as the following reasons:

- A release level is approaching its end-of-service date (that is, it has been available for about a year, and soon service will not be supported).
- Move a system to a more standardized release level when there are multiple systems in an environment with similar hardware.
- A new release has a new function that is needed in the environment.
- A scheduled maintenance action causes a platform reboot, which provides an opportunity to also upgrade to a new firmware release.

The updating and upgrading of system firmware depends on several factors, such as whether the system is stand-alone or managed by a management console, the current firmware that is 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, found at the following website:

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000179

You might also want to review the preferred practice white papers that are found at the following website:

http://www14.software.ibm.com/webapp/set2/sas/f/best/home.html

Firmware update steps

The system firmware consists of service processor microcode, Open Firmware microcode, and Systems Power Control Network (SPCN) microcode.

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, if that system is not managed by an HMC.

Power Systems has a permanent firmware boot side (A side) and a temporary firmware boot side (B side). New levels of firmware must be installed first on the temporary side to test the update's compatibility with existing applications. When the new level of firmware is approved, it can be copied to the permanent side.

For access to the initial websites that address this capability, see the Support for IBM Systems website:

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

For Power Systems, select the **Power** link.

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 used to view the levels of server firmware and power subsystem firmware that are installed and that 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 is 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.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000180

- 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.

Figure 4-3 shows the different levels in the HMC.

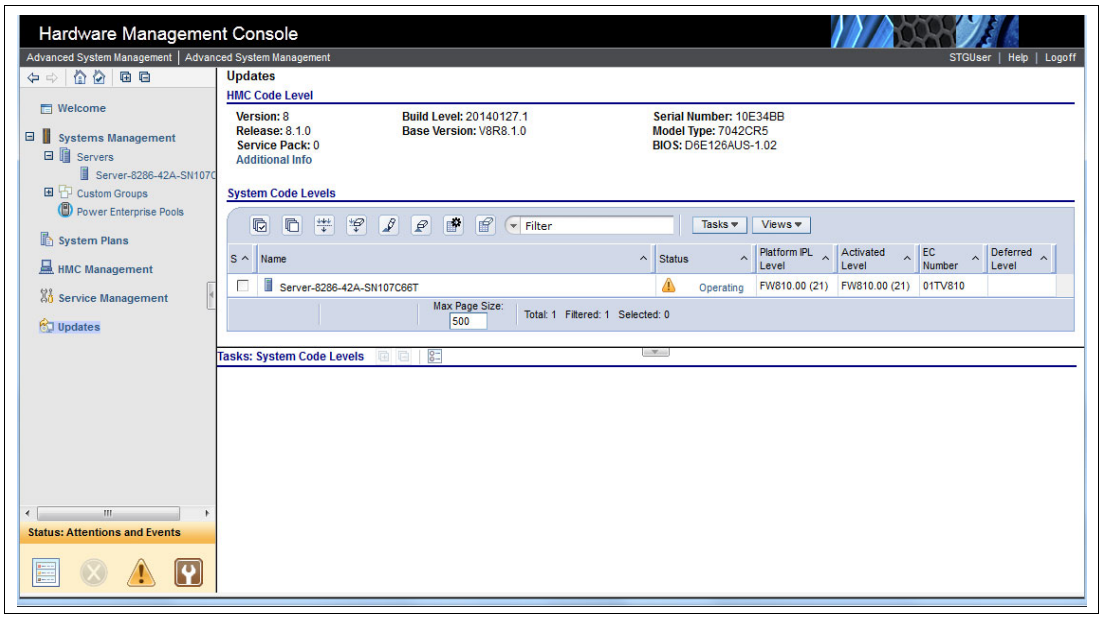

Figure 4-3 HMC System Firmware window

IBM provides the 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, are 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	-
NN	Platform and class	SV	Low end
SSS	Release indicator		
FFF	Current fix pack		
DDD	Last disruptive fix pack		

The following example uses the convention:

01SV810_030_030 = POWER8 Entry Systems Firmware for 8286-41A and 8286-42A

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000181

An installation is disruptive if the following statements are true:

- The release levels (SSS) of the currently installed and the new firmware differ.
- The service pack level (FFF) and the last disruptive service pack level (DDD) are equal in the new firmware.

Otherwise, an installation is concurrent if the service pack level (FFF) of the new firmware is higher than the service pack level that is installed on the system and the conditions for disruptive installation are not met.

4.5.3 Concurrent firmware maintenance improvements

Since POWER6, firmware service packs are concurrently applied and take effect immediately. Occasionally, a service pack is shipped where most of the features can be concurrently applied, but because changes to some server functions (for example, changing initialization values for chip controls) cannot occur during operation, a patch in this area required a system reboot for activation.

With the Power-On Reset Engine (PORE), the firmware can now dynamically power off processor components, change the registers, and reinitialize while the system is running, without discernible impact to any applications running on a processor. This potentially allows concurrent firmware changes in POWER8, which in earlier designs required a reboot to take effect.

Activating new firmware functions requires installation of a firmware release level. This process is disruptive to server operations and requires a scheduled outage and full server reboot.

4.5.4 Electronic Services and Electronic Service Agent

IBM 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 that is 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
- Electronic Service Agent

Electronic Services news page

The Electronic Services news page is a single Internet entry point that replaces the multiple entry points that traditionally are used to access IBM Internet services and support. With the news page, you can gain easier access to IBM resources for assistance in resolving technical problems.

Electronic Service Agent

The ESA is software that is on your server. It monitors events and transmits system inventory information to IBM on a periodic, client-defined timetable. The ESA 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 SSRs when they help answer your

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000182

questions or diagnose problems. Installation and use of ESA for problem reporting enables IBM to provide better support and service for your IBM server.

To learn how Electronic Services can work for you, see the following website (an IBM ID is required):

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

Here are some of the benefits of Electronic Services:

- Increased uptime

  The ESA tool enhances the warranty or maintenance agreement by providing faster hardware error reporting and uploading system information to IBM Support. This can translate to less time that is wasted monitoring the symptoms, diagnosing the error, and manually calling IBM Support to open a problem record.

  Its 24x7 monitoring and reporting mean no more dependence on human intervention or off-hours customer personnel when errors are encountered in the middle of the night.

- Security

  The ESA tool is designed to be secure in monitoring, reporting, and storing the data at IBM. The ESA tool securely transmits either through the Internet (HTTPS or VPN) or modem, and can be configured to communicate securely through gateways to provide customers a single point of exit from their site.

  Communication is one way. Activating ESA does not enable IBM to call into a customer's system. System inventory information is stored in a secure database, which is protected behind IBM firewalls. It is viewable only by the customer and IBM. The customer's business applications or business data is never transmitted to IBM.

- More accurate reporting

  Because system information and error logs are automatically uploaded to the IBM Support center with the service request, customers are not required to find and send system information, decreasing the risk of misreported or misdiagnosed errors.

  When inside IBM, problem error data is run through a data knowledge management system and knowledge articles are appended to the problem record.

- Customized support

  By using the IBM ID that you enter during activation, you can view system and support information by selecting **My Systems** at the Electronic Support website:

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

  My Systems provides valuable reports of installed hardware and software, using information that is collected from the systems by ESA. Reports are available for any system that is associated with the customers IBM ID. Premium Search combines the function of search and the value of ESA information, providing advanced search of the technical support knowledge base. Using Premium Search and the ESA information that was collected from your system, your clients can see search results that apply specifically to their systems.

  For more information about how to use the power of IBM Electronic Services, contact your IBM SSR, or see the following website:

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

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000183

Service Event Manager

The Service Event Manager allows the user to decide which of the Serviceable Events are called home with the ESA. It is possible to lock certain events. Some customers might not allow data to be transferred outside their company. After the SEM is enabled, the analysis of the possible problems might take longer.

- The SEM can be enabled by running the following command:

  ```
  chhmc -c sem -s enable
  ```

- You can disable SEM mode and specify what state in which to leave the Call Home feature by running the following commands:

  ```
  chhmc -c sem -s disable --callhome disable
  chhmc -c sem -s disable --callhome enable
  ```

You can do the basic configuration of the SEM from the HMC GUI. After you select the Service Event Manager, as shown in Figure 4-4, you must add the HMC console.

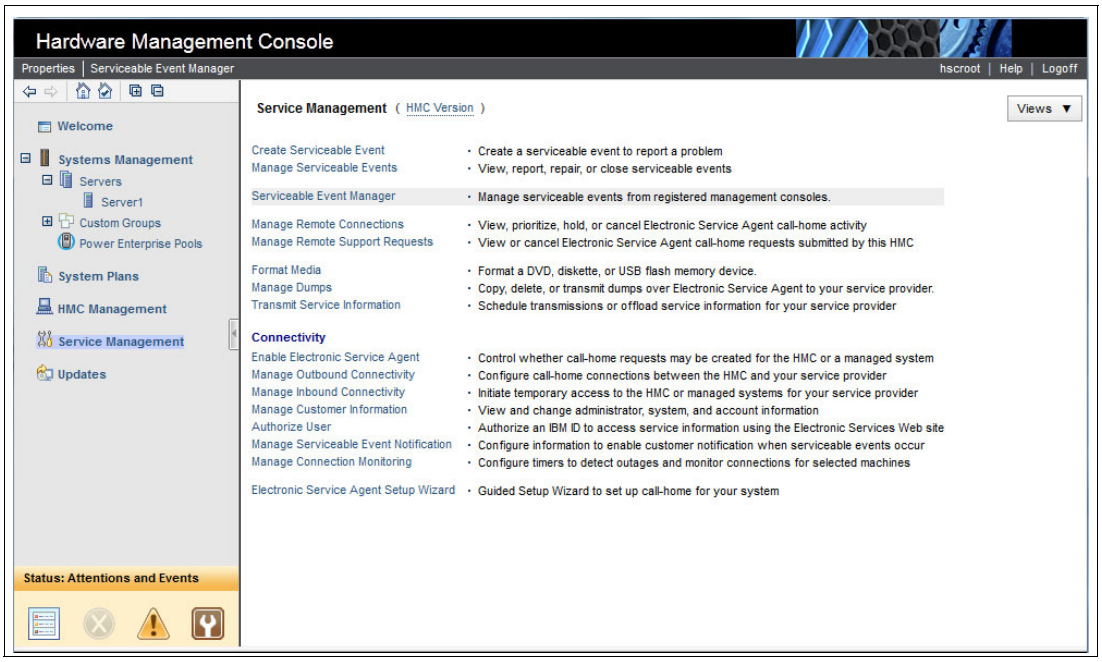

Figure 4-4 HMC selection for Service Event Manager

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000184

In the next window, you can configure the HMC that is used to manage the Serviceable Events and proceed with further configuration steps, as shown in Figure 4-5.

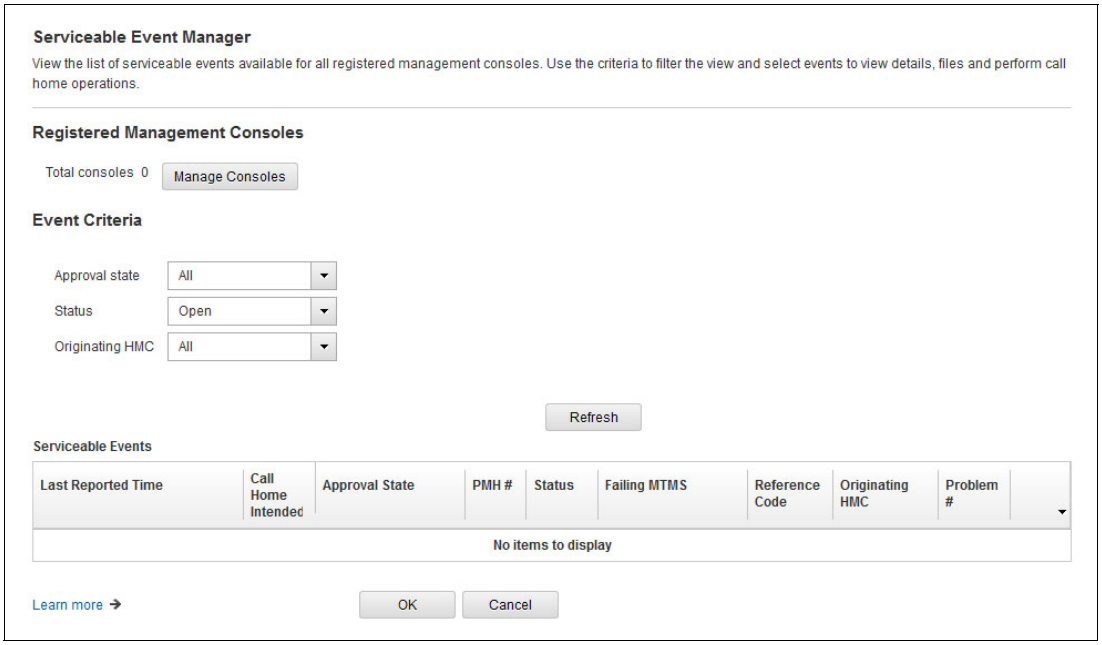

Figure 4-5 Initial SEM window

Here are detailed descriptions of the different configurable options:

- Registered Management Consoles

  “Total consoles” lists the number of consoles that are registered. Select **Manage Consoles** to manage the list of RMCs.

- Event Criteria

  Select the filters for filtering the list of serviceable events that are shown. After the selections are made, click **Refresh** to refresh the list based on the filter values.

- Approval state

  Select the value for approval state to filter the list.

- Status

  Select the value for the status to filter the list.

- Originating HMC

  Select a single registered console or **All consoles** to filter the list.

- Serviceable Events

  The Serviceable Events table shows the list of events based on the filters that are selected. To refresh the list, click **Refresh**.

The following menu options are available when you select an event in the table:

- View Details...

  Shows the details of this event.

- View Files...

  Shows the files that are associated with this event.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000185

- Approve Call Home

  Approves the Call Home of this event. This option is available only if the event is not approved already.

The Help / Learn more function can be used to get more information about the other available windows for the Serviceable Event Manager.

4.6 Selected POWER8 RAS capabilities by operating system

Table 4-2 provides a list of the Power Systems RAS capabilities by operating system. The HMC is an optional feature on scale-out Power Systems servers.

Table 4-2 Selected RAS features by operating system

RAS feature	AIX V7.1 TL3 SP3 V6.1 TL9 SP3	IBM i V7R1M0 TR8 V7R2M0	Linux RHEL6.5 RHEL7 SLES11SP3 Ubuntu 14.04
Processor			
FFDC for fault detection/error isolation	X	X	X
Dynamic Processor Deallocation	X	X	X[a]
Core Error Recovery			
► Alternative processor recovery	X	X	X[a]
► Partition Core Contained Checkstop	X	X	X[a]
I/O subsystem			
PCI Express bus enhanced error detection	X	X	X
PCI Express bus enhanced error recovery	X	X	X[b]
PCI Express card hot-swap	X	X	X[a]
Memory availability			
Memory Page Deallocation	X	X	X
Special Uncorrectable Error Handling	X	X	X
Fault detection and isolation			
Storage Protection Keys	X	Not used by OS	Not used by OS
Error log analysis	X	X	X[b]
Serviceability			
Boot-time progress indicators	X	X	X
Firmware error codes	X	X	X
Operating system error codes	X	X	X[b]
Inventory collection	X	X	X

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000186

RAS feature	AIX V7.1 TL3 SP3 V6.1 TL9 SP3	IBM i V7R1M0 TR8 V7R2M0	Linux RHEL6.5 RHEL7 SLES11SP3 Ubuntu 14.04
Environmental and power warnings	X	X	X
Hot-swap DASD / media	X	X	X
Dual disk controllers / Split backplane	X	X	X
EED collection	X	X	X
SP “Call Home” on non-HMC configurations	X	X	X[a]
IO adapter/device stand-alone diagnostic tests with PowerVM	X	X	X
SP mutual surveillance with POWER Hypervisor	X	X	X
Dynamic firmware update with HMC	X	X	X
Service Agent Call Home Application	X	X	X[a]
Service Indicator LED support	X	X	X
System dump for memory, POWER Hypervisor, and SP	X	X	X
Information center / iBM Systems Support Site service publications	X	X	X
System Support Site education	X	X	X
Operating system error reporting to HMC SFP application	X	X	X
RMC secure error transmission subsystem	X	X	X
Healthcheck scheduled operations with HMC	X	X	X
Operator panel (real or virtual)	X	X	X
Concurrent Op Panel Maintenance	X	X	X
Redundant HMCs	X	X	X
Automated server recovery/restart	X	X	X
High availability clustering support	X	X	X
Repair and Verify Guided Maintenance with HMC	X	X	X
PowerVM Live Partition / Live Application Mobility With PowerVM Enterprise Edition	X	X[c]	X
EPOW			
EPOW errors handling	X	X	X[a]

a. Supported in POWER Hypervisor, but not supported in a PowerKVM environment
b. Supported in POWER Hypervisor, with limited support in a PowerKVM environment
c. For POWER8 systems, IBM i requires IBM i 7.1 TR9 and IBM i 7.2 TR1.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000187

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Some publications referenced in this list might be available in softcopy only.

- *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491
- *IBM Power Systems S812L and S822L Technical Overview and Introduction*, REDP-5098
- *IBM Power System S822 Technical Overview and Introduction*, REDP-5102
- *IBM Power System S824L Technical Overview and Introduction*, REDP-5139
- *IBM Power Systems E850 Technical Overview and Introduction*, REDP-5222
- *IBM Power Systems E870 and E880 Technical Overview and Introduction*, REDP-5137
- *IBM Power Systems SR-IOV Technical Overview and Introduction*, REDP-5065
- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Enhancements What is New in VIOS 2.2.3*, SG24-8198
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *Performance Optimization and Tuning Techniques for IBM Processors, including IBM POWER8*, SG24-8171

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:

- *Active Memory Expansion: Overview and Usage Guide*

  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03037usen/POW03037USEN.PDF
- *IBM EnergyScale for POWER8 Processor-Based Systems* white paper

  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03039usen/POW03039USEN.PDF
- IBM Power Facts and Features - IBM Power Systems, IBM PureFlex System, and Power Blades

  http://www.ibm.com/systems/power/hardware/reports/factsfeatures.html
- IBM Power System S812L server specifications

  http://www.ibm.com/systems/power/hardware/s812l-s822l/specs.html

© Copyright IBM Corp. 2014. All rights reserved.

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000189

- IBM Power System S814 server specifications
  http://www.ibm.com/systems/power/hardware/s814/specs.html
- IBM Power System S822 server specifications
  http://www.ibm.com/systems/power/hardware/s822/specs.html
- IBM Power System S822L server specifications
  http://www.ibm.com/systems/power/hardware/s812l-s822l/specs.html
- IBM Power System S824 server specifications
  http://www.ibm.com/systems/power/hardware/s824/specs.html
- IBM Power System S824L server specifications:
  http://www.ibm.com/systems/power/hardware/s824l/specs.html
- IBM Power System E850 server specifications:
  http://www.ibm.com/systems/power/hardware/e850/specs.html
- IBM Power System E870 server specifications:
  http://www.ibm.com/systems/power/hardware/e870/specs.html
- IBM Power System E870 server specifications:
  http://www.ibm.com/systems/power/hardware/e870/specs.html
- Specific storage devices that are supported for Virtual I/O Server
  http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/datasheet.html
- *System RAS - Introduction to Power Systems Reliability, Availability, and Serviceability*
  http://public.dhe.ibm.com/common/ssi/ecm/en/pow03056usen/POW03056USEN.PDF

Online resources

These websites are also relevant as further information sources:

- IBM Fix Central website
  http://www.ibm.com/support/fixcentral/
- IBM Knowledge Center
  http://www.ibm.com/support/knowledgecenter/
- IBM Power Systems website
  http://www.ibm.com/systems/power/
- IBM Power Systems Hardware information center
  http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5/index.jsp
- IBM Storage website
  http://www.ibm.com/systems/storage/
- IBM System Planning Tool website
  http://www.ibm.com/systems/support/tools/systemplanningtool/
- IBM Systems Energy Estimator
  http://www-912.ibm.com/see/EnergyEstimator/

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000190

- Migration combinations of processor compatibility modes for active Partition Mobility

  http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hc3/iphc3pcmcombosact.htm

- 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

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000191

IBM Power Systems S814 and S824 Technical Overview and Introduction

Redpaper™

Outstanding performance based on POWER8 processor technology

4U scale-out desktop and rack-mount servers

Improved reliability, availability, and serviceability features

This IBM Redpaper publication is a comprehensive guide covering the IBM Power System S814 (8286-41A) and IBM Power System S824 (8286-42A) servers that support IBM AIX, IBM i, and Linux operating systems. The objective of this paper is to introduce the major innovative Power S814 and Power S824 offerings and their relevant functions:

- The new IBM POWER8 processor, available at frequencies of 3.02 GHz, 3.52 GHz, 3.72 GHz, 3.89 GHz, and 4.15 GHz
- Significantly strengthened cores and larger caches
- Two integrated memory controllers with improved latency and bandwidth
- Integrated I/O subsystem and hot-pluggable PCIe Gen3 I/O slots
- I/O drawer expansion options offers greater flexibility
- Improved reliability, serviceability, and availability functions
- IBM EnergyScale technology that provides features such as power trending, power-saving, capping of power, and thermal measurement

This publication is for professionals who want to acquire a better understanding of IBM Power Systems products.

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 S814 and Power S824 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.

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

REDP-5097-00

Advanced Micro Devices, Inc., Pensando Systems, Inc. - Ex. 1017, Page 000194