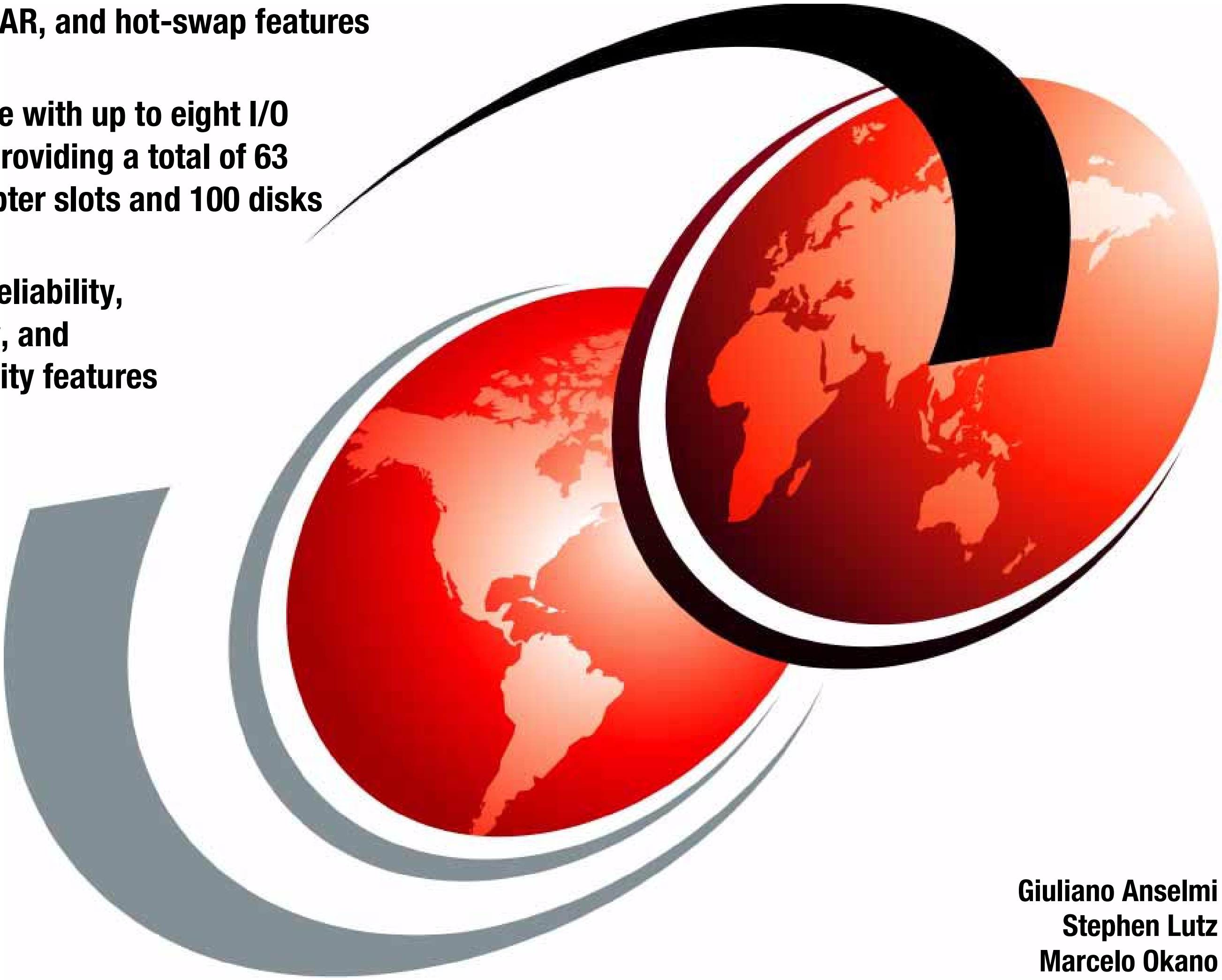
IBM eserver

pSeries 650 Model 6M2 Technical Overview and Introduction

Innovative, POWER4+ processor, RIO-2 I/O, LPAR, DLPAR, and hot-swap features

Expandable with up to eight I/O drawers, providing a total of 63 PCI-X adapter slots and 100 disks

High-end reliability, availability, and serviceability features



Giuliano Anselmi
Stephen Lutz
Marcelo Okano

Redpaper

ibm.com/redbooks

International Technical Support Organization

IBM @server pSeries 650 Model 6M2 Technical Overview and Introduction

May 2003

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

Second Edition (May 2003)

This edition applies to the IBM *@server*™ pSeries™ 650 Model 6M2, AIX® 5L™ Version 5.1, product number 5765-E61 and AIX 5L Version 5.2, product number 5765-E62.

Contents

Notices

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

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

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

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

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

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

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

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

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

Trademarks

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

AIX®
AIX 5L™
AS/400®
Balance®
Chipkill™
ClusterProven™
Electronic Service Agent™
Enterprise Storage Server™
@server ™
@server™
IBM®
IBMLink™
iSeries™
Lotus®
Notes®
PowerPC®
POWER4™
POWER4+™
pSeries™
Redbooks™
Redbooks(logo)™
RS/6000®
Service Director™
SP™
TotalStorage™
Word Pro®

The following terms are trademarks of other companies:

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

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

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

C-bus is a trademark of Corollary, Inc. in the United States, other countries, or both.

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

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

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

Preface

This document provides a comprehensive guide covering the IBM @server pSeries 650 Model 6M2 server. Major hardware offerings are introduced and their prominent functions discussed.

Professionals wishing to acquire a better understanding of IBM @server pSeries products may consider reading this document. The intended audience includes:

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

This document expands the current set of IBM @server pSeries documentation by providing a desktop reference that offers a detailed technical description of the pSeries 650 Model 6M2.

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

The team that wrote this Redpaper

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

Giuliano Anselmi is an IBM pSeries systems product engineer at the EMEA Midrange Field Support Organization in Rome, Italy, supporting the Web Server Sales Organization in EMEA, IBM sales, IBM Business Partners, and technical support organizations. He is a resolution assistant resource for critical situations escalated and related to RS/6000® and pSeries systems for the pSeries Customer Satisfaction Project Office, and IBM Dublin Enterprise Servers manufacturing plant. He has been devoted to RS/6000 and pSeries systems for 10 years.

Stephen Lutz is an IT Specialist in pre-sales technical support for IBM @server pSeries and RS/6000, part of the Web Server Sales Organization in Stuttgart, Germany. He holds a degree in Computer Science from the Fachhochschule Karlsruhe - University of Technology and is an IBM Certified Advanced Technical Expert. Stephen is a member of the High-End Technology Focus Group, supporting IBM sales, IBM Business Partners, and customers with pre-sales consultation and implementation of client/server environments.

Marcelo Okano is a Senior IT Specialist for pre-sales technical support for IBM @server pSeries and RS/6000 in Brazil. He has 10 years of experience working on pSeries products. He holds a degree in Mathematics from the Faculdade de Filosofia, Ciencias e Letras de Santo Andre. He has worked at IBM for four years. His areas of expertise include pSeries and RS/6000 systems, RS/6000 SP and Cluster 1600 systems, HACMP, HAGEO, Linux, performance tuning, and AIX.

The project that created this document was managed by:

Scott Vetter of the International Technical Support Organization, Austin Center

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

IBM USA
Ron Arroyo, Russell Bistline, Robert Bluethman, Barbara M Butler, Mark Dewalt, Jim Gallagher, John Grzinich, Ron Gonzalez, Richard B Hartin, John Hilburn, Dennis Lee, Christopher M Marroquin, Warren Maule, Jim McGaughan, Andy McLaughlin, Bill Mihaltse, Duke Paulsen, Norm Roller, Craig Shempert, Joel Sullivan, Ed Toutant, Brian Vicknair, Sterling Weaver, Dave Willoughby

IBM Germany
Vladimir Atanaskovik, Volker Haug, Ralf Vollmer

IBM Italy
Massimo Olivari

Comments welcome

Your comments are important to us.

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

- Send your comments in an Internet note to:

  `sbv@us.ibm.com`

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  ATTN: Scott Vetter
  Dept. JN9B Building 003 Internal Zip 2834
  11400 Burnet Road
  Austin, Texas 78758-3493, USA

1

General description

The IBM @server pSeries 650 Model 6M2 (referred to hereafter as the p650) is a mid-range member of the 64-bit family of symmetric multiprocessing (SMP) servers from IBM. Positioned between the pSeries 630 and the powerful pSeries 670 (referred to hereafter as p630 and p670), the p650 provides the power, capacity, and expandability required for e-business mission-critical computing. The p650 is ready for the demands of 64-bit computing, while still supporting existing 32-bit applications.

The p650 delivers a cost-efficient growth path to the future. It provides 64-bit scalability through 64-bit POWER4+ processors packaged as a 2-way card. With its four processor card slots, the p650 can be configured into 2-, 4-, 6-, or 8-way SMP configurations, at speeds of either 1.2 GHz or 1.45 GHz, and includes 8 MB (1.2 GHz) or 32 MB (1.45 GHz) of shared L3 cache per processor card. Each processor card includes eight slots for DIMMs, for a maximum system memory capacity of 64 GB (16 GB per card).

The p650 incorporates in its 8U rack drawer an I/O subsystem with one integrated 10/100 Mbps Ethernet port, one internal and one external Ultra3 SCSI port, and seven blind-swap hot-pluggable PCI slots. Six of these slots support 32-bit and 64-bit standard PCI/PCI-X adapters, and one supports 32-bit adapters only. The system also includes up to four hot-swappable internal disk drives with up to 587.2 GB of total storage (using four 146.8 GB Ultra3 SCSI drives), plus two hot-swappable media bays, four serial ports, a keyboard and mouse port, service processor, and diskette drive.

Customers can expand their capacity with the option of adding up to eight 7311-D10 or 7311-D20 I/O drawers, providing an overall total of 48 PCI/PCI-X slots on the 7311-D10 or 63 PCI/PCI-X slots and 96 disk drives on the 7311-D20. External system storage can include SCSI, SSA, or Fibre Channel storage subsystems.

Reliability and availability features include redundant hot-plug cooling fans and power supplies. Along with these hot-plug components, the p650 is designed to provide an extensive set of reliability, availability, and serviceability (RAS) features that include improved fault isolation, recovery from errors without stopping the system, avoidance of recurring failures, and predictive failure analysis. See 3.3, “Reliability, availability, and serviceability (RAS) features” on page 48.

With logical partitioning (LPAR) (configurations supporting up to eight logical partitions are available), the processors, memory, and I/O within each partition can be dynamically removed or added in an AIX 5L Version 5.2 partition without the need for a reboot. LPAR requires the use of a Hardware Management Console (HMC) that operates outside the scope of any single operating system image and is used to manage and monitor the platform resources as well as provide a service focal point. Dynamic LPAR requires AIX 5L Version 5.2 or higher running in each affected partition.

1.1 Physical package

The p650 is an 8U rack-mounted server and is intended to be installed in a 19-inch rack, such as IBM 7014 T00 or T42. The physical characteristics of p650 are the following:

- Width: 445 mm (17.5 inches)
- Depth: 760 mm (29.9 inches)
- Height: 351 mm (13.8 inches)
- Weight: 93 kg (205 lbs)

Figure 1-1 shows an inside view of the p650. On the left side are positions for the processor books. Behind the processor books are slots for the GX[1] cards to connect additional I/O drawers. On the right side of the machine, the PCI-X slots are located for optional adapters.

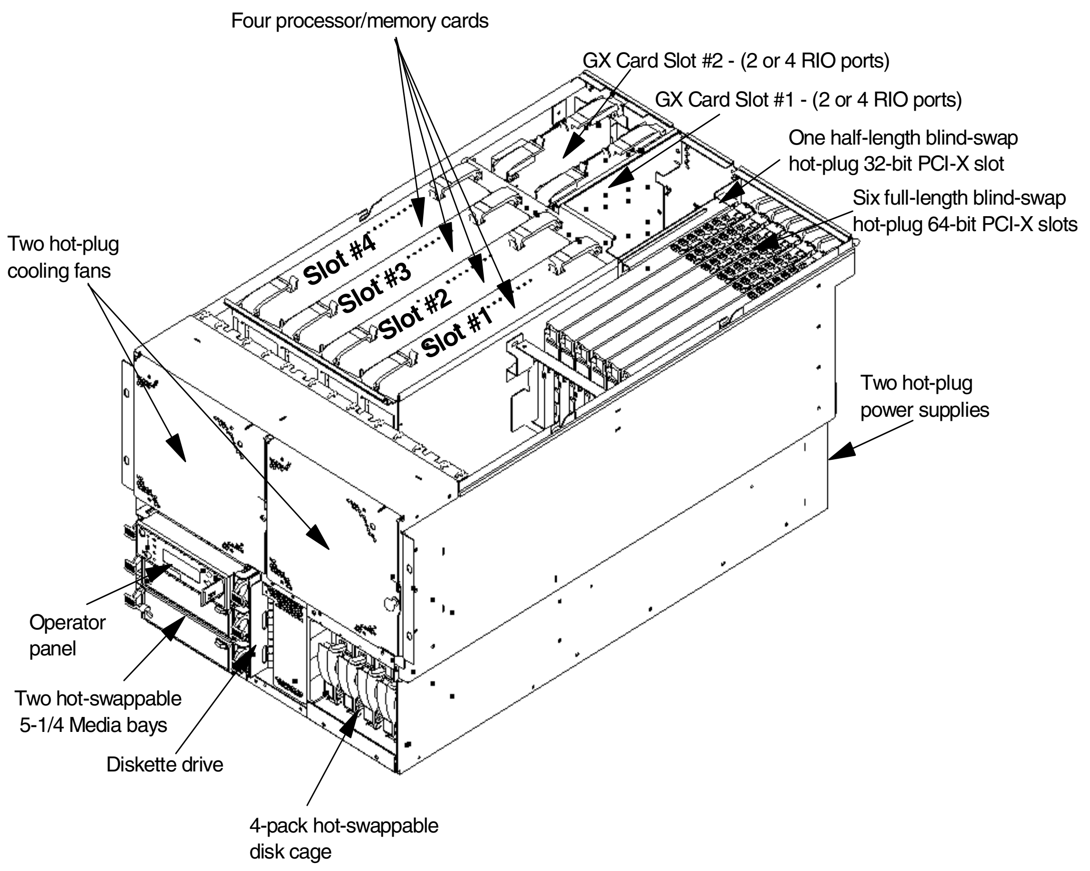

Four processor/memory cards
GX Card Slot #2 - (2 or 4 RIO ports)
GX Card Slot #1 - (2 or 4 RIO ports)
One half-length blind-swap hot-plug 32-bit PCI-X slot
Six full-length blind-swap hot-plug 64-bit PCI-X slots
Two hot-plug cooling fans
Slot #4
Slot #3
Slot #2
Slot #1
Two hot-plug power supplies
Operator panel
Two hot-swappable 5-1/4 Media bays
Diskette drive
4-pack hot-swappable disk cage

Figure 1-1 A view into the inside of the p650

[1] The GX bus is a derivative of the 6XX bus on the 620/630 processor line, where 6 has been replaced with a G for Gigaprocessor, and one of the x's has been dropped.

Figure 1-2 on page 3 shows the front view of the p650. There are two redundant hot-plug fans on top of the machine allowing concurrent operation of one fan while maintenance is performed on the other. In the lower left is the operator panel and the two hot-swappable media bays. The diskette drive is located alongside the media bays. The 4-pack disk cage with the disk locations labeled are in the lower right.

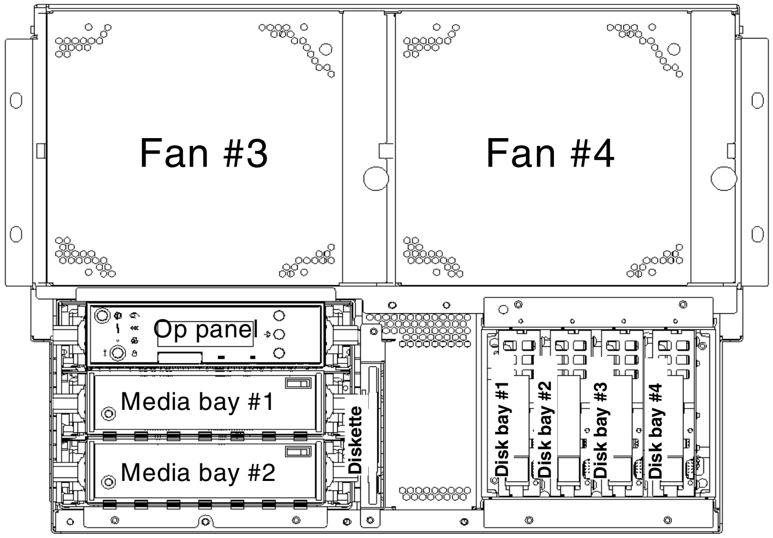


Figure 1-2 Front view of p650

Figure 1-3 shows the rear view of the p650. In the upper left are seven hot-plug PCI-X slots. In the upper right are two GX slots for connecting additional I/O drawers. In the bottom of the system are two hot-plug redundant concurrently maintainable power supplies and fans. In the middle of the rear of the machine are all of the connectors: 10/100 Mbps Ethernet, SCSI, serial ports, the rack indicator port, the HMC ports, keyboard, mouse, and SPCN ports. The reserved ports in the upper left corner as well as the debug port are for factory use only; therefore, nothing should be plugged into these ports.

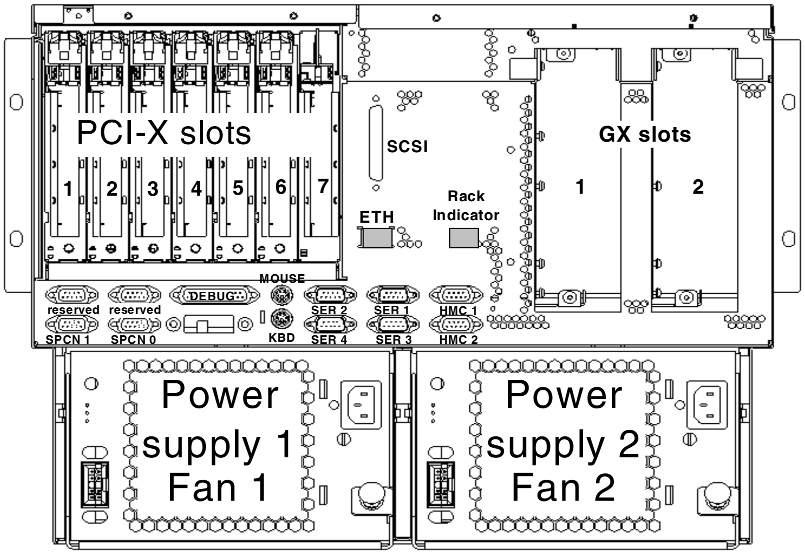


Figure 1-3 Rear view of p650

The p650 is designed to be installed and maintained by trained service representatives. However, you can hot-plug PCI/PCI-X adapters with the system up and running, provided the appropriate hot-plug tasks are followed in AIX. All adapters and devices that are part of the original order are installed and configured before shipment.

The p650 can be mounted in existing 7014 Model T00 and 7014 Model T42 racks. See 1.3, “System racks” on page 6, for further details.

1.2 Minimum and optional features

The p650 includes redundant hot-plug power supplies and cooling fans. The system supports 32-bit and 64-bit applications, and requires AIX 5L Version 5.1 at Maintenance Level 3 or AIX 5L Version 5.2 (for detailed information about AIX, see 2.9.1, “AIX 5L” on page 41).

Processors

The p650 delivers a cost-efficient growth path for the future through such capabilities as 2-, 4-, 6-, and 8-way SMP configurations.

- From a 2-way (1.2 GHz FC 5122 or 1.45 GHz FC 5208) to an 8-way
- 1.5 MB L2 cache (shared by two processors in the corresponding processor card)
- L3 cache shared per processor card
  - 32 MB L3 cache on 1.45 GHz processor card FC 5208
  - 8 MB L3 cache on 1.2 GHz processor card FC 5122

Memory

The p650 uses 208-pin DIMMs that plug into a processor card (eight DIMM slots per card). DIMMs must be populated in quads (four DIMMs) and should be balanced between processor cards. A memory feature consists of a quad. Additional quads may consist of any memory feature code (memory size).

From 2 GB up to 64 GB using the following available order combinations:

- FC 4452 2048 MB (4 x 512 MB) 208-pin 8 ns DDR SDRAM DIMMs
- FC 4453 4096 MB (4 x 1024 MB) 208-pin 8 ns stacked DDR SDRAM DIMMs
- FC 4454 8192 MB (4 x 2048 MB) 208-pin 8 ns stacked DDR SDRAM DIMMs

A system with a single processor card (2-way) may have a maximum of 16 GB of memory based on the maximum memory feature available (FC 4454, 4 x 2048 MB). The maximum memory on a 4-way is 32 GB, on a 6-way is 48 GB, and on an 8-way is 64 GB.

Disk and media

Four hot-swappable disk drive bays with 36.4 GB to 587.2 GB or optional two independent pairs of disk drives with 36.4 GB to 293.6 GB of internal disk storage are available using a split backplane. Each bay can contain one of the following disks:

- 36.4 GB Ultra3 10K RPM (FC 3273)
- 36.4 GB Ultra3 15K RPM (FC 3277)
- 73.4 GB Ultra3 10K RPM (FC 3274)
- 73.4 GB Ultra3 15K RPM (FC 3278)
- 146.8 GB Ultra3 10K RPM (FC 3275)

Two auto-docking media bays are available for hot-swappable media devices:

- 16/48X DVD-ROM auto-docking module (FC 2635)
- 40X CD-ROM auto-docking module (FC 2628)
- 4.7 GB R/W DVD-RAM auto-docking module (FC 2629)
- Tape drives
  - 4 mm 20/40 GB auto-docking module (FC 6185)
  - 8 mm 60/150 GB auto-docking module (FC 6131)
  - 8 mm 80/160 GB auto-docking module (FC 6169)

One diskette drive is also available.

At least one CD-ROM or DVD-RAM must be configured on an initial order.

Integrated I/O and expansion

The p650 basic I/O subsystem includes:

- Integrated LAN and SCSI I/O
  - One 10/100 Mbps Ethernet adapter
  - Ultra3 SCSI LVD adapter (one internal and one external 68-pin ports)
- Integrated serial, keyboard, and mouse I/O
  - Four 9-pin D-Shell serial ports
  - Keyboard and mouse ports
- Seven PCI-X slots
  - Six 64-bit 133 MHz 3.3 volt PCI-X slots, full length, blind-swap hot-plug
  - One 32-bit 66 MHz 3.3 volt PCI-X slot, half length, blind-swap hot-plug
- Systems management
  - Service processor
  - Two HMC ports
  - One rack indicator port

I/O expansion drawer

As an option, you can add up to eight 7311-D10 or 7311-D20 I/O drawers. Drawers can be intermixed on a single server provided each RIO loop contains a single drawer type. The drawers have the following attributes:

- 7311-D10 I/O drawer
  - Two RIO or RIO-2 (2x speed RIO) ports, redundant hot-plug power, redundant hot-plug cooling, and two SPCN ports.
  - Six full-size adapter slots
    - Five PCI-X slots 3.3 volt keyed 133 MHz blind-swap hot-plug
    - One PCI slot 5 volt keyed 33 MHz blind-swap hot-plug
- 7311-D20 I/O drawer
  - Two RIO or RIO-2 (2x speed RIO) ports, redundant hot-plug power, redundant hot-plug cooling, and two SPCN ports.

  - Seven full-size adapter slots
    - Seven PCI-X slots 3.3 volt keyed 133 MHz blind-swap hot-plug
  - Up to twelve hot-swappable disks.

1.3 System racks

The following descriptions provide overviews of racks available from IBM in which the p650 can be mounted.

The Enterprise Rack Models T00 and T42 are 19-inch wide racks for general use with pSeries and RS/6000 rack-based or rack drawer-based systems. The rack provides increased capacity, greater flexibility, and improved floor space utilization.

If a pSeries or RS/6000 system is to be installed in a non-IBM rack or cabinet, you should ensure that the rack conforms to the EIA standard EIA-310-D (see 1.3.4, “OEM racks” on page 8).

1.3.1 IBM RS/6000 7014 Model T00 Enterprise Rack

The 1.8 meter (71 inches) Model T00 is compatible with past and present pSeries and RS/6000 racks, and is designed for use in all situations that have previously used the older rack Models R00 and S00. The T00 rack has the following features:

- 36 EIA units (36U) of usable space.
- Optional removable side panels.
- Optional classic or sculptured front door.
- Optional side-to-side mounting hardware for joining multiple racks.
- Standard black or optional white color in OEM format.
- Increased power distribution and weight capacity.
- Optional reinforced (ruggedized) rack feature (FC 6080) provides added earthquake protection with modular rear brace, concrete floor bolt-down hardware, and bolt-in steel front filler panels.
- Optional Rack Security Kit (FC 6580) secures the side panels and allows you to add locks to the front and rear doors.
- Support for both AC and DC configurations.
- DC rack height is increased to 1926 mm (75.8 inches) due to the presence of the power distribution panel fixed to the top of the rack.
- Up to four Power Distribution Units (PDUs). See 1.3.3, “AC Power Distribution Units for rack models T00 and T42” on page 7.
- Optional rack status beacon (FC 4690). This beacon is designed to be placed on top of a rack and cabled to servers, such as a p650, and other components, such as a 7311-D10 I/O drawer, inside the rack. Servers can be programmed to illuminate the beacon in response to a detected problem or changes in system status.

  A rack status beacon junction box (FC 4693) should be used to connect multiple servers and I/O drawers to the beacon. This feature provides six input connectors and one output connector for the rack. To connect the servers or other components to the junction box or the junction box to the rack, status beacon cables (FC 4691) are necessary. Multiple junction boxes can be linked together in a series using daisy chain cables (FC 4692).

- Weight:
  - T00 base empty rack: 244 kg (535 pounds)
  - T00 full rack: 816 kg (1795 pounds)

1.3.2 IBM RS/6000 7014 Model T42 Enterprise Rack

The 2.0 meter (79.3 inches) Model T42 is the rack that will address the special requirements of customers who want a tall enclosure to house the maximum amount of equipment in the smallest possible floor space. The features that differ in the Model T42 rack from the Model T00 include the following:

- 42 EIA units (42U) of usable space
- AC power support only
- Weight:
  - T42 base empty rack: 261 kg (575 pounds)
  - T42 full rack: 930 kg (2045 pounds)

1.3.3 AC Power Distribution Units for rack models T00 and T42

For rack models T00 and T42 different Power Distribution Units (PDUs) are available. Previously, the only AC PDUs for these racks were PDUs with six outlets (FCs 9171, 9173, 9174, 6171, 6173, 6174). These PDUs do not support the power requirements of the p650.

Required: Required for the p650 are new PDUs with nine outlets available. These PDUs can only be mounted on the sides of the rack. A T00 configured for the maximum number of power outlets would have four PDUs, for a total of 36 power outlets. A T42 has an optional fifth and sixth additional PDU mounted horizontally in the rack. The base PDU (FC 9176, 9177, or 9178) determines the additional PDU (FC 7176, 7177, or 7178).

These new PDUs are *mandatory* for the p650 because of the potentially higher power loads of fully configured servers, but they also support other pSeries models. Figure 1-4 shows the new PDU.

An IBM 7014-T00 or T42 rack must have at least one Power Distribution Unit (PDU) per p650 system installed. It is recommended that each power supply in a p650 system be connected to a different PDU. No more than two p650 power supplies may be connected to the same PDU. Each PDU has nine C13 power connectors, in groups of three. When a p650 power supply is connected to one PDU power outlet, the other two connectors in that group of three must not be used.

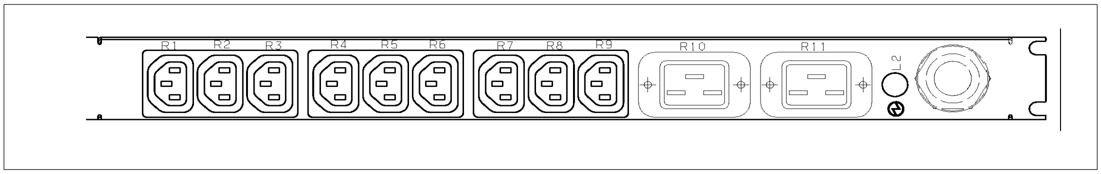


Figure 1-4 New PDU for p650 and other pSeries servers

Note: The installation of p650 into an existing rack may require the installation of the new PDUs, if they are not already present.

For p650, the initial PDU must be selected out of one of the following:

- FC 9176 Power Distribution Unit, base/side mount, single phase, L6-30 connector
- FC 9177 Power Distribution Unit, base/side mount, single phase, IEC-309 connector
- FC 9178 Power Distribution Unit, base/side mount, three phase, IEC-309 connector

Additional PDUs with nine power outlets can added to a configuration out of the following, based on which initial PDU (FC 9176, 9177, or 9178) was selected:

- FC 7176 Power Distribution Unit, side mount, single phase, L6-30 connector
- FC 7177 Power Distribution Unit, side mount, single phase, IEC-309 connector
- FC 7178 Power Distribution Unit, side mount, three phase, IEC-309 connector

1.3.4 OEM racks

The p650 can be installed in a suitable OEM rack, provided that the rack conforms to the EIA-310-D standard. This standard is published by the Electrical Industries Alliance, and a summary of this standard is available in the publication *Site and Hardware Planning Information*, SA38-0508. An online copy of this document can be found at:

http://www.ibm.com/servers/eserver/pseries/library/hardware_docs

Key points mentioned in this standard are as follows:

- Any rack used must be capable of supporting 15.9 kg (35 pounds) per EIA unit (44.5 mm (1.75 inch) of rack height).
- To ensure proper rail alignment, the rack must have mounting flanges that are at least 494 mm (19.45 inches) across the width of the rack and 719 mm (28.3 inches) between the front and rear rack flanges.
- It may be necessary to supply additional hardware, such as fasteners, for use in some manufacturers' racks.

1.3.5 Rack-mounting rules for p650, 7311-D10, and 7311-D20 I/O drawers

There are rules that should be followed when mounting drawers into a rack. The primary rules are as follows.

p650 server

The primary rules for the p650 are:

- The p650 is designed to be placed at any location in the rack. For rack stability reasons, it is advisable to start filling a rack from the bottom.
- A p650 is 8U in height, so a maximum of five p650 fit in a T42 rack or up to four p650 in a T00 rack. A fifth top-mount PDU is available in the T42 rack in order to meet the one PDU per p650 requirement.

7311-D10 I/O drawer

The primary rules for the 7311-D10 I/O drawer are:

- The 7311-D10 I/O drawer is a 4U half-wide drawer. It requires an enclosure (FC 7311) for mounting. The enclosure can accommodate two 7311-D10 drawers, side-by-side, but it may also be used with only one 7311-D10 drawer installed.

- 7311-D10 is designed to be placed at any location in the rack. For rack stability reasons, it is advisable to start filling an empty rack from the bottom and place I/O drawers above system units.
- The I/O drawers could be in the same rack as the p650 server or in an adjacent rack, although it is recommended that the I/O drawers be located in the same rack as the server for service purposes.
- A 7311-D10 I/O drawer enclosure is 4U in height, so a maximum of 18 7311-D10 (nine enclosures) fit in a T00 rack or 20 7311-D10 (10 enclosures) in a T42 rack.

7311-D20 I/O drawer

The primary rules for the 7311-D20 I/O drawer are:

- The 7311-D20 I/O drawer is a 4U full-wide drawer.
- 7311-D20 is designed to be placed at any location in the rack. For rack stability reasons, it is advisable to start filling an empty rack from the bottom and place I/O drawers above system units.
- The I/O drawers could be in the same rack as the p650 server or in an adjacent rack, although it is recommended that the I/O drawers be located in the same rack as the server for service purposes.
- A 7311-D20 I/O drawer enclosure is 4U in height, so a maximum of nine 7311-D20 fit in a T00 rack or ten 7311-D20 in a T42 rack.

Any remaining space in the rack can be used to install other systems or peripherals provided that the maximum permissible weight of the rack is not exceeded.

1.3.6 Flat panel display options

The IBM 7316-TF2 Flat Panel Console Kit may be installed in the system rack. This 1U console uses a 15-inch thin film transistor (TFT) LCD with a viewable area of 304.1 mm x 228.1 mm and a 1024 x 768 resolution. The 7316-TF2 Flat Panel Console Kit has the following attributes:

- Flat Panel Color Monitor.
- Rack tray for keyboard, monitor, and optional VGA switch with mounting brackets.
- IBM Space Saver 2 14.5-inch Keyboard that mounts in the Rack Keyboard Tray and is available as a feature in sixteen language configurations (the track point mouse is integral to the keyboard).

> **Note:** It is recommended that the 7316-TF2 be installed between EIA 20 to 25 of the rack for ease of use. The 7316-TF2 or any other graphics monitor requires a POWER GXT135P graphics accelerator (FC 2848 or FC 2849) to be installed in the server.

1.3.7 VGA switch

The VGA switch for the IBM 7316-TF2 allows for the connection of up to eight servers to a single keyboard, display, and mouse.

To help minimize cable clutter, multi-connector cables in lengths of 7, 12, and 20 feet are available. These cables can be used to connect the graphics adapter (required in each attached system), keyboard port, and mouse port of the attached servers to the switch or to connect between multiple switches in a tiered configuration. Using a two-level cascade arrangement, as many as 64 systems can be controlled from a single point.

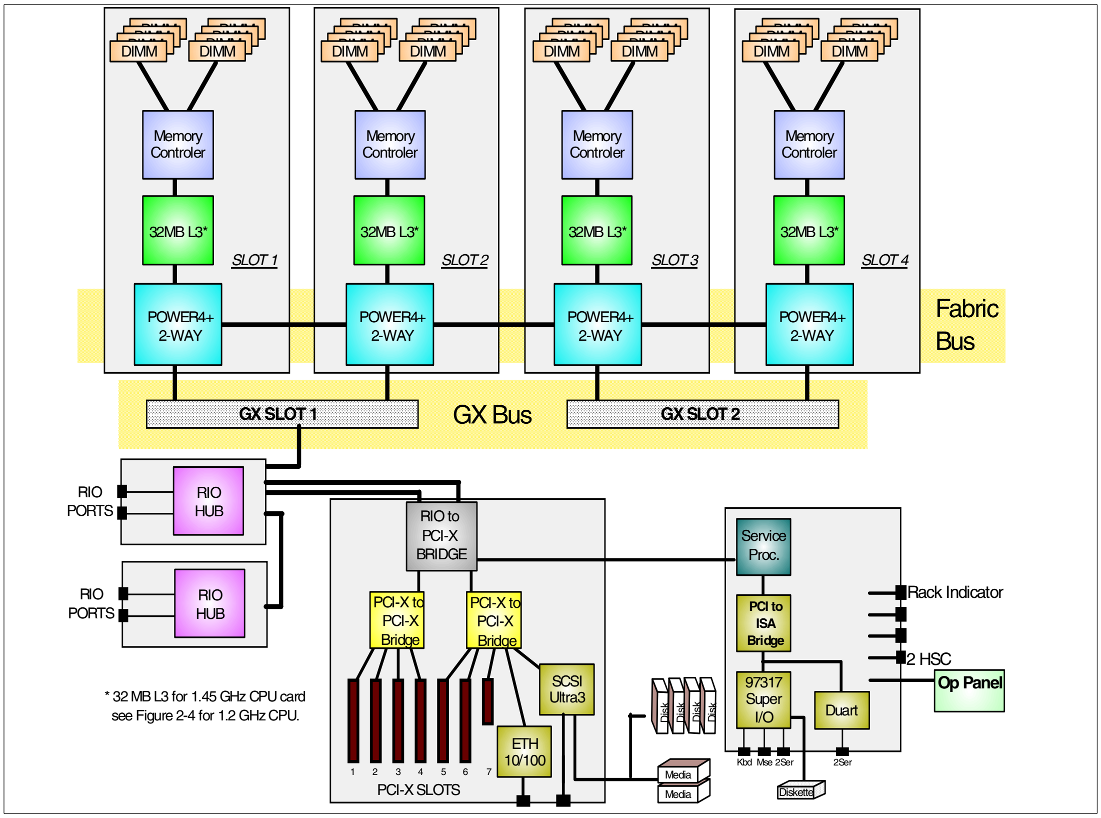
This dual-user switch allows attachment of one or two consoles, one of which must be an IBM 7316-TF2. An easy-to-use graphical interface allows fast switching between systems and supports six languages (English, French, Spanish, German, Italian, or Brazilian Portuguese).

The VGA switch is only 1U high and can be mounted in the same tray as the IBM 7316-TF2, thus conserving valuable rack space. It supports a maximum video resolution of 1600 x 1280, which facilitates the use of graphics-intensive applications and large monitors.

2

Architecture and technical overview

This chapter discusses the overall system architecture represented by Figure 2-1. The major components of this diagram will be described in this chapter. The bandwidths provided throughout this chapter are theoretical maximums provided for reference. It is always recommended to obtain real-world performance measurements using production workloads.



* 32 MB L3 for 1.45 GHz CPU card see Figure 2-4 for 1.2 GHz CPU.

Figure 2-1 Conceptual diagram of the p650 system architecture

2.1 Processor and cache

The p650 processor card contains one Single Chip Module (SCM), L3 cache, and memory. An SCM contains only one POWER4+ chip (a chip includes two cores; see Figure 2-4 on page 14) in contrast to Multichip Modules (MCMs), which contain four POWER4+ chips on one module and are used in the pSeries 655, 670, or 690. Each SCM is a Ceramic Column Grid Array (CCGA) package where the chip carrier is raised slightly from its board mounting by small metal solder columns that provide the required connections and improved thermal resilience characteristics.

2.1.1 PowerPC architecture

The p650 system complies with the RS/6000 platform architecture, which is an evolution of the PowerPC Common Hardware Reference Platform (CHRP) specifications. It is based on a POWER4+ processor chip. A view of the RS/6000 and pSeries processor evolution is shown in Figure 2-2.

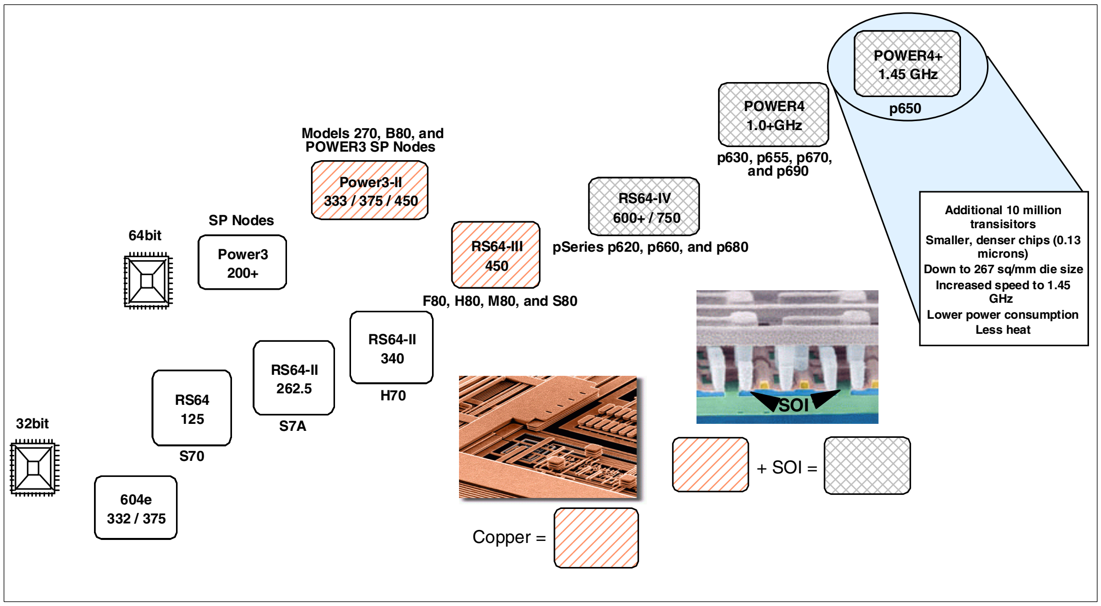


Figure 2-2 RS/6000 and IBM @server pSeries microprocessor direction

2.1.2 Copper and CMOS technology

The POWER4+ processor chip takes advantage of IBM's leadership technology. It is made using 0.13 µm-lithography CMOS[1] fabrication with seven levels of copper interconnect wiring. POWER4+ also uses Silicon-on-Insulator (SOI) technology.

In designing the POWER4+, the IBM engineering team was guided by a set of principles that included the following:

- *SMP optimization:* The system must be optimized for SMP operation, therefore advanced function is provided to allow SMP scaling.
- *Full-system design approach:* To optimize the system, the full design was in mind up front. This marriage of process technology, packaging, and micro-architecture was designed to allow software to exploit them. The entire system was designed together, from the processor to memory and I/O bridge chips.

[1] Complementary Metal Oxide Semiconductor

- *Very-high-frequency design:* The chip design was revamped with new transistor-level tools, and has transformed complex control logic into regular dataflow constructs, to deliver best-of-breed operating frequencies. The system design permits system balance to be preserved as technology improvements become available, allowing even higher processor frequencies to be delivered.
- *Leadership in reliability, availability, and serviceability:* Servers have evolved toward continuous operation. Our newer systems have RAS features previously seen only in mainframe systems. Where possible, if an error occurs, hard machine stops (checkstops) are now transformed into synchronous machine interrupts to software to allow the system to circumvent problems.
- *Balanced technical and commercial performance:* Balancing the system, the design could handle a varied and robust set of workloads. This is especially important as the e-business world evolves and data-intensive demands on systems merge with commercial requirements. The need to satisfy high performance computing (HPC) requirements with historical high-bandwidth demands and commercial requirements with their data-sharing and SMP scaling requirements required a single design to address both environments.

More information can be downloaded from the Internet at:

http://www.research.ibm.com/journal/rd/461/tendler.html

2.1.3 Processor cards

The processor chip (consisting of two processors, an L2 cache, and a bus controller), L3 cache, memory, memory controller, and card are all mounted in a rugged metal enclosure collectively named a *processor book*. This enclosure protects the contents (both in and out of the server), secures the card, and helps manage airflow used for cooling. Figure 2-3 shows a p650 1.45 GHz POWER4+ processor card.

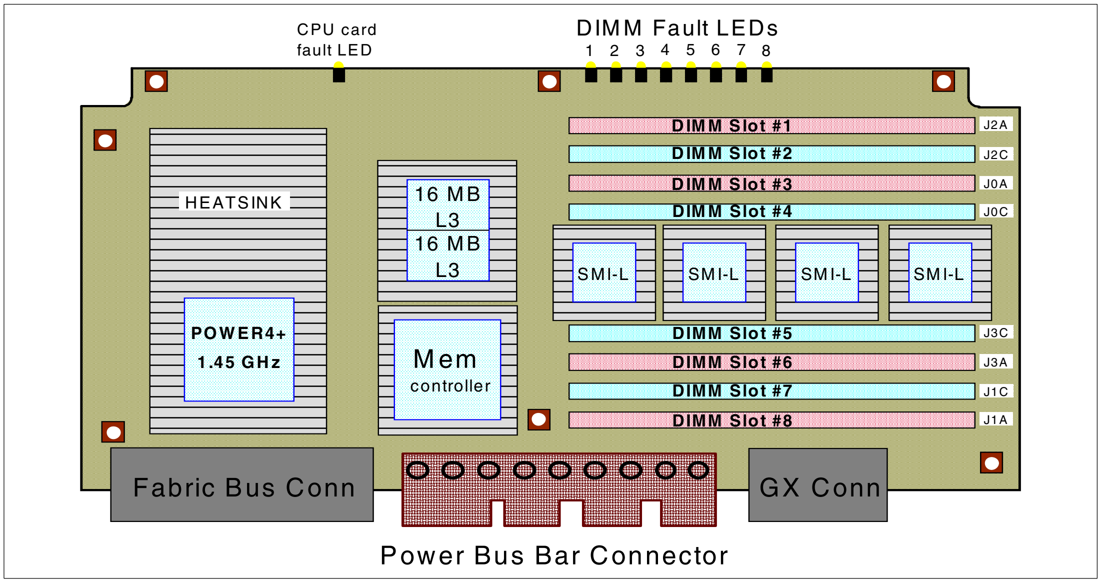


Figure 2-3 p650 1.45 GHz processor book

The POWER4+ processors used in p650 have either two 1.2 GHz or two 1.45 GHz processor cores. The p650 can contain up to four processor cards, and therefore four processor books, providing the system with either two, four, six, or eight processors.

> **Note:** A processor book should not be temporally removed from a system for testing purposes without changing the fabric bus backplane, because the fabric bus backplane must match the number of processor books present in the system. For details see Figure 2-7 on page 19.

Memory access is through the on-chip L2 cache and L3 cache directory controller to the off-chip L3 cache and finally through the memory controller and synchronous memory interface (SMI) to the memory DIMMs, as represented in Figure 2-4.

There is a fabric bus used for communication between the different SCMs. The processor to fabric bus clock speed ratio is 2:1. The fabric bus uses 8 bytes to read and 8 bytes to write, giving an aggregate peak bandwidth of 5.8 GB/s for the 1.45 GHz processor (23.2 GB/s in an 8-way system) and 4.8 GB/s for the 1.2 GHz processor (19.2 GB/s in an 8-way system).

The processor I/O path is through the GX bus. The processor-to-GX bus clock speed ratio is 3:1. The GX bus uses 8 bytes to read and 8 bytes to write, giving a bandwidth of 3.87 GB/s for a 1.45 GHz processor (15.47 GB/s in an 8-way system) and 3.2 GB/s for a 1.2 GHz processor (12.8 GB/s in an 8-way system).

Figure 2-4 shows a conceptual diagram of the processor and cache.

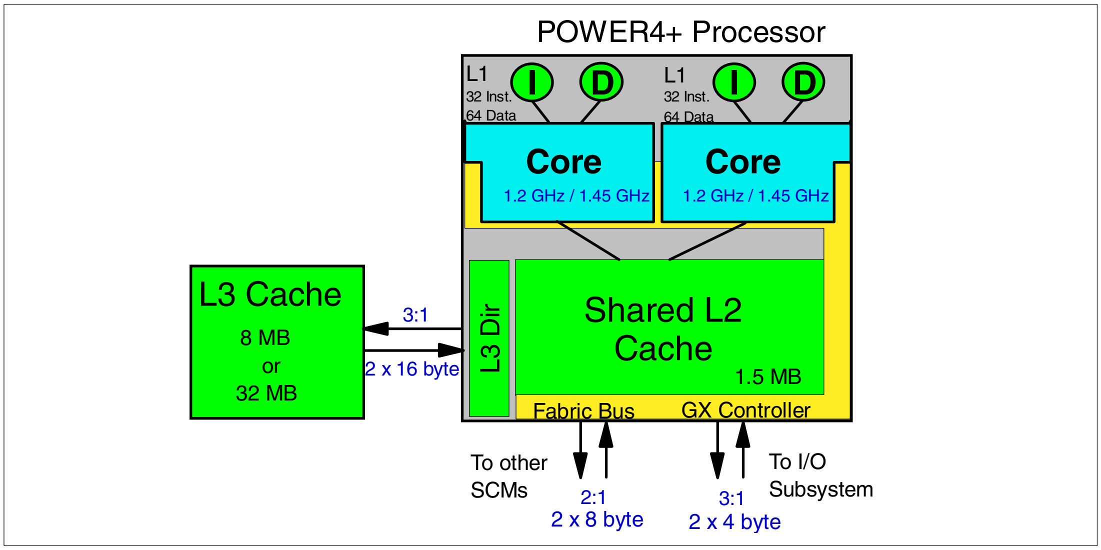


Figure 2-4 Conceptual diagram of processor and cache

Processor clock rate

The p650 operates with a processor clock rate of 1.2 GHz or 1.45 GHz. To determine the processor characteristics on a running system, use one of the following commands.

`pmcycles -m` This command (AIX 5L Version 5.1 and higher) uses the Performance Monitor cycle counter and the processor real-time clock to measure the actual processor clock speed in MHz. The following is the output of a 2-way p650 1.45 GHz system:

```
Cpu 0 runs at 1450 MHz
Cpu 1 runs at 1450 MHz
```

> **Note:** The **`pmcycles`** command is part of the bos.pmapi fileset. First check if that component is installed using the **`lslpp -l bos.pmapi`** command.

lsattr -El proc*X* Where *X* is the number of the processor (for example, proc0 is the first processor in the system). The output from the command[2] would be similar to the following (False, as used in this output, signifies that the value cannot be changed through an AIX command interface):

```
state enable           Processor state  False
type powerPC_POWER4    Processor type   False
frequency 145600000    Processor Speed  False
```

2.1.4 L1, L2, and L3 cache

The POWER4+ storage subsystem consists of three levels of cache and the memory subsystem. The first two levels of cache are onboard the POWER4+ chip. The first level is 64 KB of Instruction (I) and 32 KB of Data (D) cache per processor core. The second level is 1.5 MB of L2 cache, which is a small increase over the POWER4 processor (1.44 MB). Both cores share the L2 cache. All caches have either full ECC[3] or parity protection on the data arrays, and the L1 cache has the ability to re-fetch data from the L2 cache in the event of soft errors detected by parity checking.

The Level 3 (L3) cache consists of two components: The L3 cache controller/directory and the L3 data array. The L3 controller and cache directory are on the POWER4+ chip. For the 1.45 GHz processor card, the L3 data array is 32 MB in size and consists of two 16 MB embedded DRAM (eDRAM) chips on a separate module. For the 1.2 GHz processor card, the L3 cache is 8 MB in size and is included on one module together with the memory controller.

The processor-to-L3 cache clock speed ratio is 3:1; therefore, the L3 bandwidth for the 1.45 GHz processor is 15.47 GB/s (61.9 GB/s for an 8-way system) and 12.8 GB/s for the 1.2 GHz processor (51.2 GB/s for an 8-way system).

2.1.5 Shared cache mode

For the two- and four-processor card (4-way and 8-way) 1.45 GHz p650, the L3 can be configured in either private or shared mode. In shared mode, the L3s on each of the processor cards act as a single unified L3. In this mode each L3 caches' data is retrieved from the memory on the card the L3 is located on, independent of which processor makes the request. In this way there is no duplication of data in any of the L3s. In a shared L3 mode 2-card, 4-way system, the L3 acts as a single 64 MB L3. Similarly, in a shared L3 mode 4-card, 8-way system, the L3 acts as a single 128 MB L3. In this mode, from an L3 perspective, the whole system acts much like it is a single MCM on the p670 and p690. In all other configurations, the L3 can only operate in private mode. When in private mode, the L3s on each processor card act like the L3s attached to separate MCMs on the p670 and p690. When in private mode, the user has options for configuring AIX memory allocation schemas. Reference AIX 5L product documentation for additional details.

If the system has the right configuration of CPU cards and memory, L3 shared mode is automatically enabled by default and it can be changed for a private L3 during the initial system boot using the L3 Mode menu. The system continues to operate in this mode until changed. Changes take effect on the next system reboot. See the following article for more detail:

http://www.research.ibm.com/journal/rd/461/tendler.html

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

[3] ECC single error correct, double error detect.

2.2 Memory subsystem

The memory subsystem is different for the two processor cards. Figure 2-5 shows the conceptual diagram of the two different memory subsystems.

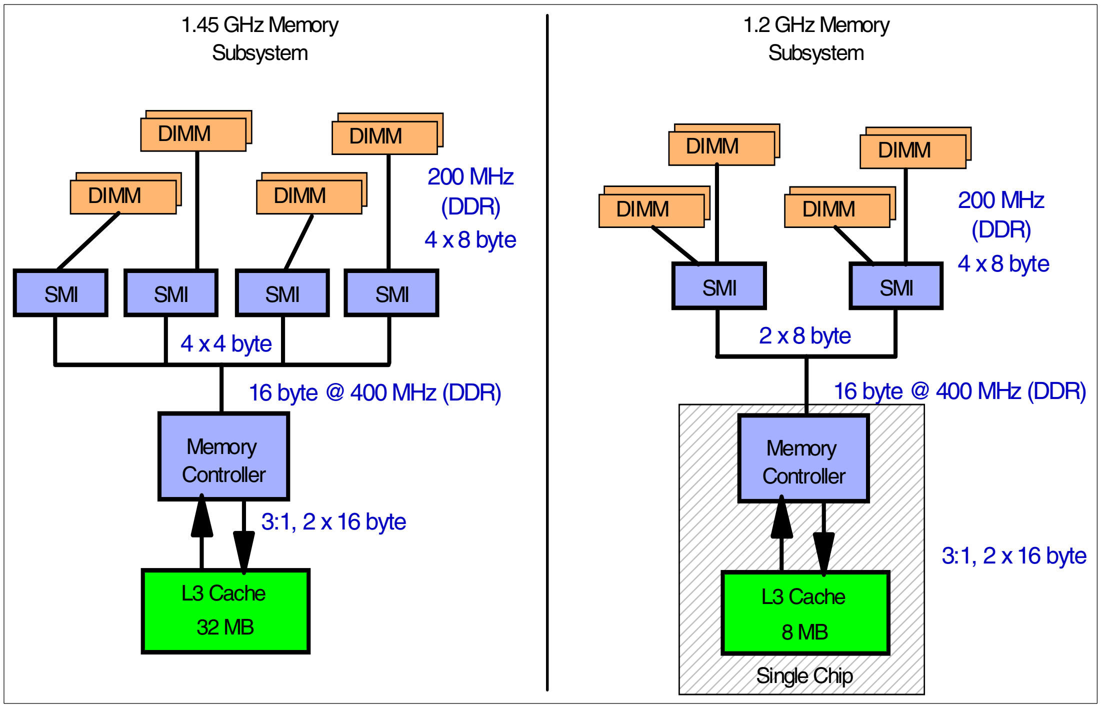


Figure 2-5 Conceptual diagram of memory subsystem

The 1.45 GHz system is equipped with a separate memory controller. This memory controller is designed to work with a 32 MB L3 cache module, instead of 8 MB in an 1.2 GHz system, enabling better performance.

Tip: In systems with two or four processor cards, the option exists to operate the L3 caches in shared mode, providing a single 64 MB or 128 MB cache, respectively, shared across all processors. In many applications this provides improved throughput over the private L3 cache configuration.

The conceptual diagram of a 1.45 GHz system's memory subsystem shows four 8-byte data paths to memory with an aggregated bandwidth of 6.4 GB/s (25.6 GB/s in an 8-way system). Each processor card can hold up to eight double data rate (DDR) synchronous DRAM (SDRAM) DIMMs, which must be populated in quads. Figure 2-3 on page 13 shows the two possible quads that are highlighted in different colors. Slots 1, 3, 5, and 7 represent the first quad, and slots 2, 4, 6, and 8 represent the second quad.

In a 1.2 GHz system, the 8 MB L3 cache and the memory controller are on one chip. This implementation also provides four 64-bit data paths to memory with the same aggregated bandwidth of 6.4 GB/s (25.6 GB/s in an 8-way system). There are also eight DIMM slots, which also have to be equipped in quads following the same rules as in a 1.45 GHz system.

DDR memory can theoretically double memory throughput at a given clock speed by providing output on both the rising and falling edges of the clock signal (rather than just on the rising edge).

Memory must be balanced across the processor cards. See 2.2.1, “Memory options” on page 17 and Table 2-1 on page 17 for further details.

2.2.1 Memory options

The following memory features for the p650 are available:

FC 4452	2048 MB (4 x 512 MB) 208 pin 8 ns DDR SDRAM DIMMs
FC 4453	4096 MB (4 x 1024 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs
FC 4454	8192 MB (4 x 2048 MB) 208 pin 8 ns stacked DDR SDRAM DIMMs

Each memory feature consists of four DIMMs, or a quad. The minimum amount of memory for p650 is 2 GB (one FC 4452); the maximum with all four processor cards fully populated is 64 GB.

Table 2-1 shows the recommended memory configurations. The values in the chart represent the memory size in GB, not the quantity. Notice they are balanced configurations.

Table 2-1 Recommended memory configurations

Total memory	**One proc. card**	**Two processor cards**		**Three processor cards**			**Four processor cards**			
		1st	**2nd**	**1st**	**2nd**	**3rd**	**1st**	**2nd**	**3rd**	**4th**
2 GB	2									
4 GB	4	2	2							
6 GB	4 + 2			2	2	2				
8 GB	4 + 4	4	4				2	2	2	2
10 GB	8 + 2									
12 GB	8 + 4	4 + 2	4 + 2	4	4	4				
16 GB	8 + 8	4 + 4	4 + 4				4	4	4	4
18 GB				4 + 2	4 + 2	4 + 2				
20 GB		8 + 2	8 + 2							
24 GB		8 + 4	8 + 4	8	8	8	4 + 2	4 + 2	4 + 2	4 + 2
30 GB				8 + 2	8 + 2	8 + 2				
32 GB		8 + 8	8 + 8				8	8	8	8
36 GB				8 + 4	8 + 4	8 + 4				
40 GB							8 + 2	8 + 2	8 + 2	8 + 2
48 GB				8 + 8	8 + 8	8 + 8	8 + 4	8 + 4	8 + 4	8 + 4
64 GB							8 + 8	8 + 8	8 + 8	8 + 8

2.3 System buses

The p650 has three main backplane cards building the internal system buses: GX backplane, the fabric bus backplane, and the I/O subsystem that is made by a sandwich of two cards (the

I/O backplane and the CSP[4] card). Each processor card has two connectors (see Figure 2-3 on page 13): The fabric bus connector goes into the fabric bus backplane, and the GX connector provides the GX bus, which is used to connect to the I/O subsystem. The GX bus provides an interface to a single device like a RIO hub. The GX bus and GX slots are contained on the GX backplane.

Note: For original orders with a single 2-way processor card, the fabric bus backplane is not required, but one of three different fabric bus backplanes is required for an order with more than one 2-way processor card. See 2.3.1, “Fabric bus and fabric bus backplane” on page 18, for more details.

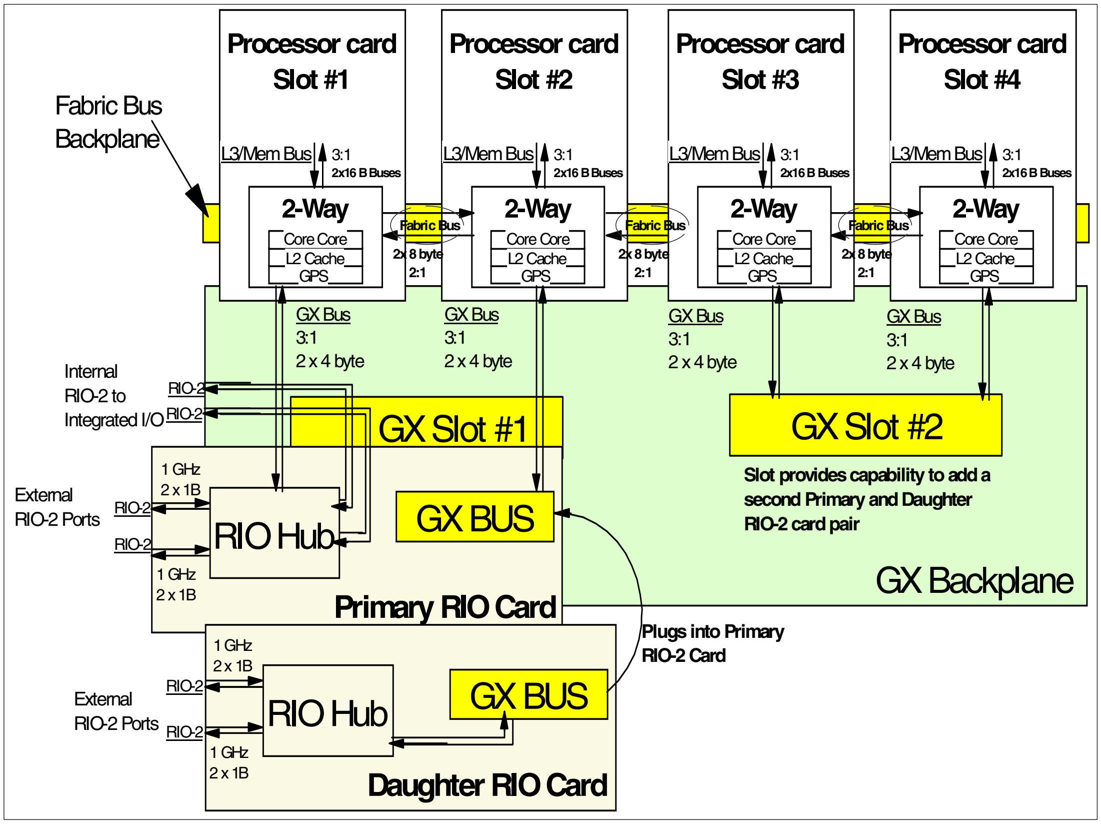


Figure 2-6 GX buses, fabric buses, and RIO-2 features

The p650 processor is supported with a distributed memory architecture that provides high memory bandwidth. Although each processor can address all memory and access a single shared memory resource, memory is distributed locally behind individual 2-way processor cards. As such, two processors (one card) and their associated L3s and memory are packaged inside the processor card. Access to memory associated with another processor is accomplished through the fabric bus.

2.3.1 Fabric bus and fabric bus backplane

The fabric bus consists of four parallel busses that are point-to-point but similar in design to the 6XX bus used in previous RS/6000 products. The primary difference from the 6XX bus is the point-to-point configuration, which means that the fabric bus is not traversed during

[4] CSP stands for Converged Service Processor. Since the release of 7025 Model F80, 7026 Model H80, and M80, the RS/6000 (pSeries) Service Processor design converged to AS/400 (iSeries) Service Processor design.

normal operations, there are no arbitration signals on the bus, and transfers are not seen by all bus occupants at the same time. Since the fabric bus is a looped bus, there are multiple versions of the fabric bus backplane for each processor card configuration. The fabric bus backplane must match the number of processor cards present in the system. The fabric bus going between the SCMs is architecturally identical to the inter-MCM bus in the p690.

Table 2-2 lists the required features to support the fabric bus.

Table 2-2 Fabric bus backplane requirements

Processor configuration	FC required	Card description
1 x 2-way processor card	None	No fabric bus backplane required
2 x 2-way processor card	5120	2-slot fabric bus backplane
3 x 2-way processor card	5123	3-slot fabric bus backplane
4 x 2-way processor card	5124	4-slot fabric bus backplane

The p650 supports four pluggable 2-way processor cards. Processor upgrades are achieved by simply adding processor cards and adding or replacing the appropriate fabric bus backplane. See Figure 2-7 for reference.

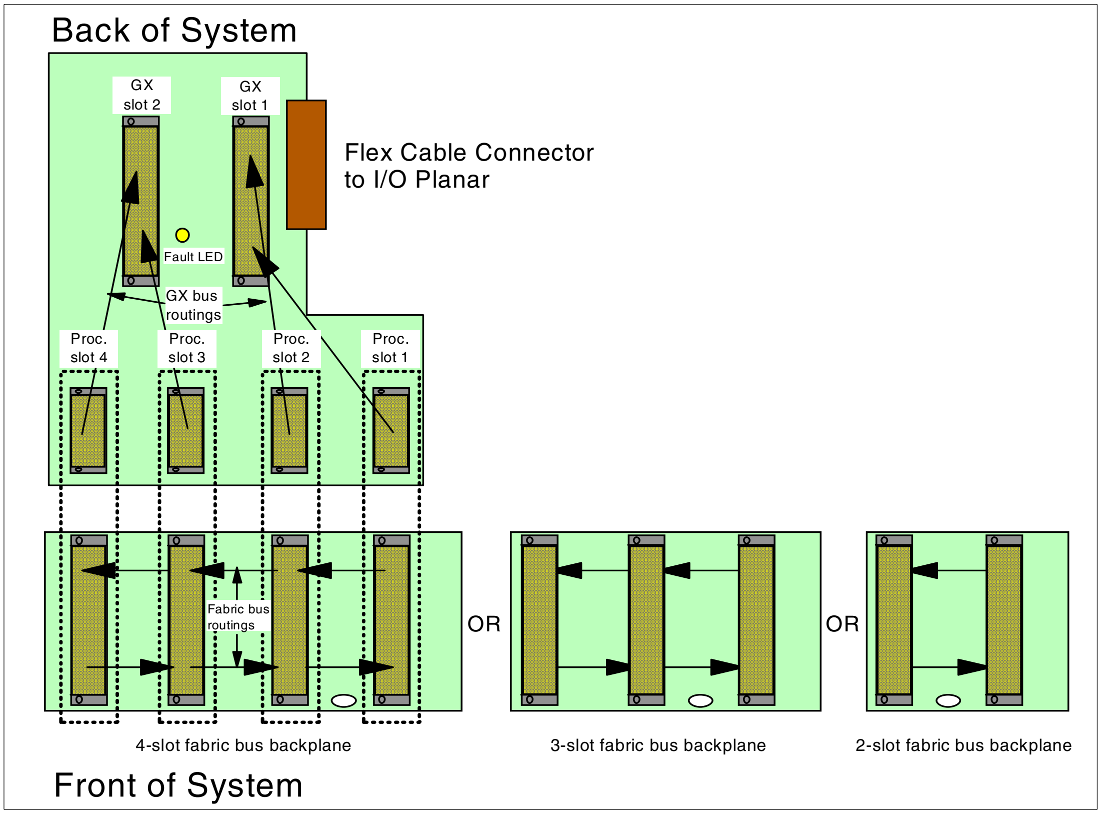


Figure 2-7 GX backplane and fabric bus backplanes

2.3.2 GX bus

The p650 has one GX backplane with two GX slots; each slot contains two GX buses. I/O connects to the processor and memory subsystem using the GX bus. Each processor card provides a single GX bus for a total system capability of four GX buses. GX slot 1 contains the

GX buses routed from the processor cards in slots 1 and 2, and GX slot 2 contains GX buses routed from the processor cards in slots 3 and 4. GX slots are not active unless the corresponding processor card is installed.

2.3.3 RIO buses and GX cards

There are two GX cards that provide RIO or RIO-2 I/O expansion. One is the Primary RIO or RIO-2 card and the other is the Daughter card:

- Primary RIO card (FC 6411) or RIO-2 card (FC 6415).
- Daughter RIO card (FC 6412) or RIO-2 card (FC 6416).

Each RIO port can operates at 500 MHz (RIO) or 1 GHz (RIO-2) in bidirectional mode and is capable of passing data in each direction on each cycle of the port. Therefore, the maximum data rate is 2 GB/s (RIO) or 4 GB/s (RIO-2) per I/O drawer in double barrel mode.

The first Primary RIO card plugs into GX slot 1 and is required as a base feature card for p650 (see Figure 2-6 on page 18). This feature provides two Remote I/O ports for attaching up to four 7311-D10 or up to two 7311-D20 I/O drawers to the system in a single loop. It connects to the GX bus from the processor card in the first processor slot and it provides four RIO ports, two of which are used to drive the p650’s internal I/O, and two RIO ports are routed to an external connector for I/O drawer expansion.

The Daughter RIO card is a feature card that plugs into the Primary RIO card, providing two remote I/O ports for attaching up to four 7311-D10 or up to two 7311-D20 I/O drawers to the system in a single loop, and it is connected to the GX bus from the processor in slot 2.

The Primary RIO card routes the second GX bus from the GX slot on the GX backplane to a daughter connector for the Daughter card. The RIO hub on the base Primary RIO card is configured for two active RIO ports to the internal I/O Host Bridge chip. The internal RIO bus runs in double barrel mode, which means that the RIO loop is dedicated only to the internal I/O subsystem. External RIO buses will run in double barrel mode as long as only one I/O drawer is present in a RIO loop. The maximum supported number of Primary RIO cards is two, the same as for the Daughter RIO card.

Note: The number of Primary and Daughter RIO cards configured may depend on the number of I/O drawers and the system performance that you want to achieve in a new system order, and not only by the number of processor cards.

Figure 2-8 provides an example of optimized configurations with the maximum number of 7311-D10 I/O drawers.

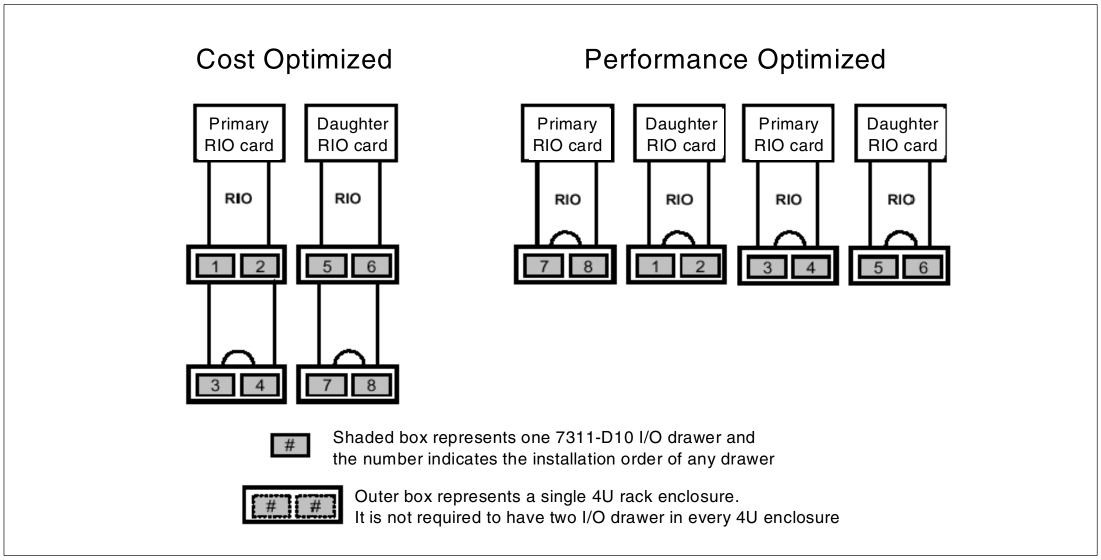

Cost Optimized
Performance Optimized
Primary RIO card
Daughter RIO card
RIO
1
2
3
4
5
6
7
8
#
Shaded box represents one 7311-D10 I/O drawer and the number indicates the installation order of any drawer
Outer box represents a single 4U rack enclosure. It is not required to have two I/O drawer in every 4U enclosure

Figure 2-8 7311-D10 I/O drawer attached per RIO loop versus system performance

Figure 2-9 on page 21 provides an example of optimized configurations with the maximum number of 7311-D20 I/O drawers.

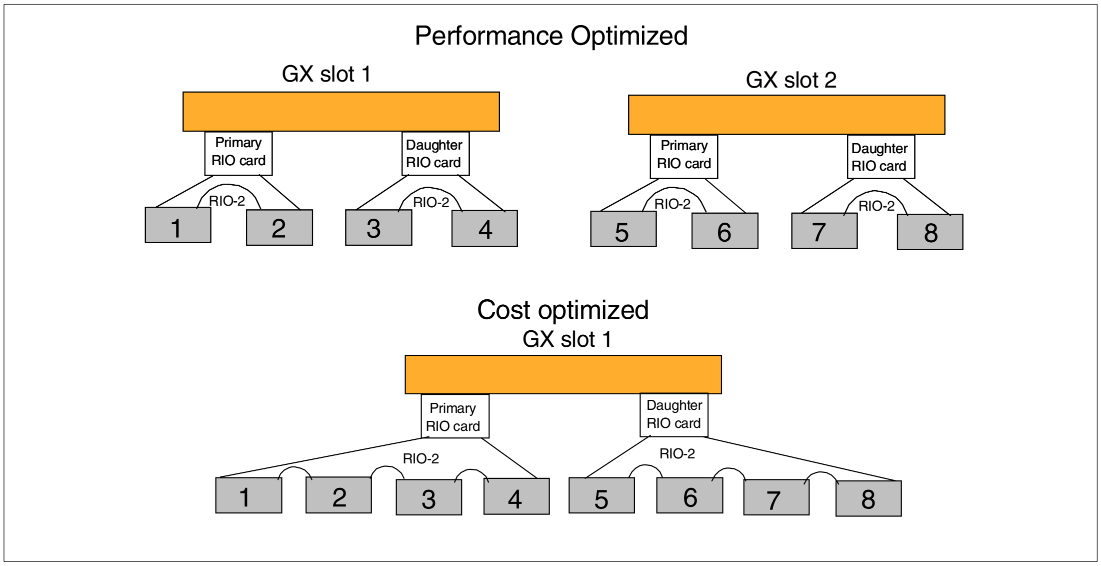


Figure 2-9 7311-D20 I/O drawer attached per RIO loop versus system performance

Optimized solutions are not required in all cases (see “General order rules, restrictions” on page 22).

Any RIO loop supports a maximum of four 7311-D10 or two 7311-D20 I/O drawers; a p650 supports a maximum of eight I/O drawers.

Table 2-3 shows how an initial order configuration would plug the I/O drawers for best performance according to the number of Primary RIO cards (FC 6411), Daughter RIO cards (FC 6412), and the number of 7311-D10 I/O drawers in the system.

Table 2-3 RIO GX card and I/O drawer order configuration

# of processor cards =>		One				Two				Three				Four			
# of 7311-D10	**RIO card =>**	**1P**	**1D**	**2P**	**2D**	**1P**	**1D**	**2P**	**2D**	**1P**	**1D**	**2P**	**2D**	**1P**	**1D**	**2P**	**2D**
1		1					1				1				1		
2		2				1	1				1	1			1	1	
3		3				1	2			1	1	1			1	1	1
4		4				2	2			1	2	1		1	1	1	1
5						2	3			1	2	2		1	2	1	1
6						3	3			2	2	2		1	2	2	1
7						3	4			2	3	2		1	2	2	2
8						4	4			2	3	3		2	2	2	2

Legend:
1P=GX slot 1 Primary RIO card FC 6411 or RIO-2 card FC 6415
1D=GX slot 1 Daughter RIO card FC 6412 or RIO-2 card FC 6416
2P=GX slot 2 Primary RIO card FC 6411 or RIO-2 card FC 6415
2D=GX slot 2 Daughter RIO card FC 6412 or RIO-2 card FC 6416

Example 1

If you have five 7311-D10 I/O drawers, one Primary RIO card, and one Daughter RIO card, the suggested way to plug the drawers into the RIO cards is two I/O drawers in the first Primary loop and three I/O drawers in the first Daughter loop.

Example 2

If you have five 7311-D10 I/O drawers, two Primary RIO cards, and one Daughter RIO card, then plug one I/O drawer into the first Primary loop, two I/O drawers into the first Daughter loop, and two I/O drawers into the second Daughter loop.

General order rules, restrictions

The following list defines the configuration rules and restrictions:

- A Primary RIO or RIO-2 card consumes a GX slot and hence a maximum of two per machine.
- The sum of Primary RIO or RIO-2 cards and Daughter RIO or RIO-2 cards must be less than or equal to the sum of the processor cards.
- You do not need to configure more than the initial Primary RIO or RIO-2 card on a system with two or more processor cards unless expansion or performance requirements mandate it.
- A second Primary RIO or RIO-2 card without first a Daughter RIO or RIO-2 card is not allowed.
- Two Primary RIO or RIO-2 cards require three or more processors.
- The p650 requires at least one Primary RIO card.
- Every Daughter RIO or RIO-2 card requires a Primary RIO or RIO-2 card.
- A maximum of eight 7311-D10 or 7311-D20 I/O drawers per p650.
- The first Primary RIO or RIO-2 card can have zero to four 7311-D10 or zero to two 7311-D20 I/O drawers.

- The first Daughter card can have zero to four 7311-D10 or zero to two 7311-D20 I/O drawers, and the total on the first Daughter card must be greater or equal to the I/O drawers cabled to the second Primary RIO or RIO-2 card.
- The second Primary RIO or RIO-2 card can have zero to four 7311-D10 or zero to two 7311-D20 I/O drawers, and the total must be greater or equal to the second Daughter RIO card.
- The second Daughter RIO or RIO-2 card can have zero to two I/O drawers, and the total must be lesser than or equal to the second Primary RIO card.
- The 7311-D10 and 7311-D20 I/O drawers cannot be in the same I/O loop.
- The system must have a RIO-2 Enabled System Planar (FC 9581 or FC 6581) and FC 6415 in the first primary RIO card slot to achieve the higher RIO performance of RIO-2 for internal I/O.
- All adapters in a RIO loop must be RIO-2 to achieve RIO-2 performance.
- If RIO and RIO-2 adapters are in the same loop, it will operate at the standard RIO speed.

2.4 Internal I/O subsystem

The internal I/O subsystem consists of a sandwich of two cards, packaged together as a single field replacement unit (FRU). There is an internal RIO bus enhanced flex cable that comes from the GX backplane to the I/O subsystem flex cable connector (see Figure 2-10). The logic is partitioned such that the bottom card contains the Converged Service Processor function and the top card contains the PCI-X slots and Integrated I/O.

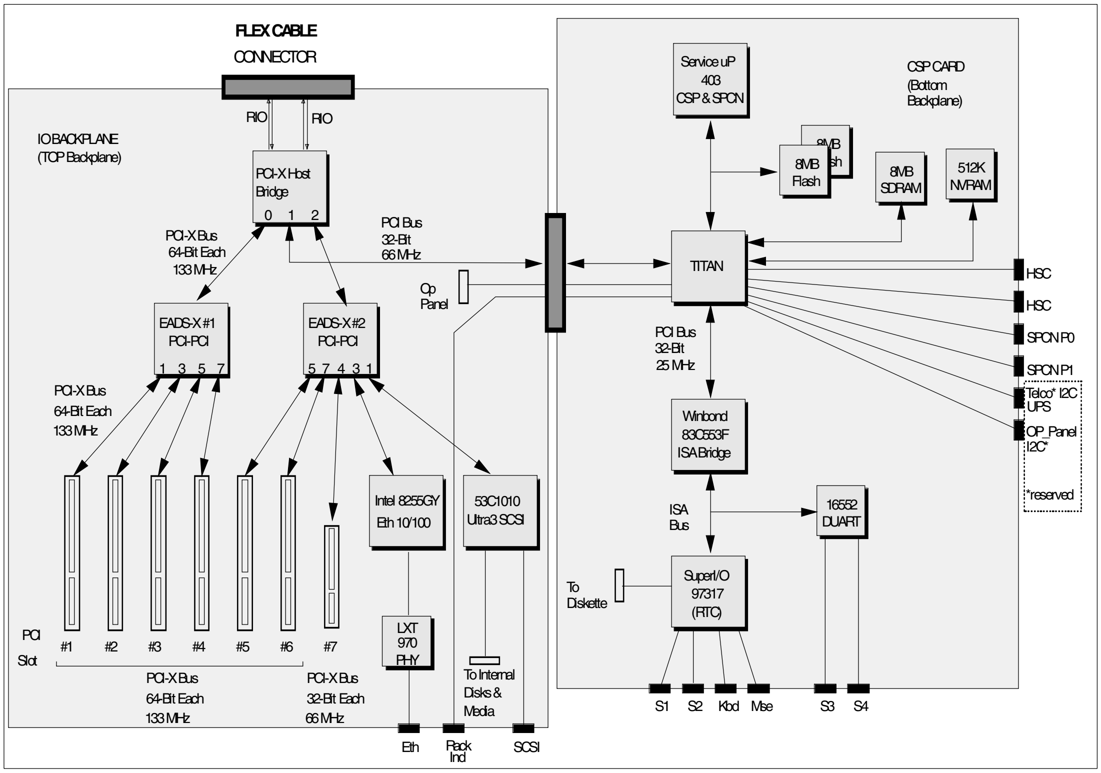


Figure 2-10 p650 internal I/O subsystem

2.4.1 PCI-X subsystem

The p650 internal RIO loop connects the first Primary RIO or RIO-2 GX card to the I/O subsystem through the PCI-X Host bridge. It includes three PCI-X buses. Bus 0 and bus 2 are connected to the PCI-X to PCI-X chip (EADS-X). The first EADS-X chip is connected to the first four PCI-X slots. The second EADS-X chip is connected to the last three slots and the integrated adapters.

PCI-X is the latest version of PCI bus technology, using a higher clock speed (133 MHz) to deliver a bandwidth of up to 1 GB/s supporting 3.3 volt adapters. The PCI-X slots in the p650 support blind-swap hot-plug and Extended Error Handling (EEH). EEH-capable adapters respond to a specially generated data packet from a PCI/PCI-X slot with a problem. This packet is analyzed by the system firmware, which then allows the device driver to reset the adapter or slot, isolating the error and reducing the need for the system reboot. In addition to that, the p650 also gives the ability to concurrently change or add PCI/PCI-X adapters and the disks within the system in the same way, without the need for a system reboot.

The addition, removal, or changing of a PCI/PCI-X adapter can be accomplished by using a system management tool, such as the Web-based System Manager or the PCI Hot-Plug Manager (using SMIT). The PCI hot-plug tasks can be accomplished also by using the Hot-Plug Task of the online diagnostics task selection menu (using the **`diag`** command). Each of these tools provide a method so that a PCI/PCI-X slot can be identified first, powered off to enable removal or insertion of an adapter, and then powered back on to enable the device to be configured. The p650 also provides an updated blind-swap cassette mechanism to change (replace), add, or remove an adapter for both internal PCI-X slots and I/O drawer PCI/PCI-X slots (see 2.5, “I/O drawer” on page 27, for cassette FCs and more information).

2.4.2 Adapters

In general, PCI-X slots also support existing 3.3 volt or universal signalling PCI/PCI-X adapters, which include both 64-bit and 32-bit adapters packaged in new blind-swap cassettes. The only limitation is for a 5 volt PCI adapter (not universal) that is only supported in the first PCI slot of the 7311-D10 I/O drawer. See 2.5, “I/O drawer” on page 27, for more information.

Choosing between 32-bit and 64-bit adapters influences slot placement and affects performance. Higher speed adapters use 64-bit slots because they can transfer 64 bits of data for each data transfer phase. 32-bit adapters can typically function in 64-bit PCI-X slots. However, 32-bit adapters still operate in 32-bit mode and achieve no performance advantage in a 64-bit slot.

For a full list of the PCI/PCI-X adapters that the Model p650 supports, and for important information regarding the rules of adapter placement, see the *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538, for additional information. You can find this publication at:

http://www.ibm.com/servers/eserver/pseries/library/hardware_docs/pci_adp_pl.html

Integrated adapters

The Ethernet controller (PCI, 32-bit) operates at 33 MHz.

The SCSI controller (PCI, 64-bit) operates at 66 MHz and has dual low-voltage differential (LVD) ports. The external port attaches to the bulkhead using a 68-pin cable. The internal port connects to the disks backplane also through a 68-pin SCSI cable.

Graphics adapters

The p650 is a server and is not intended to serve as a workstation. Therefore, the POWER GXT135P (FC 2848 or FC 2849) is the only graphics adapter available. This adapter offers 2D function or 2D function with digital support for business graphics, Internet applications, or for those applications that require a graphics display for installation and management.

LAN adapters

Since the p650 is usually connected to a local area network (LAN), the internal 10/100 Mbps Ethernet integrated adapter, situated on the system board, can be used to accomplish that.

> **Tip:** In conjunction with certain network switches, you can use the Cisco Systems' EtherChannel feature of AIX to build one virtual Ethernet interface with increased bandwidth using up to eight Ethernet interfaces (adapters or integrated).

Other LAN connection options include: Gigabit Ethernet, dual-port Gigabit Ethernet, 4-port Ethernet, token-ring, and ATM. IBM supports an installation with NIM using Ethernet (including the integrated adapter; see 2.6.5, "Additional boot options" on page 34), token-ring adapters, or FDDI adapters (use chrp as the platform type).

Storage adapters

System storage can be added by attaching SCSI, SSA, or Fibre Channel adapters to the storage subsystems, in addition to the dual port integrated Ultra3 SCSI adapter.

2.4.3 Disks and media backplanes

The standard 4-pack disk backplane is composed of two cards, while the media backplane is a single card. Both backplanes are connected to the internal SCSI controller, which is not RAID capable, but the backplanes contain a SCSI Enclosure Service (SES) module to support hot-swap for disks and media. The SES processor is the SCSI hot-swap manager; it provides the control mechanism for the device hot-swap options such as identify/replace/remove.

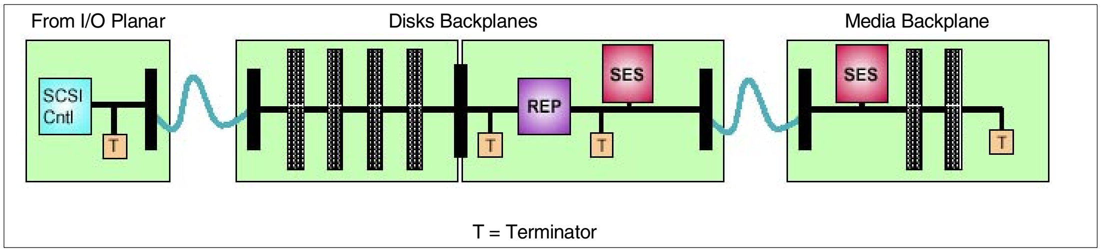


Figure 2-11 Disk and media backplanes

The 4-pack disk backplanes and the disk cage are combined into a single field replacement unit (FRU) and they provide support for the following:

- Four 1-inch disk drives
- Ultra3 speed capability (SE not supported)
- SES module for hot-swap capability
- U3 Carriers (U2 also for migration)
- Repeater module to allow connection to media devices
- VPD module

The media backplane supports the following:

- Ultra2 and SE capability
- VPD support
- Auto-docking media bays (can be removed and replaced with simple snap-in action)
- Connection from the disk backplane

An optional split SCSI disk backplane (FC 6579) is available for the p650. It divides the internal disks into two sets of two independent groups of two disks, ideal when the minimum requirements for two partitions must be met with internal I/O. It must be connected to the cable assembly, which includes a special cassette that occupies PCI slot 7, and provides two external SCSI ports for connecting the two sides of the split disk backplane to separate SCSI adapters, each side supporting two disks. External connections to these ports are provided by cable (FC 4262) that are connected to the external integrated SCSI port or an Ultra3 SCSI Adapter.

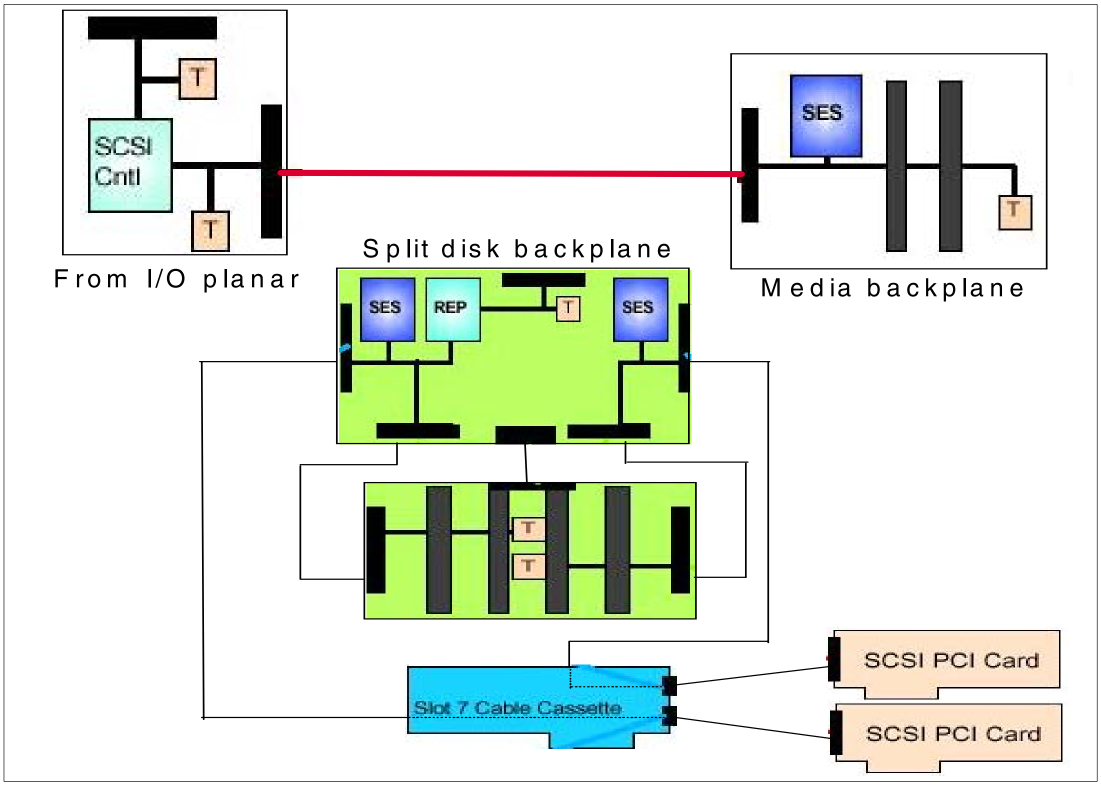


Figure 2-12 Split disk and media backplanes

Note: In the p650, if the system includes a split disk backplane, the media backplane must be connected to the integrated internal SCSI port.

2.4.4 Internal hot-swappable SCSI devices

The p650 can have up to four hot-swappable drives using 4-pack disk backplane or two independent pair of hot-swappable drives using split disk backplane in the front hot-swappable disk bays and up to two hot-swappable media devices in the front hot-swappable media bay. The hot-swap process is controlled by the SCSI enclosure services (SES).

Prior to the hot-swap of a disk or media in the hot-swappable bay, all necessary operating system actions must be undertaken to ensure that the disk or media is capable of being deconfigured.

Once the disk drive or media has been deconfigured, the SCSI enclosure device will power off the slot, enabling safe removal of the device. You should ensure that appropriate planning has been given to any operating system related disk layout, such as the AIX Logical Volume Manager, when using disk hot-swap capabilities. For more information, see *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496.

Note: The new auto-docking mechanism for the hot-swappable media devices requires dedicated media feature codes. Existing media bay devices from earlier systems are not interchangeable. Refer to “Disk and media” on page 4.

2.5 I/O drawer

The p650 CEC[5] has seven internal PCI-X slots, which is enough in many cases. If more PCI-X slots are needed, especially well-suited for LPAR mode, up to eight 7311-D10 or 7311-D20 I/O drawers can be added to the system.

2.5.1 7311-D10 I/O drawer

The 7311-D10 is a 4U half-wide drawer that should be mounted in a rack enclosure where two 7311-D10 could be mounted side-by-side. It features five hot-pluggable PCI-X slots and one standard hot-plug PCI slot. The 7311-D10 I/O drawer includes redundant concurrently maintainable power and cooling and is the only I/O drawer that can be attached to the p650. The 7311-D10 I/O drawer does not slide out from the enclosure on rails, and therefore must be removed for service.

The 7311-D10 I/O drawer is connected with a Remote I/O (RIO) or Remote I/O 2 (RIO-2) loop to the CEC. Each RIO port can operates at 500 MHz (RIO) or 1 GHz (RIO-2) in bidirectional mode and is capable of passing up to eight bits of data in each direction on each cycle of the RIO port. Therefore, the maximum data rate is 2 GB/s (RIO) or 4 GB/s (RIO-2) per I/O drawer in double barrel mode.

The p650 to achieve the higher Remote I/O performance of RIO-2 must have a RIO-2 Enabled System Planar (FC 9581 or FC 6581), all adapters (FC 6415 and FC 6416) in a RIO loop must be RIO-2 and 7311-D10 must have RIO-2 riser adapter (FC 6413).

The I/O drawer enclosure has the following physical characteristics:

- Width: 217 mm (8.6 inches)
- Depth: 711 mm (28.0 inches)
- Height: 168 mm (6.6 inches)
- Weight: 16.8 kg (37 lbs.)

Two I/O drawers in a 7311 rack-mounted enclosure have the following characteristics:

- Width: 445 mm (17.5 inches)
- Depth: 711 mm (28.0 inches)
- Height: 175 mm (6.9 inches)

[5] Central Electronics Complex (CEC)

- Weight: 39.1 kg (86 lbs.)

Figure 2-13 on page 28 shows the different views of the 7311-D10 I/O drawer.

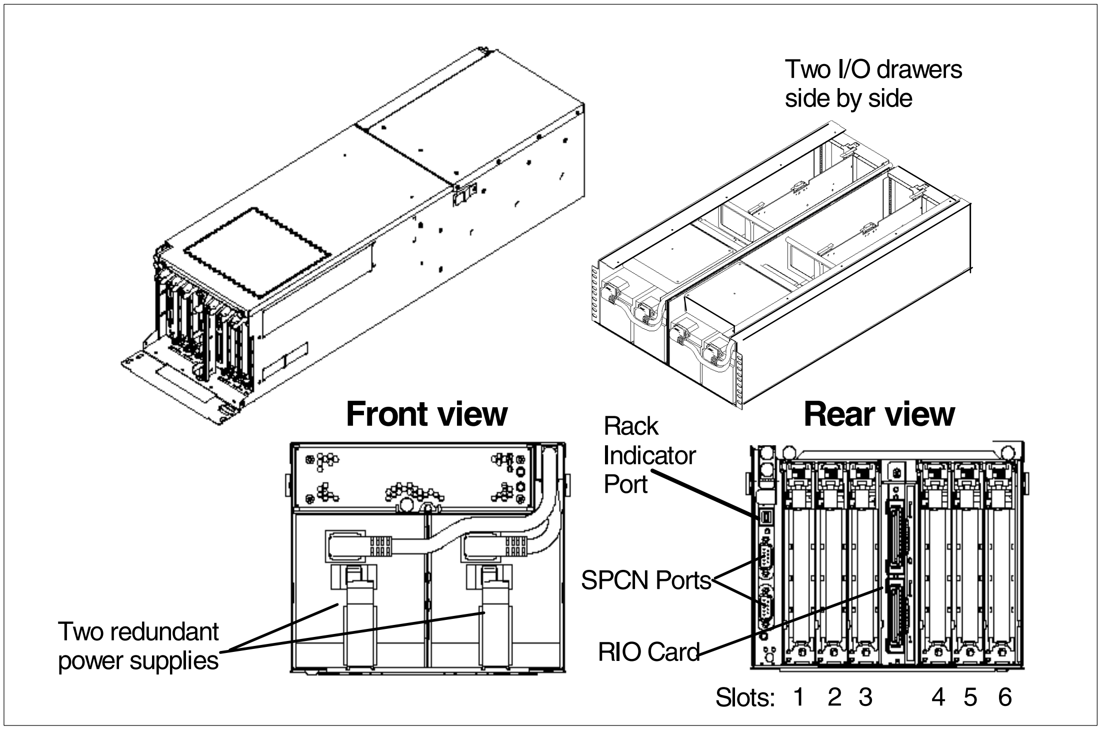


Figure 2-13 7311-D10 I/O drawer views

In the 7311-D10 I/O drawer, PCI-X and PCI/PCI-X adapters could be hot-plugged by using blind-swap cassettes: PCI/PCI-X blind-swap cassette kit, single wide, universal (FC 4599). The single-wide, universal cassette could be used for almost every adapter, except very wide adapters, where the double-wide adapter is needed.

Certain adapters require a retaining clip to be attached to the PCI blind swap carrier to ensure the adapter seat correctly in the I/O drawer. Blank PCI blind swap carriers that are factory-installed do not have the clip. If a blank carrier is used for an adapter that is transferred from a system other than p690 or p670, the clip will be needed then ask IBM services representative for assistance.

FC 4597 is available for the half-length slot in the p650 and not used in the I/O drawer.

2.5.2 7311-D20 I/O drawer

The 7311-D20 is a 4U full-size drawer, which must be mounted in a rack. It features seven hot-pluggable PCI-X slots and up to 12 hot-swappable disks arranged in two 6-packs. Redundant concurrently maintainable power and cooling is an optional feature (FC 6268).

The hot-plug mechanism is the same as on the Models 6F1, 6H1, or 6M1; therefore, PCI cards are inserted from the top of the I/O drawer down into the slot. The installed adapters are protected by plastic separators, designed to prevent grounding and damage when adding or removing adapters.

The 7311-D20 I/O drawer is connected with a Remote I/O (RIO) or Remote I/O 2 (RIO-2) loop to the CEC. Each RIO port can operates at 500 MHz (RIO) or 1 GHz (RIO-2) in bidirectional mode and is capable of passing up to eight bits of data in each direction on each cycle of the

RIO port. Therefore, the maximum data rate is 2 GB/s (RIO) or 4 GB/s (RIO-2) per I/O drawer in double barrel mode.

The p650 to achieve the higher Remote I/O performance of RIO-2 must have a RIO-2 Enabled System Planar (FC 9581 or FC 6581), all adapters (FC 6415 and FC 6416) in a RIO loop must be RIO-2 and 7311-D20 must have RIO-2 riser adapter (FC 6417).

The I/O drawer has the following physical characteristics:

- Width: 482 mm (19 inches)
- Depth: 610 mm (24.0 inches)
- Height: 178 mm (7 inches)
- Weight: 45.9 kg (101 lb.)

Figure 2-14 on page 29 shows the different views of the 7311-D20 I/O drawer.

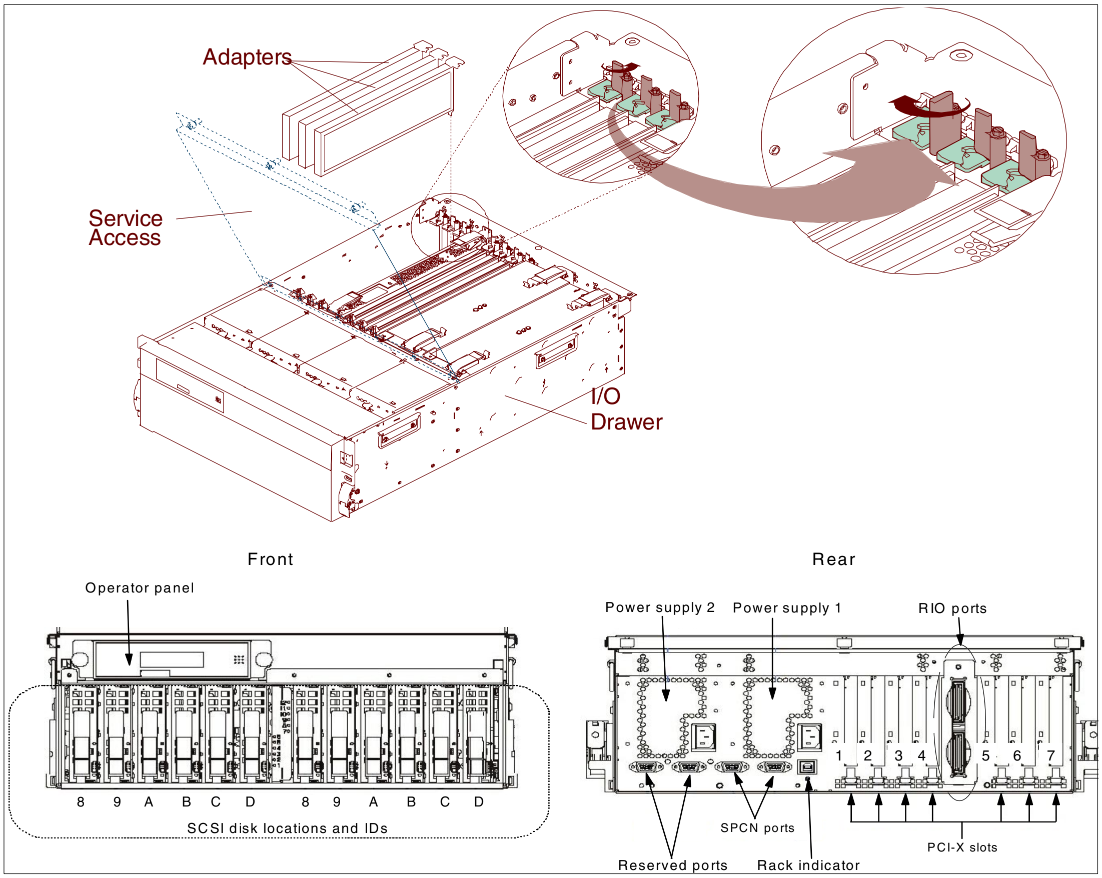


Figure 2-14 7311-D20 I/O drawer views

> **Note:** The 7311-D10 or 7311-D20 I/O drawer is designed to be installed by an IBM service representative.

2.5.3 I/O drawer ID assignment

A basic understanding of I/O drawer ID assignment is vital for locating devices and providing your service representative with the proper information.

Each I/O drawer belonging to the p650 is uniquely identified by an ID, such as U0.*x*, where *x* is the I/O drawer ID starting from 2, since the p650 is identified as U0.1.

The first ID assignment calls all the I/O drawers attached to the p650 according to the sequence they have in the SPCN loop. At this time, the system firmware creates a table in NVRAM reporting the I/O drawer ID assignment against the I/O drawer system serial number.

If you have a system A with three I/O drawers, the assigned addresses could be U0.2, U0.3, and U0.4. If you want to move the second I/O drawer, U0.3, away for any reason, the system maintains the names of the other I/O drawers, U0.2 and U0.4, and their related parts, without any change of the AIX physical locations related to the features inside those drawers. If you add a brand new I/O drawer or an I/O drawer from system B that was never used with system A, system A assigns to the I/O drawer the first address after the last configured I/O drawer, in our case U0.5 and not U0.3. If you move the removed I/O drawer, U0.3, back, the system checks its table as to whether the serial number is known. Since it is, the system recognizes it as U0.3 again.

2.5.4 PCI-X subsystem inside 7311-D10 I/O drawer

The PCI-X host bridge inside the I/O drawer provides two primary 64-bit PCI-X buses running at 133 MHz. Therefore, a maximum bandwidth of 1 GB/s is provided by each of the busses. To avoid overloading an I/O drawer, the recommendations in the *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538 should be followed.

Each primary PCI-X bus is connected to a PCI-X-to-PCI-X bridge, which provides three slots with Extended Error Handling (EEH) for error recovering. Slot 1 is a standard PCI slot, which operates at 33 MHz and 5 volt signaling. Slots 2 to 6 are PCI-X slots that operate at 133 MHz and 3.3 volt signaling. Figure 2-15 shows a conceptual diagram of the 7311-D10 I/O drawer.

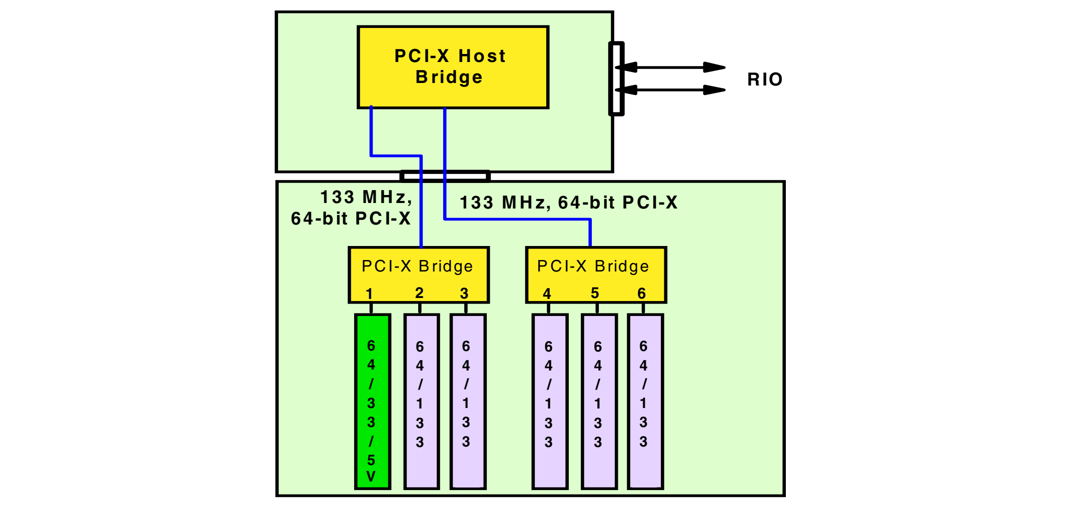


Figure 2-15 Conceptual diagram of the 7311-D10 I/O drawer

2.5.5 PCI-X subsystem inside 7311-D20 I/O drawer

The PCI-X host bridge inside the I/O drawer provides two primary 64-bit PCI-X buses running at 133 MHz. Therefore, a maximum bandwidth of 1 GB/s is provided by each of the busses. To avoid overloading an I/O drawer, the recommendations in the *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538 should be followed.

Each primary PCI-X bus is connected to a PCI-X-to-PCI-X bridge, which provides three or four slots with Extended Error Handling (EEH) for error recovering. Slots 1 to 7 are PCI-X slots that operate at 133 MHz and 3.3 volt signaling. Figure 2-16 on page 31 shows a conceptual diagram of the 7311-D20 I/O drawer.

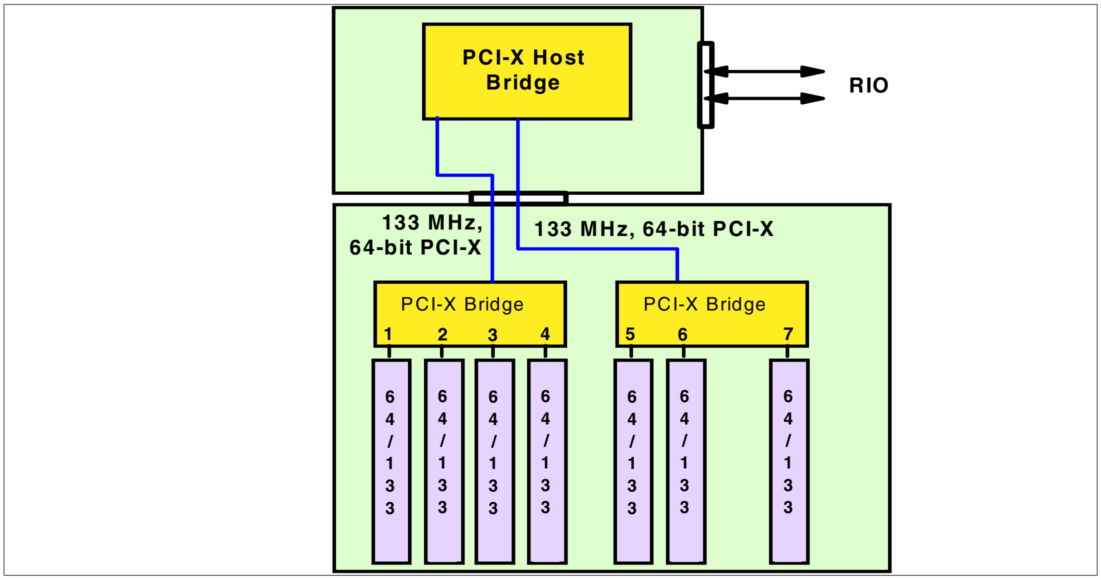


Figure 2-16   Conceptual diagram of the 7311-D20 I/O drawer

2.5.6 I/O drawer cabling

For power control and monitoring of the I/O drawers, SPCN cables are used. The SPCN cables form a loop similar to the way the RIO cables do, but there is no correlation between SPCN and RIO cabling required. Regardless how many I/O drawers are in the system, the SPCN system forms only one loop. The cabling starts from SPCN port 0 on the CEC to SPCN port 0 on the first I/O drawer connecting them from port 1 to port 0 of further I/O drawers or back to the CEC. For further information about SPCN refer to 3.3.6, “System Power Control Network (SPCN), power supplies, and cooling” on page 54.

The RIO cabling works similarly; however, there could be more loops for performance reasons. The CEC connects from a RIO port 0 to port 0 on the I/O drawer. From port 1 at the I/O drawer additional drawers can be connected. The last I/O drawer in a RIO loop connect from port 1 back to port 1 on the CEC. Figure 2-17 on page 32 shows the cabling of the RIO cables as well as the cabling of SPCN.

The RIO cables used in p650 are different from the cables used in former systems; therefore, they could not be exchanged with cables from other systems, such as p660 or p680. The cables used in the p630 are the same cables and are interchangeable. For further details about the RIO interface refer to 2.3.3, “RIO buses and GX cards” on page 20.

SPCN cables are the same as used in former systems. There are additional short RIO and SPCN cables available for connecting two I/O half-drawers. The following RIO and SPCN cables are available:

- Remote I/O cable, 1.2 M (FC 3146)
- Remote I/O cable, 3.5 M (FC 3147)
- Remote I/O cable, 10 M (FC 3148)
- SPCN cable, 2 M (FC 6001)

- SPCN cable, 3 M (FC 6006)
- SPCN cable, 6 M (FC 6008)
- SPCN cable, 15 M (FC 6007)

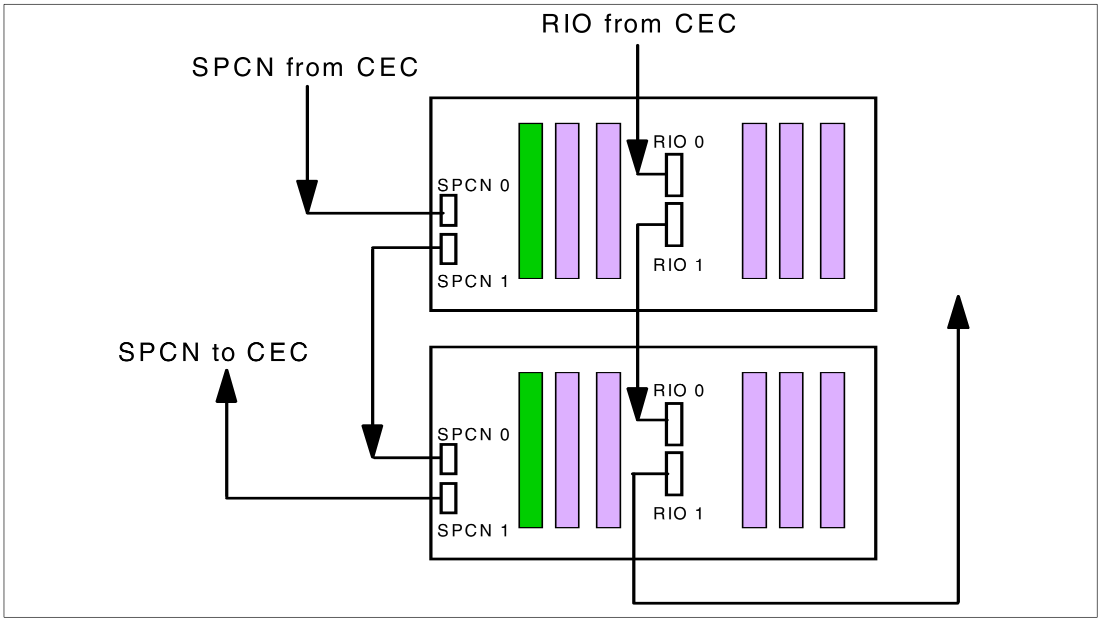


Figure 2-17 Remote I/O and SPCN cabling

2.6 External disk subsystems

The p650 can have up to four hot-swappable drives using 4-pack disk backplane or two independent pair of hot-swappable drives using split disk backplane in the front hot-swappable disk bays. Internal disks are usually used for the AIX rootvg and paging space. When the system is used in full system partition mode (no LPARs) these four disks should be enough. But when the system is used in LPAR mode, each LPAR needs its own disks for the rootvg boot device. The standard 4-disk backplane is connected to the internal SCSI port, therefore it could be used in only one LPAR and could not be shared between different LPARs. The optional split SCSI backplane allows two LPARs when connected to independent SCSI controllers (two SCSI cards are recommended for LPAR).

These additional disks cannot be added by using 7311-D10 I/O-drawer, because this does not have disk bays inside, but can be added by using 7311-D20 I/O drawer.

For expansion of a p650, there are different possibilities, which will be discussed in the following sections.

2.6.1 IBM 2104 Expandable Storage Plus

The IBM 2104 Expandable Storage Plus Model DU3 is a low-cost 3U disk subsystem that supports up to 14 Ultra3 SCSI disks from 18.2 GB up to 146.8 GB at the time this publication was written. This subsystem could be used in splitbus mode meaning the bus with 14 disks could be split into two busses with seven disks each. In this configuration two additional LPARs could be provided with up to seven disks for rootvg by using one Ultra3 SCSI adapter (FC 6203) for each LPAR.

For further information about the IBM 2104 Expandable Storage Plus subsystem, visit the following Web site:

http://www.storage.ibm.com/hardsoft/products/expplus/expplus.htm

2.6.2 IBM 7133 Serial Disk Subsystem (SSA)

The IBM 7133 Serial Disk Subsystem Model D40 provides a 4U highly available storage subsystem for pSeries servers and is a good solution for providing disks for booting additional LPARs. Disks are available from 18.2 GB up to 72.8 GB at the time this publication was written. SSA disk subsystems are connected to the server in loops. Each 7133 disk subsystem provides a maximum of four loops with a maximum of four disks each. Therefore up to four additional LPARs could be provided with disks with dedicated loops for booting by using one Advanced Serial RAID Plus adapter (FC 6230) for each LPAR. Disk space for booting could be provided in JBOD (Just a Bunch Of Disks) or RAID mode.

> **Notes:** FC 6230 serial RAID adapters provide boot support from a RAID configured disk with firmware level 7000 and above.
>
> Fastwrite cache must not be enabled on the boot resource SSA adapter.
>
> For more information on SSA boot, see the SSA frequently asked questions located on the Web:
>
> http://www.storage.ibm.com/hardsoft/products/ssa/faq.html#microcode

For further information about SSA, visit the following Web site:

http://www.storage.ibm.com/hardsoft/products/7133/7133.htm

2.6.3 IBM TotalStorage FAStT Storage servers

The IBM TotalStorage FAStT Storage server family consists of three models: Model 200, 500, and 700. The Model 200 is the smallest model which scales up to 4.8 TB, Model 700 is the largest which scales up to 16.3 TB at the time this publication was written. Model 200 and 500 provide up to 16 bootable partitions, which are attached with the Gigabit Fibre Channel adapter (FC 6239). Model 700 provides up to 64 bootable partitions. In most cases, both the FAStT storage server and p650 or the I/O drawer 7311-D10 are connected to a Storage Area Network (SAN).

If only space for the rootvg is needed, the FAStT Model 200 is a good solution. At the time this publication was written the smallest configuration with FAStT Model 200 is only 3U high and can provide up to 730 GB disk space without expansion enclosures. This space could be shared between up to 16 LPARs to store the rootvgs. Therefore one FAStT Model 200 is sufficient to provide disk space for booting for up to two p650 with eight LPARs each.

> **Note:** To boot p650 or any other pSeries server from a SAN using FC 6239, the adapter microcode 3.22A1 or later is required. Boot support is provided from direct attached FastT as well as from SAN attached FastT.

For support of additional features and for further information about the FAStT family refer to the following Web site:

http://www.storage.ibm.com/hardsoft/disk/fastt/index.html

2.6.4 IBM TotalStorage Enterprise Storage Server (ESS)

The IBM TotalStorage Enterprise Storage Server is the high-end storage server, providing from 420 GB up to 27.9 TB of disk space. An ESS system could also be used to provide disk space for booting LPARs of p650. An ESS Storage Server is usually connected to a Storage Area Network (SAN) to which p650 is also connected by using Gigabit Fibre Channel adapters (FC 6239).

For further information about ESS refer to the following Web site:

http://www.storage.ibm.com/hardsoft/products/ess/index.html

2.6.5 Additional boot options

The most common way to boot the p650 is from an internal disk or an external attached disk, which was discussed in 2.6, “External disk subsystems” on page 32. Additional possibilities for booting the p650, especially for installing the system or for system maintenance, are the following:

- CD-ROM or DVD-RAM

  These devices can be used to boot the system so that a system can be loaded, system maintenance performed, or stand-alone diagnostics performed.

- Internal or external tape drives

  The media bay tape drive or any externally attached tape drive can be used to boot the system using a **mksysb**, for example.

- LAN boot

  LAN adapters are used to boot from a system with Network Installation Manager (NIM). NIM simplifies installation and management of multiple servers and LPARs by putting the contents of the AIX installation medias and all the fixes to a central server (NIM server). The NIM server is then used to install and maintain other AIX systems over an IP network.

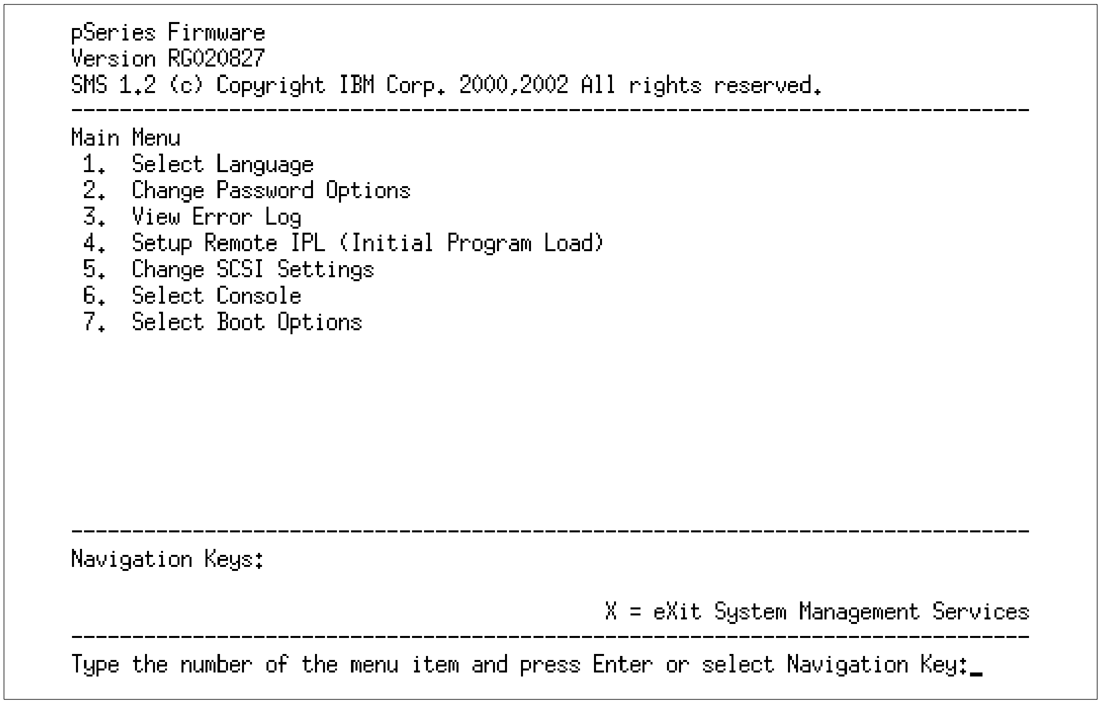
  At the time this publication was written, the following adapters are supported for a NIM boot:

  - FC 4962 10/100 Mbps Ethernet PCI Adapter II
  - FC 4961 IBM Universal 4-Port 10/100 Ethernet Adapter
  - FC 4959 IBM Token-Ring PCI Adapter
  - FC 5700 IBM Gigabit Ethernet-SX PCI-X Adapter
  - FC 5701 IBM 10/100/1000 Base-TX Ethernet PCI-X Adapter

2.7 System Management Services (SMS)

When you have the p650 with a graphics adapter connected to a graphics display plus a keyboard and mouse device, or an ASCII display terminal connected to one of the first two system serial ports, you can use the System Management Services menus to view information about the system and perform tasks such as setting a password, changing the boot list, and setting the network parameters. Graphical SMS is not supported in LPAR mode.

To start the System Management Services, press the 1 key on the terminal or in the graphics keyboard after the word `keyboard` appears and before the word `speaker` appears. After the text-based System Management Services start, the screen shown in Figure 2-18 displays.



```
pSeries Firmware
Version RG020827
SMS 1.2 (c) Copyright IBM Corp. 2000,2002 All rights reserved.
-------------------------------------------------------------------------------
Main Menu
 1.  Select Language
 2.  Change Password Options
 3.  View Error Log
 4.  Setup Remote IPL (Initial Program Load)
 5.  Change SCSI Settings
 6.  Select Console
 7.  Select Boot Options








-------------------------------------------------------------------------------
Navigation Keys:

                                            X = eXit System Management Services
-------------------------------------------------------------------------------
Type the number of the menu item and press Enter or select Navigation Key:_
```

Figure 2-18 System Management Services main menu

Note: The version of firmware currently installed in your system is displayed at the top of each screen. Processor and other device upgrades may require a specific version of firmware to be installed in your system.

On each menu screen, you are given the option of choosing a menu item and pressing Enter (if applicable), or selecting a navigation key. You can use the different options to set the password and protect our system setup, review the error log for any kind of error reported during the first phases of the booting steps, or set-up the network environment parameters if you want the system boots from a NIM server. For more information see *pSeries 650 Service Guide,* SA38-0612.

Use the Select Boot Options menu to view and set various options regarding the installation devices and boot devices:

1. Select Install or Boot a Device
2. Select Boot Devices
3. Multiboot Startup

Option 1 (Select Install or Boot a Device) allows you to select a device to boot from or install the operating system from. This selection is for the current boot only.

Option 2 (Select Boot Devices) allows you to set the boot list.

Option 3 (Multiboot Startup) toggles the multiboot startup flag, which controls whether the multiboot menu is invoked automatically on startup.

2.8 LPAR

LPAR stands for logical partitioning and is the ability to divide a physical server into *virtual* logical servers, each running in its own private copy of the operating system. The p650 can be divided into up to eight LPARs when enough resources are available.

Though it may not seem practical, running a machine with a single LPAR, compared to full system partition mode (non-LPAR), provides for a faster system restart because the hypervisor has already provided some initialization, testing, and building of device trees. In environments where restart time is critical, it we recommend that you test the single LPAR scenario to see if it meets the system recycle time objectives.

Depending on the software installed on the p650, dynamic LPAR may be available or unavailable:

Dynamic LPAR available With dynamic LPAR available, the resources can be exchanged between partitions without stopping and rebooting the affected partitions. Dynamic LPAR requires AIX 5L Version 5.2 for all affected partitions, and the HMC recovery software must be at Release 3 Version 1 (or higher). In partitions running AIX 5L Version 5.1 or Linux, the Dynamic Logical Partitioning menu is not available.

Dynamic LPAR unavailable Without dynamic LPAR, the resources in the partitions are static. Dynamic LPAR is unavailable for partitions running AIX 5L Version 5.1 or Linux. When you change or reconfigure your resource without dynamic LPAR, all the affected partitions must be stopped and rebooted in order to make resource changes effective.

A server can contain a mix of partitions that support dynamic LPAR along with those that do not.

Note: Rebooting a running partition only restarts the operating system and does not restart the LPAR. To restart an LPAR, the operating system should be shut down without reboot and afterwards restarted again.

2.8.1 Hardware Management Console (HMC)

When the p650 is partitioned, an IBM Hardware Management Console for pSeries (HMC) is necessary. Either a dedicated 7315-C01, 7315-C02, or an existing HMC from a p670 or p690 installation (FC 7316) can be used. If a p650 is only used in full system partition mode (no LPARs) outside a cluster, an HMC is not required. In this case, a p650 behaves as a non-partitionable pSeries model.

The HMC is a dedicated desktop workstation that provides a graphical user interface for configuring and operating pSeries servers functioning in either non-partitioned, LPAR, or clustered environments. It is configured with a set of hardware management applications for configuring and partitioning the server. One HMC is capable of controlling multiple pSeries servers. At the time this publication was written, a maximum of 16 non-clustered pSeries servers and a maximum of 64 LPARs are supported by one HMC.

The HMC provides two serial ports. One serial port should be used to attach a modem for the Service Agent (see 3.3.8, “Service Agent and Inventory Scout” on page 55, for details). The second port could be used to attach a server. If multiple servers should be attached to the HMC, additional serial ports are necessary. The ports could be provided by adding a maximum of two of the following features to the HMC:

- 8-port Async Adapter (FC 2943)
- 128-Port Async Controller (FC 2944)

Note: To ensure that the Async adapter is installed into HMC and not in the server, make sure that the adapter is configured as a feature of the HMC at the time of order.

The HMC is connected with special attachment cables to the HMC ports of the p650. Only one serial connection to a server is necessary despite the number of LPARs. The following cables are available:

- FC 8121 Attachment Cable, HMC to host, 15 meters
- FC 8120 Attachment Cable, HMC to host, 6 meters

With these cables, the maximum length from any server to the HMC is limited to 15 meters. To extend this distance, a number of possibilities are available:

- Another HMC could be used for remote access. This remote HMC must have a network connection to the HMC that is connected to the servers.
- AIX 5L Web-based System Manager Client could be used to connect to the HMC over the network or the Web-based System Manager PC client could be used, which runs on a Windows operating system-based or Linux operating system-based system.
- When a 128-Port Async Controller is used, customer supplied RS-422 cables (up to 330 meters) may connect to an existing RS422 RAN breakout box. The breakout box is connected to the HMC port on the server using an attachment cable (FC 8121, 8120). When the 15 meter cable is used, the maximum distance the HMC can be is 345 meters, providing the entire cable length can be used.To attach an RS422 cable to a RS232 RAN box, you must have some type of RS232 to RS422 interface converter on the line.

The HMC provides a set of functions that are necessary to manage LPAR configurations. These functions include:

- Creating and storing LPAR profiles, which define the processor, memory, and I/O resources allocated to an individual partition.
- Starting, stopping, and resetting a system partition.
- Booting a partition or system by selecting a profile.
- Displaying system and partition status. In a non-partitionable system, the LED codes are displayed in the operator panel. In a partitioned system, the operator panel shows the word `LPAR` instead of any partition LED codes. Therefore all LED codes for system partitions are displayed over the HMC.

- Virtual console for each partition or controlled system. With this feature, every LPAR can be accessed over the serial HMC connection to the server. This is a very good feature when the LPAR is not reachable across the network or a remote NIM installation should be performed.

The HMC also provides a service focal point for the systems it controls. It is connected to the service processor of the system using the dedicated serial link, and must be connected to each LPAR using an Ethernet LAN for Service Focal Point and to coordinate dynamic LPAR operations. The HMC provides tools for problem determination and service support, such as call-home and error log notification through an analog phone line.

2.8.2 LPAR minimum requirements

Each LPAR must have a set of resources available. The minimum resources that are needed are the following:

- At least one processor per partition.
- At least 256 MB of physical memory for additional partition.
- At least one disk to store the operating system (for AIX, the rootvg).
- At least one disk adapter or integrated adapter to access the disk.
- At least one LAN adapter per partition to connect to the HMC, as well as general network access.
- A partition must have an installation method, such as NIM and a means of running diagnostics, such as network diagnostics.

> **Note:** The minimum system memory required to run in LPAR mode with a single 256 MB partition is 1 GB.

2.8.3 Hardware guidelines for LPAR on p650

The p650 is capable of creating a maximum of eight LPARs when equipped with optional features. There are some limitations that should be considered when planning for LPARs, which will be discussed in the following.

Processor

There are no special considerations for processors. Each LPAR needs at least one processor.

Memory

Planning the memory for logical partitioning involves additional considerations to those discussed in 2.2.1, “Memory options” on page 17. These considerations are different when using AIX 5L Version 5.1, AIX 5L Version 5.2, or Linux.

When a machine is in full system partition mode (no LPARs) all of the memory is dedicated to AIX 5L. When a machine is in LPAR mode, some of the memory used by AIX is relocated outside the AIX-defined memory range. In the case of a single small partition (256 MB), the first 256 MB of memory will be allocated to the hypervisor; 256 MB is allocated to translation control entries (TCEs) and to hypervisor per partition page tables; and 256 MB for the first page table for the first partition. TCE memory is used to translate the I/O addresses to system memory addresses. Additional small page tables for additional small partitions will fit in the page table block. Therefore, the memory allocated independently of AIX to create a single 256 MB partition is 1 GB.

With the previous memory statements in mind, LPAR requires at least 2 GB of memory for two or more LPARs on a p650.

The following rules apply only for partitions with AIX 5L or Linux:

- The minimum memory for an LPAR is 256 MB. Additional memory can be configured in increments of 256 MB.
- The memory consumed outside AIX is from 0.75 GB up to 2 GB, depending on the amount of memory and the number of LPARs.
- For AIX 5L Version 5.1, the number of LPARs larger than 16 GB is limited to two in a system with 64 GB of installed memory, because of the memory alignment in AIX 5L Version 5.1.

  LPARs that are larger than 16 GB are aligned on a 16 GB boundary. Because the hypervisor memory resides on the lower end of the memory and TCE resides on the upper end of the memory, there are only two 16 GB boundaries available.

  The organization of the memory in a p650 must also be taken into account. Every processor card has its dedicated memory range. Processor card 1 has the range 0-16 GB, processor card 2 has the range 16-32 GB, processor card 3 32-48, and processor card 4 48-64 GB. If a processor card is not equipped with the maximum possible memory, there will be holes and the necessary 16 GB contiguous memory will not be present in the system. For example, in a system with three processor cards and 36 GB of memory, the memory is distributed into the ranges 0-12, 16-28, and 32-50. In this configuration, the only available 16 GB boundary (at 16 GB) has only 12 GB of memory, which is too small for a partition with more than 16 GB of memory and AIX 5L Version 5.1.
- According to the recommended memory configurations in Table 2-1 on page 17 and technical requirements only two possible configurations for partitions greater than 16 GB of memory running AIX 5L Version 5.1:
  - 6 processors with 48 GB permits only one partition greater than 16 GB of memory.
  - 8 processors with 64 GB permits two partitions greater than 16 GB of memory.
- With AIX 5L Version 5.2, there are no predefined limits concerning partitions larger than 16 GB, but the total amount of memory and hypervisor overhead remains a practical limit.

> **Note:** To create LPARs running AIX 5L Version 5.2 or Linux larger than 16 GB, the checkbox **Small Real Mode Address Region** must be checked (on the HMC, LPAR Profile, Memory Options dialog). Do not select this box if you are running AIX 5L Version 5.1.

I/O

The I/O devices are assigned on a slot level to the LPARs, meaning an adapter installed in a specific slot can only be assigned to one LPAR. If an adapter has multiple devices such as the 4-port Ethernet adapter or the Dual Ultra3 SCSI adapter, all devices are automatically assigned to one LPAR and cannot be shared.

The p650’s internal devices can also be assigned to LPARs, but in this case the internal connections must be taken into account. The 10/100 Mbps Ethernet port is connected directly to one of the PCI-X-to-PCI-X bridges; therefore, it can be independently assigned to a LPAR.

Without the split 4-pack SCSI backplane, the 4-pack internal disks, the media bays, and the external SCSI port are all driven by one SCSI chip on the I/O backplane. This chip is connected to one of the PCI-X-to-PCI-X bridges, which in terms of LPAR is equal to a slot. Therefore, in a standard configuration all SCSI resources in the disk and media bays, including external disks that are connected to the external SCSI port, must be assigned together to the same LPAR. There is no requirement to assign them to a particular LPAR, in fact they can remain unassigned if the LPAR minimum requirements are obtained using external devices attached to a SCSI adapter in the CEC or I/O drawer.

This can result in complications when an LPAR with the internal SCSI resources is active and a second LPAR needs to be installed using the internal media devices. In a standard configuration, this is not possible without removing them from the active LPAR that contains all the internal SCSI devices. In this scenario, when the second LPAR is installed using all the internal SCSI devices, you must be careful to not override the disks of the first LPAR.

To avoid this problem, the best solution for providing access to CD-ROMs and DVD-RAMs for different LPARs is probably to use an external attached DVD-RAM (FC 7210 Model 025) with a storage device enclosure (FC 7212 Model 102). This external DVD-RAM could be connected to a PCI SCSI adapter (FC 6203), which makes it easy to move the DVD-RAM between different LPARs. This solution also provides the advantages of sharing this DVD-RAM between several servers by attaching it to SCSI adapters in different servers.

Two independent pairs of hot-swappable drives are optionally available using split 4-pack SCSI backplane. This configuration is available for systems that require two logical partitions using under the covers DASD. When configured using a single SCSI card and the external SCSI port, the internal media bays will be grouped with the two drives controlled by the external SCSI port. To free the media devices for assignment to LPARS, it is recommended to use two SCSI adapters to support the split SCSI backplane. In this configuration, the media devices will be attached to the internal SCSI controller and be free to be assigned to any partition and not affect a configuration using the internal disks.

For additional LPARs, external disk space is necessary, which can be accomplished by using 7311-D20 I/O drawer or external disk subsystems as described in 2.6, “External disk subsystems” on page 32. The external disk space must be attached with a separate adapter for each LPAR by using SCSI, SSA, or Fibre Channel adapters, depending on the subsystem.

The four internal serial ports, diskette drive, keyboard, and mouse are connected to an ISA bus that is in the end connected to the RIO to PCI-X host bridge. Therefore these ports and the diskette drive could only be assigned together to one LPAR, but these resources are independent from the SCSI resources.

The number of RIO or RIO-2 cards installed has no affect on the number of LPARs supported other than the limitations related to the total number of I/O drawers supported, and the ability to meet the LPAR minimum requirements in a particular configuration.

There are limits to dynamic LPAR. The ISA I/O resources can neither be added nor removed using dynamic LPAR, including any devices sharing the same PCI-X bridge, such as serial ports, native keyboard and mouse ports, and the diskette drive. Not all resources can be removed using dynamic LPAR; for example, you cannot go below the minimum configuration for processors, memory, or I/O (for example, removing a resource such as the rootvg, paging disks, or other critical resources).

2.9 Operating system requirements

The p650 is capable of running IBM AIX 5L for POWER and supports appropriate versions of Linux. AIX 5L has been specifically developed and enhanced to exploit and support the extensive RAS features on IBM @server pSeries systems, and AIX 5L Version 5.2 supports dynamic logical partitioning.

2.9.1 AIX 5L

The p650 requires AIX 5L Version 5.2 or AIX 5L Version 5.1 at Maintenance Level 3 (APAR IY32749) or AIX 5L Version 5.2 initial CD-set and a license of AIX 5L Version 5.1.

In order to boot from the CD, make sure that one of the following media is available:

- AIX 5L Version 5.1 5765-E61, dated 10/2002 (CD# LCD4-1061-04) and update CD dated 12/2002 (CD# LCD4-1103-06) or later.
- AIX 5L Version 5.2 5765-E62, initial CD-set (CD# LCD4-1133-00) or later.

Note: Ensure that the appropriate levels of CDs, as previously described, are ordered. The p650 will not be able to boot from earlier levels of AIX install CDs.

IBM periodically releases maintenance packages for the AIX 5L operating system. These packages are available on CD-ROM (FC 0907) or they can be downloaded from the Internet at:

`http://techsupport.services.ibm.com/server/fixes`

You can also get individual operating system fixes and information on how to obtain AIX 5L service at this site. If you have problems downloading the latest maintenance level, ask your IBM Business Partner or IBM representative for assistance.

Tip: To check the current AIX level, enter the **`oslevel -r`** command. The output for the AIX 5L Version 5.1 minimum maintenance level is `5100-03`.

AIX 5L application binary compatibility

IBM AIX 5L Version 5.1 preserves binary compatibility for 32-bit application binaries from previous levels of AIX Version 4 and AIX 5L, and for 64-bit applications compiled on previous levels of AIX 5L. 64-bit applications compiled on Version 4 must be recompiled to run on AIX 5L.

2.9.2 Linux

SuSE Linux Enterprise Server 8 with Service Pack 1 supports the p650 in Full System Partition (non-LPAR), and LPAR mode. For more information, see:

`http://www.suse.com/us/business/products/server/sles/i_pseries.html`

Linux can be ordered through IBM when an order for a Linux ready Express Configuration is placed. 5639-LNX is the feature code.

IBM expects other Linux distributions to support the p650 later in 2003.

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

`http://www.ibm.com/servers/eserver/pseries/linux/whitepapers/linux_pseries.html`

For all the latest in IBM Linux news, see:

`http://www-1.ibm.com/servers/eserver/pseries/linux/`

3

Availability, investment protection, expansion, and accessibility

This chapter provides more detailed information about configurations, upgrades, and design features that will help lower the total cost of ownership.

3.1 Capacity on Demand

The pSeries 650 systems can be shipped with non-activated resources (processors and/or memory), which may be purchased and activated at a certain point in time without affecting normal machine operation.

The following sections outline the methods available, namely:

- Capacity Upgrade on Demand
- Memory Capacity Upgrade on Demand
- On/off Capacity on Demand
- Trial Capacity Upgrade on Demand

3.1.1 Capacity Upgrade on Demand

Capacity Upgrade on Demand (CUoD) is a method where up to six CUoD 1.45 GHz processors can be installed and later activated in increments of two processors

The following basic rules apply for a pSeries 650 system with the CUoD processor configuration:

- Up to six CUoD 1.45 GHz processors can be installed in increments of 2-way processors (FC 7014).
- Processors are activated in increments of 2-way processors and can be permanently activated (FC 7011) or temporarily (FC 7013), each 7011 or 7013 activates 2-way processors of one 7014.
- The 1.2 GHz processors are not available for CUoD.
- Processor Capacity Upgrade On Demand is supported by AIX 5L Version 5.1 or AIX 5L Version 5.2.
- CUoD processors which have not been activated are available to p650 server for dynamic processor sparing when running AIX 5l Version 5.2, see Section 3.1.5, “Dynamic Processor Sparing” on page 46.
- Dynamic processor sparing is limited to systems and/or partitions which are running AIX 5.2. Systems running only AIX 5L Version 5.1 provide processor sparing for a failed processor on a system IPL.

3.1.2 Memory Capacity Upgrade on Demand

Memory Capacity Upgrade on Demand is a method where up to 56 GB of CUoD memory can be installed and later activated in increments of 4 GB using encrypted keys.

Two new feature codes are required for this feature:

- FC 7057 is a 4 GB memory feature with 0 GB active
- FC 7058 is a 8 GB memory feature with 0 GB active

The following basic rules apply for a pSeries 650 system with the CUoD memory configuration:

- Up to 56 GB of CUoD memory can be installed and it is provided by CUoD DIMM features FC 7057 (4096 MB) and FC 7058 (8192 MB). Each consists of a set of four DIMMs, with none of the memory activated.

- CUoD memory is activated in increments of 4 GB processors and can be permanently activated ordering:
  - FC 7052 activates all 4096 MB of memory on feature 7057 (4096 MB).
  - FC 7053 activates 4096 MB of memory on feature 7058 (8192 MB).
  - Two FC 7053 activate all 8192 MB of memory on feature 7058 (8192 MB).
- CUoD memory features are supported only on 1.45 GHz processor cards. They can not be used with 1.2 GHz processor cards.
- Memory Capacity Upgrade On Demand requires AIX 5L Version 5.2.

3.1.3 On/off Capacity on Demand

Normally, once CUoD features are activated, they are applied to the system for the remainder of the system life. With On/off Capacity on Demand, groups of two CUoD processors can be activated and their usage, based on an hourly granularity, is deducted from a 30-day activation increment.

After the 30 day usage, the processors are moved back to inactive CUoD status upon a reboot or LPAR reconfiguration.

On/Off CoD 30-Day Two Processor Activations (FC 7013) feature provides thirty days of usage for two processors on CUoD (Capacity Upgrade on Demand) processor feature (FC 7014).

Any quantity of On/Off CoD 30-Day Two Processor Activations (FC 7013) can be ordered and used as needed. Usage is continuously monitored and the remaining time available is reported in increments of one hour. Processors can be turned off when they are not in use, to preserve the remaining time available.

Note: On/Off CoD 30-Day Two Processor Activation for Processor (FC 7013) will be available on September, 2003.

3.1.4 Trial Capacity on Demand

In the past, customers with CUoD featured systems must purchase the activation codes from IBM before the non-activated CUoD resources can be activated to meet the increased workload. With this feature, customers can now activate the required non-activated CUoD resource immediately and after that, proceed to purchase those resources from IBM.

When viewed on the HMC menus, this feature is named Activate Immediate.

A one-time no cost activation for a maximum period of 30 consecutive days is available as a complementary service when access to CUoD resources are required immediately. It is useful when a CUoD permanent activation purchase is pending.

The following basic rules apply for a pSeries 650 system:

- After the CUoD resources are activated, the customer must either buy part or all of the activated CUoD resources from IBM or return the activated CUoD resources back to the system within 30 days.
- Trial CoD can only be used once:
  - When the system first boot up or

  - After the customer has purchased the activated CUoD resources which were activated through trial function previously.

There are several advantages of using this feature:

- Improve the response time to meet unpredictable increase in workload.
- Customer can monitor the performance of the system after activating the CUoD resources before placing the order for activation codes.

3.1.5 Dynamic Processor Sparing

When you have a CUoD system (that is, a system with at least one CUoD 2-way processor card) a feature called Dynamic Processor Sparing is automatically provided with it. Dynamic Processor Sparing is the capability of the system to disable a failing processor and enable a non-activated CUoD processor if a standby processor is available. Non-activated CUoD processors are processors that are physically installed in the system, but not activated. They can not run jobs or tasks unless they are become activated via CUoD processor activation code.

When Dynamic Processor sparing is performed, the system will be out of CUoD compliance for a period of time between when the spare processor is allocated and the failed processor is deallocated. By design, AIX must request the spare processor, and in doing so AIX is committing to returning the failing processor in a timely fashion.

> **Note:** Dynamic Processor Sparing requires AIX 5L version 5.2 or higher, and the system must be run in a partitioned environment. The Dynamic Processor Sparing requires that the CPU guard attribute is set to enable, see Section 3.3.4, “Dynamic or persistent deallocation” on page 52

3.2 Autonomic computing

The IBM autonomic computing initiative is about using technology to manage technology. This initiative is an ongoing effort to create servers that respond to unexpected capacity demands and system glitches without human intervention. The goal: New highs in reliability, availability, and serviceability, and new lows in downtime and cost of ownership.

Today's pSeries offers some of the most advanced self-management features for UNIX servers on the market today.

Autonomic computing on IBM @server pSeries servers[6] describes the many self-configuring, self-healing, self-optimizing, and self-protecting features that are available on IBM @server pSeries servers.

Self-configuring

Autonomic computing provides self-configuration capabilities for the IT infrastructure. Today IBM systems are designed to provide this at a feature level with capabilities like plug and play devices, and configuration setup wizards. Examples include:

- Virtual IP address (VIPA)
- IP multipath routing
- Microcode discovery services/inventory scout
- Hot-swappable disks

[6] http://www-3.ibm.com/autonomic/index.shtml

- Hot-swap PCI
- Wireless/pervasive system configuration
- TCP explicit congestion notification

Self-healing

For a system to be self-healing, it must be able to recover from a failing component by first detecting and isolating the failed component, taking it off-line, fixing or isolating the failed component, and reintroducing the fixed or replacement component into service without any application disruption. Examples include:

- Multiple default gateways
- Automatic system hang recovery
- Automatic dump analysis and e-mail forwarding
- Ether channel automatic failover
- Graceful processor failure detection and failover
- HACMP and HAGeo
- First failure data capture
- Chipkill ECC Memory, dynamic bit steering
- Memory scrubbing
- Automatic, dynamic deallocation (processors, LPAR, PCI buses/slots)
- Electronic Service Agent - *call-home* support

Self-optimization

Self-optimization requires a system to efficiently maximize resource utilization to meet the end-user needs with no human intervention required. Examples include:

- Static LPAR
- Dynamic LPAR
- Workload Manager enhancement
- Extended memory allocator
- Reliable, scalable cluster technology (RSCT)
- PSSP cluster management and Cluster Systems Management (CSM)

Self-protecting

Self-protecting systems provide the ability to define and manage the access from users to all the resources within the enterprise, protect against unauthorized resource access, detect intrusions and report these activities as they occur, and provide backup/recovery capabilities that are as secure as the original resource management systems. Examples include:

- Kerberos Version 5 Authentication (authenticates requests for service in a network)
- Self-protecting kernel
- LDAP directory integration (LDAP aids in the location of network resources)
- SSL (manages Internet transmission security)
- Digital certificates
- Encryption (prevents unauthorized use of data)

3.3 Reliability, availability, and serviceability (RAS) features

Excellent quality and reliability are inherent in all aspects of the p650 design and manufacture, and the fundamental principle of the design approach is to minimize outages. The RAS features help to ensure that the systems operate when required, perform reliably, and efficiently handle any failures that may occur. This is achieved using capabilities provided by both the hardware and the AIX 5L operating system.

Mainframe-class diagnostic capability based on internal error checkers, First-Failure Data Capture (FFDC), and run-time analysis is provided. This monitoring of all internal error check states is provided for processor, memory, I/O, power, and cooling components, and is aimed at eliminating the need to try to recreate failures later for diagnostic purposes. The unique IBM RAS capabilities are important for the availability of your server.

The following features provide the p650 with UNIX industry-leading RAS:

- Fault avoidance through highly reliable component selection, component minimization, and error handling technology designed into the chips.
- Improved reliability through processor operation at a lower voltage enabled by the use of copper chip circuitry and Silicon-on-Insulator.
- Fault tolerance with redundancy, dual line cords, and concurrent maintenance for power and cooling (using standard redundant hot-swap power supplies and fans).
- Automatic First Failure Data Capture (FFDC) and diagnostic fault isolation capabilities.
- Concurrent run-time diagnostics based on FFDC.
- Predictive failure analysis on processors, caches, memory, and disk drives.
- Dynamic error recovery.
- Service processor and SPCN for constantly monitoring the system.
- Error Checking and Correction (ECC) or equivalent protection (such as refetch) on main storage, all cache levels (1, 2, and 3), and internal processor arrays.
- Dynamic processor deallocation based on run-time errors.
- Persistent processor deallocation (processor boot-time deallocation based on run-time errors).
- Persistent deallocation extended to memory.
- Chipkill correction in memory.
- Memory scrubbing and redundant bit-steering for self-healing.
- Industry-leading PCI/PCI-X bus parity error recovery.
- Hot-plug functionality of the PCI-X bus I/O subsystem.
- PCI/PCI-X bus and slot deallocation.
- Disk drive fault tracking.
- Avoiding checkstops with process error containment.
- Environmental monitoring (temperature and power supply).
- Auto-reboot.
- Disk mirroring (RAID 1) and disk controller duplexing capability are provided by the AIX operating system.

Some of the RAS features of the p650 are covered in more detail in the following sections.

3.3.1 Service processor

The converged service processor is a specialized device for managing the system that is situated on the CSP backplane (see Figure 2-10 on page 23). The CSP provides a non-graphic user interface by attaching a supported ASCII terminal to serial port 1 and serial port 2. The p650 implements the same Converged Service Processor implementation as previously implemented in the previous system platforms. Also the System Power Control Network (SPCN) function is not implemented in a separate controller, but instead integrated with the CSP microprocessor functions.

When the system is powered down but still plugged into an active AC power source, the service processor and SPCN functions are still active under standby power. This means that all service processor menu functions (using a local, remote, or terminal concentrator console), as well as dial-out capability, are available even if the system is powered down or unable to power up. While the system is powered off, the service processor is in an idle state waiting for either a power on command or a keystroke from any of the TTYs attached to either serial port 1 or serial port 2. Depressing the Enter key or space bar will cause the service processor to display the service processor main menu.

Immediately after power on, the SPCN controls the powering up of all devices needed during the boot process. When the SPCN has completed its tasks, the CSP, using its onboard processor, checks for processor and memory resources and then tests them. After the processor and memory tests have completed, the service processor then hands the rest of the boot process over to system firmware. This changeover occurs when, on the system operator panel, the 9*xxx* LED codes become E*xxx* codes.

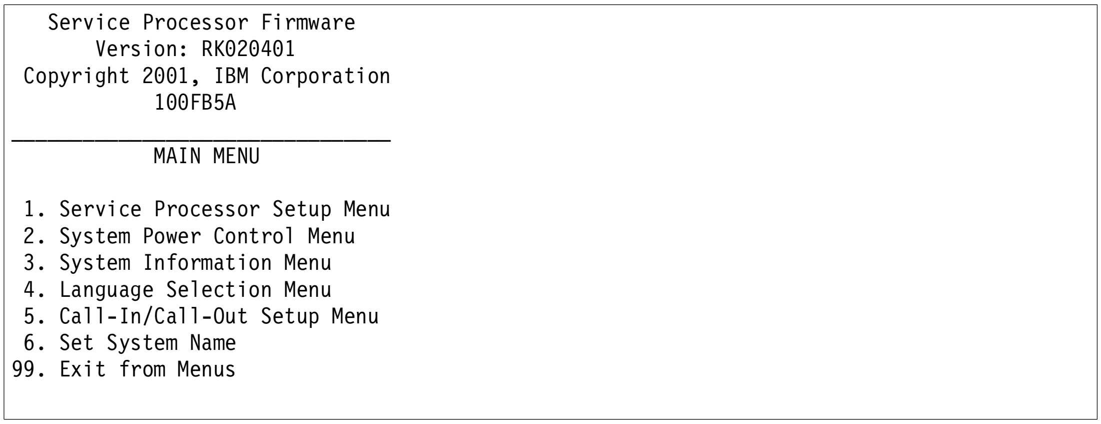
If you do not invoke the SMS menu, once system firmware has completed the test tasks, it reads the current boot list. If there is a valid boot device with AIX, firmware passes the control of the system to the operating system. When AIX controls the machine, the service processor is still working and checking the system for errors. Also, the surveillance function of the service processor is monitoring AIX to check that it is still running and has not stalled.

The CSP design for the p650 includes the ability to reset the service processor. This enables the system firmware to force a hard reset of the service processor if it detects a loss of communication. As this would typically occur while the system is already up and running, the service processor reset will be accomplished without impacting system operation.

When the system does not boot, a good source of error log information to perform a problem root cause or to involve the IBM technical support, is the history log maintained by the service processor, which keeps a limited history of error and information events, including any error information about the last failed boot.

Service processor main menu

The service processor menu and subsequent menus are only visible on an ASCII screen attached to the native serial ports. Figure 3-1 shows the service processor main menu.



```
    Service Processor Firmware
        Version: RK020401
  Copyright 2001, IBM Corporation
           100FB5A
________________________________
           MAIN MENU

 1. Service Processor Setup Menu
 2. System Power Control Menu
 3. System Information Menu
 4. Language Selection Menu
 5. Call-In/Call-Out Setup Menu
 6. Set System Name
99. Exit from Menus
```

Figure 3-1 Service processor main menu

Useful information contained within this menu is defined as follows:

RK020401 Service processor firmware level and date code. The RK020401 firmware indicated in this example is a preproduction level. Generally available versions will reflect a later date code.

100FB5A Machine serial number.

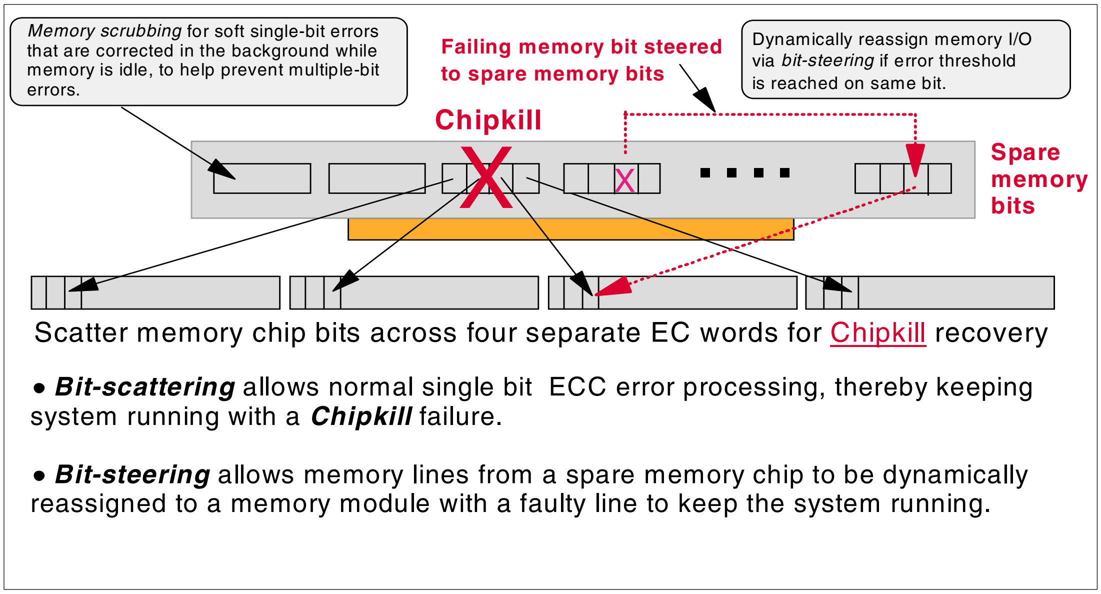
Service Processor Setup menu These menus allow you to set passwords to provide added security to your system and update machine firmware.

System Power Control menu These menus allow you to control some aspects of how the machine powers on and how much of the boot process you wish to complete (that is, you may decide to stop the current boot into system management services, SMS, menu).

System Information menu From this menu option, information about the previous boot sequence and errors produced can be viewed. Additionally, this menu provides a means of checking and altering the availability of processors and memory.

3.3.2 Memory reliability, fault tolerance, and integrity

The p650 uses Error Checking and Correcting (ECC) circuitry for system memory to correct single-bit and to detect double-bit memory failures. Detection of double-bit memory failures helps maintain data integrity. Furthermore, the memory chips are organized such that the failure of any specific memory module only affects a single bit within a four bit ECC word (*bit-scattering*), thus allowing for error correction and continued operation in the presence of a complete chip failure (*Chipkill recovery*). The memory DIMMs also utilize *memory scrubbing* and thresholding to determine when spare memory modules within each bank of memory should be used to replace ones that have exceeded their threshold of error counts (*dynamic bit-steering*). Memory scrubbing is the process of reading the contents of the memory during an idle time and checking and correcting any single-bit errors that have accumulated by passing the data through the ECC logic. This function is a hardware function on the memory controller chip and does not influence normal system memory performance.


Memory scrubbing for soft single-bit errors that are corrected in the background while memory is idle, to help prevent multiple-bit errors.
Failing memory bit steered to spare memory bits
Dynamically reassign memory I/O via *bit-steering* if error threshold is reached on same bit.
Chipkill
X
X
. . . .
Spare memory bits

Scatter memory chip bits across four separate EC words for Chipkill recovery

- ***Bit-scattering*** allows normal single bit ECC error processing, thereby keeping system running with a ***Chipkill*** failure.
- ***Bit-steering*** allows memory lines from a spare memory chip to be dynamically reassigned to a memory module with a faulty line to keep the system running.

Figure 3-2 Main storage ECC and extensions

3.3.3 First Failure Data Capture, diagnostics, and recovery

If a problem should occur, the ability to correctly diagnose it is a fundamental requirement upon which improved availability is based. The p650 incorporates un-matched capability in start-up diagnostics (see 3.3.1, “Service processor” on page 49) and in run-time FFDC based on strategic error checkers built into the chips.

Any errors detected by the pervasive error checkers are captured into Fault Isolation Registers (FIRs), which can be interrogated by the service processor. The service processor in the p650 has the capability to access system components using special purpose service processor ports or by access to the error registers (see Figure 3-3 on page 51).

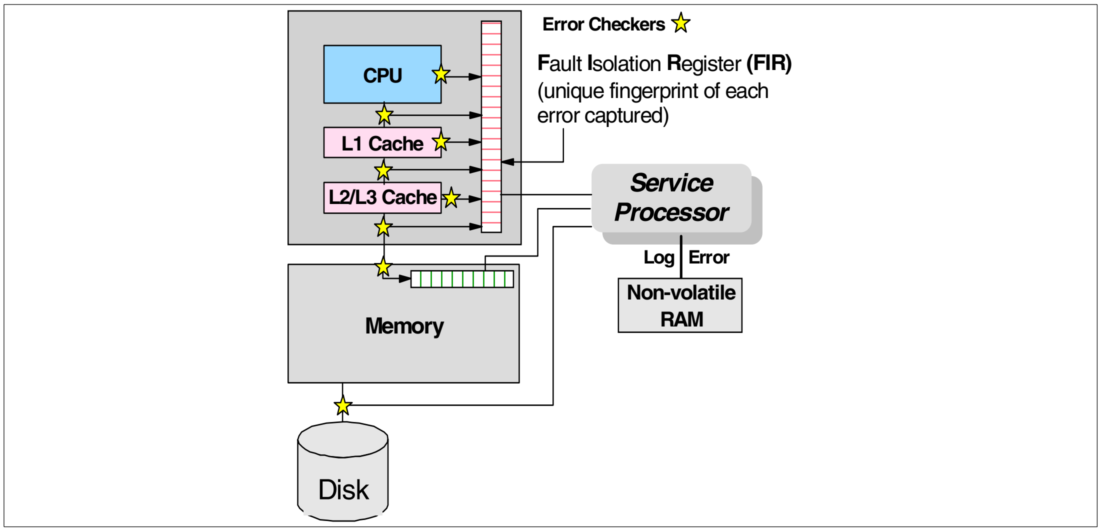


Figure 3-3 Fault Isolation Register

The FIRs are important because they enable an error to be uniquely identified, thus enabling the appropriate action to be taken. Appropriate actions might include such things as a bus retry, ECC correction, or system firmware recovery routines. Recovery routines could include dynamic deallocation of potentially failing components such as a processor or L2 cache.

Errors are logged into the system non-volatile random access memory (NVRAM) and the service processor event history log, along with a notification of the event to AIX for capture in the operating system error log. Diagnostic Error Log Analysis (diagela) routines analyze the error log entries and invoke a suitable action such as issuing a warning message. If the error can be recovered, or after suitable maintenance, the service processor resets the FIRs so that they can accurately record any future errors.

The ability to correctly diagnose any pending or firm errors is a key requirement before any dynamic or persistent component deallocation or any other reconfiguration can take place.

Also, the p650 provides processor cards, DIMMs, PCI/PCI-X adapters, and disk LED indicators to assist service people. These LEDs are a visual help for FRU identification. They are not being turned on by the system when a failure occurs, but system administrator or service people can identify the failing FRU in the AIX error log entry by looking for the AIX physical location or by turning on the related LED from online diagnostics or from a SMIT menu.

3.3.4 Dynamic or persistent deallocation

Dynamic deallocation of potentially failing components is non-disruptive, allowing the system to continue to run. Persistent deallocation occurs when a failed component is detected and is then deactivated at subsequent boot time.

Dynamic deallocation functions include:

- Processor
- L3 cache line delete
- PCI/PCI-X bus and slots

For dynamic processor deallocation, the service processor performs a predictive failure analysis based on any recoverable processor errors that have been recorded. If these transient errors exceed a defined threshold, the event is logged and the processor is deallocated from the system while the operating system continues to run. This feature (named *cpuguard*) enables maintenance to be deferred to suitable time. Processor deallocation can only occur if there are sufficient functional processors.

To verify whether cpuguard has been enabled, run the following command:

```
lsattr -El sys0 | grep cpuguard
```

If enabled, the output will be similar to the following:

```
cpuguard    enable      CPU Guard     True
```

If the output shows cpuguard as disabled, enter the following command to enable it:

```
chdev -l sys0 -a cpuguard='enable'
```

Note: The use of cpuguard is only effective on systems with three or more functional processors on AIX 5L Version 5.1, or two or more with AIX 5L Version 5.2.

Cache or cache-line deallocation is aimed at performing dynamic reconfiguration to bypass potentially failing components. The p650 provides this feature for both L2 and L3 caches. The L1 data cache and L2 data and directory caches can provide dynamic detection and correction of hard or soft array cell failures. Dynamic run-time deconfiguration is provided if a threshold of L1 or L2 recovered errors is exceeded.

In the case of a L3 cache run-time array single-bit solid error, the spare chip resources are used to perform a line delete on the failing line.

System bus recovery (retry) is provided for any address or data parity errors on the GX bus or for any address parity errors on the fabric bus.

PCI/PCI-X hot-plug slot fault tracking helps prevent slot errors from causing a system machine check interrupt and subsequent reboot. This provides superior fault isolation and the error affects only the single adapter. Run-time errors on the PCI/PCI-X bus caused by failing adapters will result in recovery action. If this is unsuccessful, the PCI/PCI-X device will be gracefully shut down. Parity errors on the PCI/PCI-X bus itself will result in bus retry and, if uncorrected, the bus and any I/O adapters or devices on that bus will be deconfigured.

The Model 6M2 supports PCI/PCI-X Extended Error Handling (EEH) if it is supported by the PCI/PCI-X adapter. In the past, PCI bus parity errors caused a global machine check interrupt, which eventually required a system reboot in order to continue. In the p650 system, new hardware, system firmware, and AIX interaction has been designed to allow transparent recovery of intermittent PCI/PCI-X bus parity errors, and graceful transition to the I/O device available state in the case of a permanent parity error in the PCI/PCI-X bus.

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

Persistent deallocation functions include:

- Processor
- Memory
- Deconfigure or bypass failing I/O adapters

Following a hardware error that has been flagged by the service processor, the subsequent reboot of the system will invoke extended diagnostics. If a processor or L3 cache has been marked for deconfiguration by persistent processor deallocation, the boot process will attempt to proceed to completion with the faulty device automatically deconfigured. Failing I/O adapters will be deconfigured or bypassed during the boot process.

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

3.3.5 UE-Gard

The UE-Gard (Uncorrectable Error-Gard) is a RAS feature that enables AIX 5L Version 5.2, in conjunction with hardware and firmware support, to isolate certain errors that would previously have resulted in a condition where the system had to be stopped (checkstop condition). The isolated error is being analyzed to determine if AIX can terminate the process that suffers the hardware data error instead of terminating the entire system.

UE-Gard is not to be confused with (dynamic) CPU Guard. CPU Guard takes a CPU dynamically offline after a threshold of recoverable errors is exceeded, to avoid system outages.

On memory errors the firmware will analyze the severity and record it in a RTAS[7] log.

AIX will be called from firmware with a pointer to the log. AIX will analyze the log to determine whether the error is recoverable. If the error is recoverable then AIX will resume. If the error is not fully recoverable then AIX will determine whether the process with the error is critical. If the process is not critical then it will be terminated by issuing a SIGBUS signal with an UE siginfo indicator. In the case where the process is a critical process, then the system will be terminated as a machine check problem.

3.3.6 System Power Control Network (SPCN), power supplies, and cooling

Environmental monitoring related to power, fan operation, and temperature is performed by the System Power Control Network. Critical power events, such as a power loss, trigger appropriate signals from hardware to impacted components to assist in the prevention of data loss without operating system intervention. Non-critical environmental events are logged and reported to the operating system.

A SYSTEM_HALT warning is issued to the operating system if the sensor chips in various parts of the hardware detect that the air temperature rises above a preset maximum limit, or if two or more system fan units are running slowly or have stopped. This warning will result in an immediate system shutdown action.

Hot-plug power supplies

The p650 provides redundancy as the standard option. With redundant power supplies, the failed power supply can be replaced concurrently and with minimum disruption. Note that when ambient temperatures exceed 32 C (92 F), it is advisable that you shut down the machine prior to replacing the faulty power supply.

Hot-plug cooling fans

The p650 has four fans that can be changed concurrently. To assist in the identification of a failing fan, each unit is equipped with an amber LED that will illuminate when a fault is detected. Two fans (fan #3 and fan #4) are located in the front and provide cooling for the processor, memory, and I/O subsystem; the other two fan units (fan #1 and fan #2) are located in the rear side as part of the power supplies, below the I/O and system planars, and provide cooling for disks and the media subsystem. Each fan can provide redundancy in the event of a single fan failure (the remaining processor cooling fan will increase speed when required). Any failed fan can be replaced concurrently, therefore eliminating the need for system downtime.

3.3.7 Early Power-Off Warning (EPOW)

Both critical and not critical power supply and fan failures generate a signal that SPCN reports to AIX through the service processor interface as an EPOW error message.

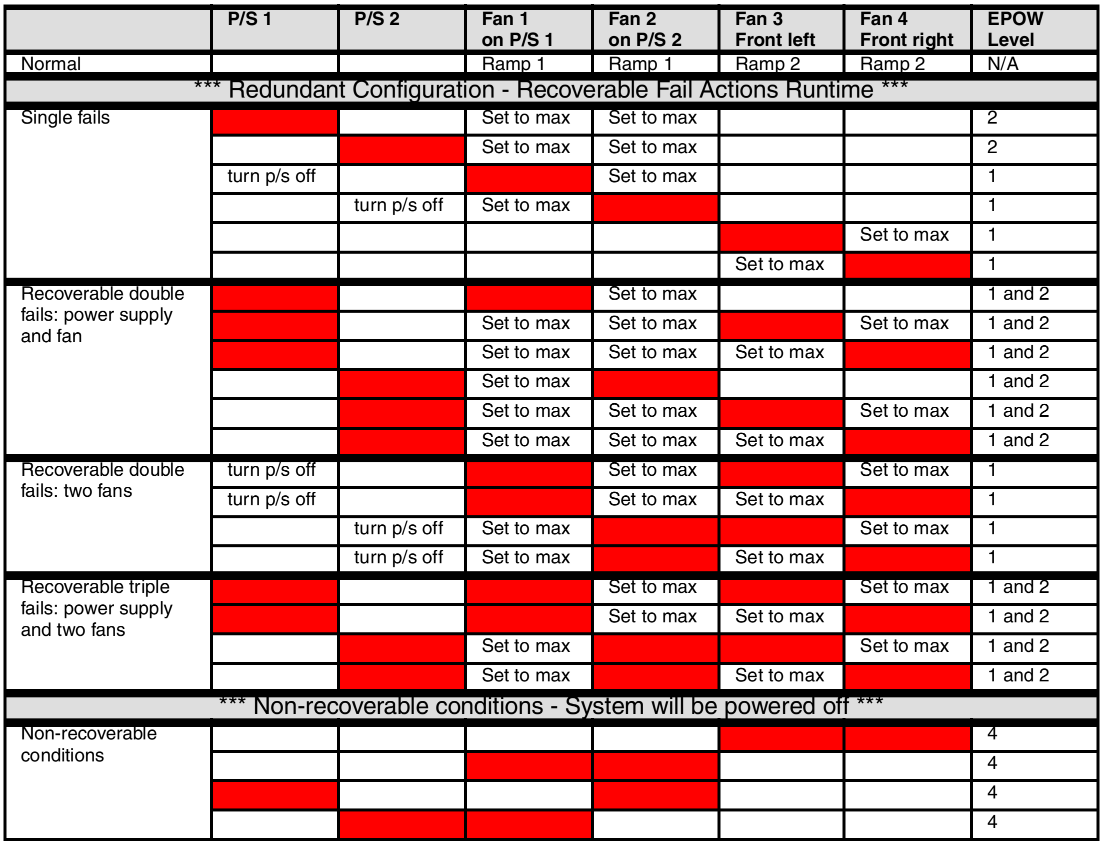
The following is a summary of the p650 EPOW functions:

- EPOW 1 - Warn Cooling: This type of fault occurs when one of the system fans is not working. SPCN sends an alert to CSP and CSP flags for Event_Scan to pick up and place an entry in the error log.
- EPOW 2 - Warn Power: This type of fault occurs when one of the system's redundant supplies stops working. SPCN sends an alert to CSP and CSP flags for Event_Scan to pick up and place an entry in the error log.
- EPOW 4 - System Halt (will start shut down 20 seconds from the time AIX gets the EPOW): This type of EPOW is flagged if the system must shut down for thermal reasons.

[7] RunTime Abstraction Services (RTAS) is the local firmware that is replicated to each LPAR of the system.

This happens when SPCN detects that two or more fans are not performing (or missing). For this type of message, SPCN powers off the offending domain if AIX does not power off the system first. The actual time may require longer than twenty seconds depending on the configuration and implementation.

There are two independent cooling domains in the p650 drawer, a top and bottom section. As such, some multiple fan fail conditions are not critical. Table on page 55 describes how the power and cooling redundancy works for the p650 drawer. You can match the EPOW level generated by the failures with the events reported inside the table.



	P/S 1	**P/S 2**	**Fan 1 on P/S 1**	**Fan 2 on P/S 2**	**Fan 3 Front left**	**Fan 4 Front right**	**EPOW Level**
Normal			Ramp 1	Ramp 1	Ramp 2	Ramp 2	N/A
*** Redundant Configuration - Recoverable Fail Actions Runtime ***							
Single fails	🟥		Set to max	Set to max			2
		🟥	Set to max	Set to max			2
	turn p/s off		🟥	Set to max			1
		turn p/s off	Set to max	🟥			1
					🟥	Set to max	1
					Set to max	🟥	1
Recoverable double fails: power supply and fan	🟥		🟥	Set to max			1 and 2
	🟥		Set to max	Set to max	🟥	Set to max	1 and 2
	🟥		Set to max	Set to max	Set to max	🟥	1 and 2
		🟥	Set to max	🟥			1 and 2
		🟥	Set to max	Set to max	🟥	Set to max	1 and 2
		🟥	Set to max	Set to max	Set to max	🟥	1 and 2
Recoverable double fails: two fans	turn p/s off		🟥	Set to max	🟥	Set to max	1
	turn p/s off		🟥	Set to max	Set to max	🟥	1
		turn p/s off	Set to max	🟥	🟥	Set to max	1
		turn p/s off	Set to max	🟥	Set to max	🟥	1
Recoverable triple fails: power supply and two fans	🟥		🟥	Set to max	🟥	Set to max	1 and 2
	🟥		🟥	Set to max	Set to max	🟥	1 and 2
		🟥	Set to max	🟥	🟥	Set to max	1 and 2
		🟥	Set to max	🟥	Set to max	🟥	1 and 2
*** Non-recoverable conditions - System will be powered off ***							
Non-recoverable conditions					🟥	🟥	4
			🟥	🟥			4
	🟥			🟥			4
		🟥	🟥				4

Figure 3-4   Recoverable/non-recoverable fail matrix

A white box indicates a device that is present and working correctly. A painted box indicates a failed or not present resource.

Note: When the EPOW level is 4, hardware responds as the description (shuts down within a few seconds). However, the error is logged and reported to the AIX error log as an EPOW 2.

3.3.8 Service Agent and Inventory Scout

Service Agent and Inventory Scout are two tools that can be used on the Model 6M2 to enable you to maintain the maximum availability of your system. Each item performs a different task, as discussed in the following sections.

Service Agent

Service Agent is the successor to Service Director. It is an application program that operates on a pSeries or RS/6000 computer and monitors it for hardware errors. When the p650 is operated in LPAR mode, Service Agent runs in the attached HMC. Service Agent reports detected errors, assuming they meet certain criteria for criticality. If any alertable problem is detected, then Service Agent (if configured to do so and connected to a suitable modem and analog telephone line) will alert the IBM service organization to request service. Also, if required, Service Agent can also be configured to send a notification by e-mail to the system administrator. Along with the request for service, sense data is also transmitted to enable an action plan to be prepared by the product support specialists. All data to be sent can be monitored by you before sending, if required, and a manual call initiated. IBM can provide support to set up the environment and parameters. Since licenses are checked by the IBM Service Agent Server, whenever a call is made to IBM, only machines under IBM Warranty or a Maintenance Agreement can use the Service Agent to report errors. This tool will be set up by your Customer Engineer upon your request. The latest User‘s Guide is available at the Web url:

ftp://ftp.software.ibm.com/aix/service_agent_code/AIX

What a Service Agent does

Here are some of the key things you can accomplish using the Electronic Service Agent for pSeries and RS/6000:

- Automatic problem analysis.
- Problem-definable threshold levels for error reporting.
- Automatic problem reporting; service calls placed to IBM without intervention.
- Automatic customer notification.
- Commonly viewed hardware errors; you can view hardware event logs for any monitored machine on the network from any Service Agent host user interface.
- High Availability Cluster Multiprocessing (HACMP) support for full fallback; includes High Availability Control Workstation (HACWS).
- Network environment support with minimum telephone lines for modems.
- VPD data can be sent to IBM using Performance Management.

Inventory Scout

Inventory Scout is being shipped with AIX 5L as a fileset or can be downloaded from the Internet. This tool will enable you to check the firmware or microcode levels of all of the devices in your system and advise you as to which code levels require attention. For more information, view the following Web page:

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

The option for Electronic Service Agent, or the service processor, to place a call to IBM is available at no additional cost, provided that the system is under warranty or an IBM service or maintenance contract is in place. The Electronic Service Agent monitors and analyzes system errors. For non-critical errors, Service Agent can place a service call automatically to IBM. For critical system failures (and if AIX is not operating), the dial-out is performed by the service processor itself, which also has the ability to send out an alert automatically using the telephone line to dial a paging service. This function is set up and controlled by the customer, not by IBM. It is not enabled by default. A hardware fault will also turn on the two attention indicators (one located on the front and the other located on the rear of the system) to alert the user of a hardware problem.

Note: Service Agent and Inventory Scout require that the p650 system has a graphic adapter, with keyboard and mouse or an ASCII terminal. Service Agent also requires Java Versions 1.1.6 or higher and X11 libraries on system that run the Graphic User Interface. ASCII User Interface does not require X11 libraries.

3.4 High-availability solution

For even greater availability and reliability, the p650 supports the IBM High Availability Cluster Multiprocessing (HACMP) software clustering solution. This solution helps to minimize downtime of systems and applications, providing a superior base for high availability. This is an essential ingredient of business-critical environments.

The p650 logically has four serial ports, all located at the back of the system (see Figure 1-3 on page 3). It is recommended that HACMP or UPS functions use the S3 or S4 port, because ports S1 and S2 are monitored by the service processor. If the user types any key from an ASCII terminal and that character comes in, the service processor selects that port to show the service processor menu to the ASCII terminal. This menu could confuse HACMP or an attached UPS; therefore, these ports should not be used.

Note: Order FC 3124 (HACMP serial to serial cable - drawer to drawer 3.7 meter) or FC 3125 (HACMP serial to serial cable - rack to rack 8 meter) for the serial non-IP heartbeat connections. FC 3925 converters are required for each end of either cable to attach it to the system.

3.5 IBM @server Cluster 1600 and SP switch attachment

The Model 6M2 is supported in either a non-switched IBM @server Cluster 1600 or a switched Cluster 1600 system using the SP Switch2 adapter (FC 8398).

Up to 32 p650s running in non-LPAR (full system partition) mode are supported in a cluster. A Cluster 1600 can scale up to 128 LPARs. The cluster management server can be running on any server or LPAR running AIX 5L Version 5.2 using IBM Cluster Systems Management for AIX 5L Version 1.3.1. PSSP[8] Version 3.5 is required to support SP Switch2 adapter (FC 8398) with AIX 5L Version 5.2 or Version 5.1.

To attach a Model 6M2 to a Cluster 1600, an HMC is required. One HMC can also control several Model 6M2s that are part of the cluster. If a Model 6M2 configured in LPAR mode is part of the cluster, all LPARs must be part of the cluster. It is not possible to use selected LPARs as part of the cluster and use others for non-cluster use.

The HMC uses a dedicated connection to the Model 6M2 to provide the functions needed to control the server, such as powering the system on and off. The HMC must have an Ethernet connection to the CWS. Each LPAR in Model 6M2 must have an Ethernet adapter to connect to the CWS *trusted* LAN.

Information regarding HMC control of clustered servers under the control of IBM Parallel Systems Support Programs for AIX (PSSP) or Cluster Systems Management for AIX (CSM) can be found in the Scaling Statement section of the Family 9078+01 IBM @server Cluster 1600, 9078-160 sales manual, accessible on IBMlink:

`http://www.ibmlink.ibm.com`

[8] Parallel System Support Programs (PSSP)

Related publications

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

IBM Redbooks

For information on ordering these publications, see “How to get IBM Redbooks” on page 61.

- *AIX Logical Volume Manager from A to Z: Introduction and Concepts,* SG24-5432
- *AIX Logical Volume Manager from A to Z: Troubleshooting and Commands,* SG24-5433
- *IBM @server pSeries 670 and pSeries 690 System Handbook,* SG24-7040
- *Practical Guide for SAN with pSeries,* SG24-6050
- *Problem Solving and Troubleshooting in AIX 5L*, SG24-5496
- *Understanding IBM @server pSeries Performance and Sizing,* SG24-4810

Other resources

These publications are also relevant as further information sources:

- *7014 Series Model T00 and T42 Rack Installation and Service Guide*, SA38-0577, contains information regarding the 7014 Model T00 and T42 Rack, in which this server may be installed.
- *Flat Panel Display Installation and Service Guide*, SA23-1243, contains information regarding the 7316-TF2 Flat Panel Display, which may be installed in your rack to manage your system units.
- *IBM @server pSeries 650 Model 6M2 Installation Guide*, SA38-0610, contains detailed information on installation, cabling, and verifying server operation.
- *IBM @server pSeries 650 Model 6M2 Service Guide*, SA38-0612, contains reference information, maintenance analysis procedures (MAPs), error codes, removal and replacement procedures, and a parts catalog.
- *IBM @server pSeries 650 Model 6M2 User’s Guide*, SA38-0611, contains information to help users use the system, use the service aids, and solve minor problems.
- *RS/6000 Adapters, Devices, and Cable Information for Multiple Bus Systems*, SA38-0516, contains information about adapters, devices, and cables for your system. This manual is intended to supplement the service information found in the Diagnostic Information for Multiple Bus Systems documentation.
- *RS/6000 and @server pSeries Diagnostics Information for Multiple Bus Systems*, SA38-0509, contains diagnostic information, service request numbers (SRNs), and failing function codes (FFCs).
- *RS/6000 and pSeries PCI Adapter Placement Reference*, SA38-0538, contains information regarding slot restrictions for adapters that can be used in this system.
- *System Unit Safety Information*, SA23-2652, contains translations of safety information used throughout the system documentation.

Referenced Web sites

These Web sites are also relevant as further information sources:

- Autonomic computing on IBM eServer pSeries servers

  http://www.ibm.com/autonomic/index.shtml
- Ceramic Column Grid Array (CCGA), see IBM Chip Packaging

  http://www.ibm.com/chips/micronews
- Copper circuitry

  http://www.ibm.com/chips/bluelogic/showcase/copper/
- ESS information

  http://www.storage.ibm.com/hardsoft/products/ess/index.html
- FAStT family: Support of additional features and for further information

  http://www.storage.ibm.com/hardsoft/disk/fastt/index.html
- Frequently asked SSA-related questions

  http://www.storage.ibm.com/hardsoft/products/ssa/faq.html
- Hardware documentation

  http://www.ibm.com/servers/eserver/pseries/library/hardware_docs
- IBM @server support: Fixes

  http://techsupport.services.ibm.com/server/fixes
- IBM @server support: Tips for AIX administrators

  http://techsupport.services.ibm.com/server/aix.techTips
- IBM Linux news

  http://www-1.ibm.com/servers/eserver/pseries/linux/
- SuSE Linux Enterprise Server 8

  http://www.suse.com/us/business/products/server/sles/i_pseries.html
- IBM Storage homepage

  http://www.storage.ibm.com/
- Linux for IBM @server pSeries

  http://www.ibm.com/servers/eserver/pseries/linux/
- Microcode discovery service

  http://techsupport.services.ibm.com/server/aix.invscoutMDS
- Pervasive system management

  http://www.ibm.com/servers/pervasivesm/
- POWER4 system mircoarchitecture

  http://www.research.ibm.com/journal/rd/461/tendler.html
- POWER4 system microarchitecture: Comprehensively described in the IBM Journal of Research and Development, Vol 46 No.1 January 2002.

  http://www.research.ibm.com/journal/rd46-1.html
- PowerPC Microprocessor Common Hardware Reference Platform (CHRP): A System Architecture

  http://www.mkp.com/books_catalog/catalog.asp?ISBN=1-55860-394-8

- SCSI terms and terminology

  http://www.scsita.org/terms/scsiterms.html

- Silicon on Insulator (SOI) technology

  http://www.ibm.com/chips/bluelogic/showcase/soi

How to get IBM Redbooks

You can order hardcopy Redbooks, as well as view, download, or search for Redbooks at the following Web site:

ibm.com/redbooks

You can also download additional materials (code samples or diskette/CD-ROM images) from that site.

IBM Redbooks collections

Redbooks are also available on CD-ROMs. Click the CD-ROMs button on the Redbooks Web site for information about all the CD-ROMs offered, as well as updates and formats.

IBM @server pSeries 650 Model 6M2 Technical Overview and Introduction

Redpaper

Innovative, POWER4+ processor, RIO-2 I/O, LPAR, DLPAR, and hot-swap features

Expandable with up to eight I/O drawers, providing a total of 63 PCI-X adapter slots and 100 disks

High-end reliability, availability, and serviceability features

This document provides a comprehensive guide covering IBM @server pSeries 650 servers. Major hardware offerings are introduced and their prominent functions discussed.

Professionals wishing to acquire a better understanding of IBM @server pSeries products may consider reading this document. The intended audience includes:

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

This document expands the current set of pSeries documentation by providing a desktop reference that offers a detailed technical description of the IBM @server pSeries 650.

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