IBM Power E1080
Technical Overview and Introduction

Giuliano Anselmi

Manish Arora

Ivaylo Bozhinov
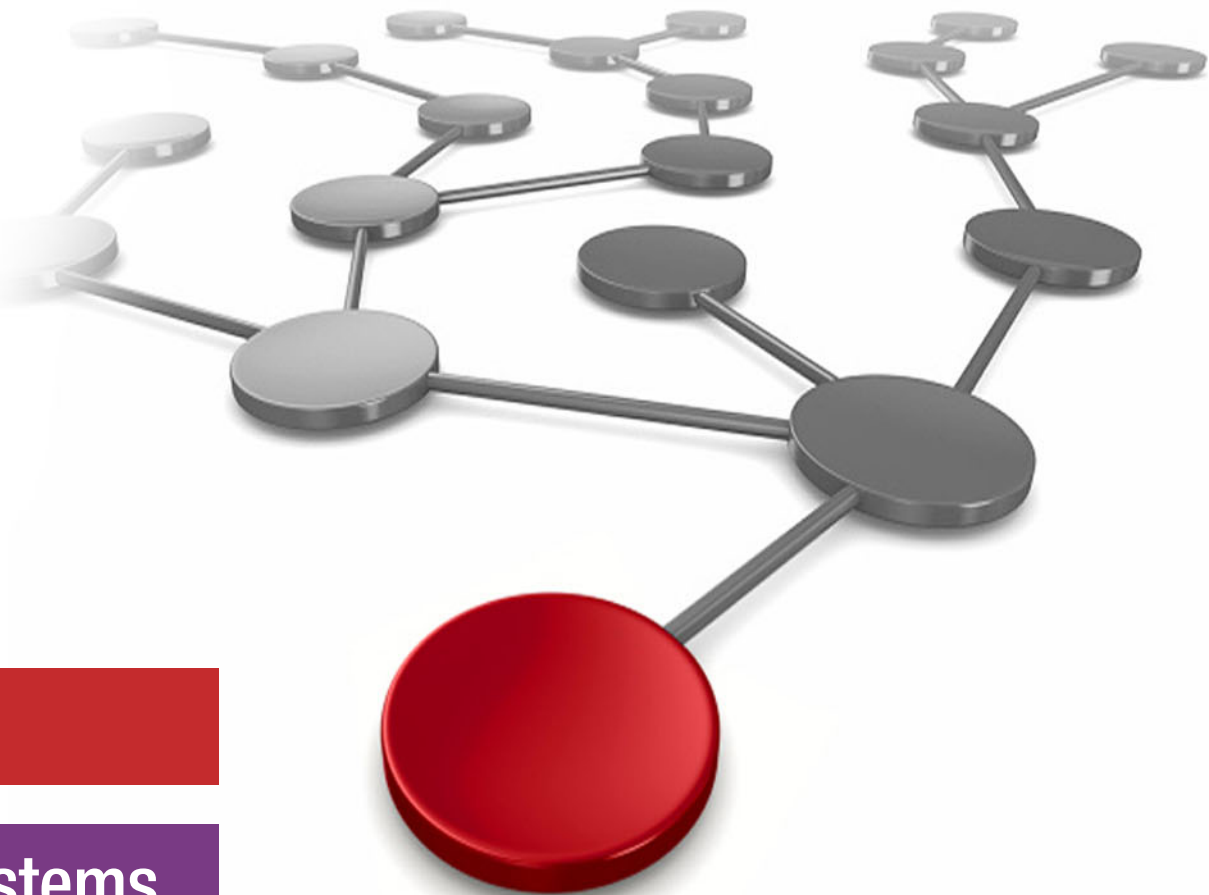
Dinil Das

Turgut Genc

Bartlomiej Grabowski

Madison Lee

Armin Röll

Cloud

Power Systems

IBM IBM Redbooks

IBM Power E1080: Technical Overview and Introduction

October 2021

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

First Edition (October 2021)

This edition applies to IBM Power E1080 (9080-HEX) server and the Hardware Management Console (HMC) Version 10 Release 1.

Important: At time of publication, this book is based on a pre-GA version of a product. For the most up-to-date information regarding this product, consult the product documentation or subsequent updates of this book.

This document was created or updated on August 30, 2023.

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

Redbooks (logo) ®	IBM FlashCore®	Power Architecture®
AIX®	IBM FlashSystem®	Power10™
C3®	IBM Spectrum®	POWER8®
Db2®	IBM Z®	POWER9™
Easy Tier®	Interconnect®	PowerHA®
IBM®	Micro-Partitioning®	PowerVM®
IBM Cloud®	POWER®	Redbooks®

The following terms are trademarks of other companies:

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

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

Ansible, OpenShift, Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

VMware, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

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

Preface

This IBM® Redpaper® publication provides a broad understanding of a new architecture of the IBM Power® E1080 (also known as the Power E1080) server that supports IBM AIX®, IBM i, and selected distributions of Linux operating systems. The objective of this paper is to introduce the Power E1080, the most powerful and scalable server of the IBM Power portfolio, and its offerings and relevant functions:

► Designed to support up to four system nodes and up to 240 IBM Power10™ processor cores

  The Power E1080 can be ordered as a one, two, three, or four node configuration with each node providing up to 60 Power10 processor cores to a maximum 240 Power10 processor cores in a four node configuration.

► Designed to supports up to 64 TB memory

  The Power E1080 can be initially ordered with the total memory RAM capacity up to 8 TB and supports up to 64 TB in a full combined four system nodes server.

► Designed to support up to 32 Peripheral Component Interconnect® (PCIe) Gen 5 slots in a full combined four system nodes server and up to 192 PCIe Gen 3 slots with expansion I/O drawers

► Up to over 4,000 directly attached serial-attached SCSI (SAS) disks or solid-state drives (SSDs)

► Up to 1,000 virtual machines (VMs) with logical partitions (LPARs) per system

► System control unit, providing redundant system master Flexible Service Processor (FSP)

► Supports IBM Power System Private Cloud Solution with Dynamic Capacity

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

► Customers
► Sales and marketing professionals
► Technical support professionals
► IBM Business Partners
► Independent software vendors (ISVs)

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

Authors

This paper was produced by a team of specialists from around the world working at IBM Redbooks, Austin Center.

Giuliano Anselmi is an IBM Power Digital Sales Technical Advisor in IBM Digital Sales Dublin. He joined IBM and focused on Power processor-based technology. For almost 20 years, he covered several technical roles. He is an important resource for the mission of his group and serves a reference for Business Partners and customers.

Brian W. Thompto, Madhavi Valluri, Madeline Vega
IBM

Now you can become a published author, too!

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

ibm.com/redbooks

► Send your comments in an email to:

redbooks@us.ibm.com

► Mail your comments to:

IBM Corporation, IBM Redbooks
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

► Find us on LinkedIn:

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

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

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

► Stay current on recent Redbooks publications with RSS Feeds:

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

Introduction to Power E1080

The Power E1080 is the newest addition to IBM Power family, the industry's best-in-class server platform for security and reliability. The Power E1080 introduces the essential enterprise hybrid cloud platform, which is uniquely designed to help you securely and efficiently scale core operational and artificial intelligence (AI) applications anywhere in a hybrid cloud.

The Power E1080 simplifies end-to-end encryption and brings AI where your data is stored for faster insights. This configuration helps enable greater workload deployment flexibility and agility while accomplishing more work.

The Power E1080 can help you to realize the following benefits:

► Protect trust from core to cloud

Protect data that is in-transit and at-rest with greatly simplified end-to-end encryption across hybrid cloud without affecting performance.

► Enjoy enterprise quality of service

The Power E1080 can detect, isolate, and recover from soft errors automatically in the hardware without taking an outage or relying on an operating system to manage the faults.

► Drive greater efficiency with sustainable and scalable compute

The processor performance, massive system throughput, and memory capacity qualify the Power E1080 server to be the perfect workload consolidation platform. This performance Sentails significant savings in floor space, energy consumption, and operational expenditure costs.

This chapter includes the following topics:

1.1 System overview

The Power E1080, also referred to by its 9080-HEX machine type-model designation, represents the most powerful and scalable server in the IBM Power portfolio. It is comprised of a combination of CEC enclosures that are called *nodes* (or *system nodes*) and more units and drawers.

1.1.1 System nodes, processors, and memory

In this section, we provide a general overview of the system nodes, processors, and memory. For more information about the system nodes, see 1.2, "System nodes" on page 6.

A system node is an enclosure that provides the connections and supporting electronics to connect the processor with the memory, internal disk, adapters, and the interconnects that are required for expansion.

A combination of one, two, three, or four system nodes per server is supported.

Each system node provides four sockets for Power10 processor chips and 64 differential DIMM (DDIMM) slots for Double Data Rate 4 (DDR4) technology DIMMs.

Each socket holds one Power10 single chip module (SCM). An SCM can contain 10, 12, or 15 Power10 processor cores. It also holds the extra infrastructure logic to provide electric power and data connectivity to Power10 processor chip.

The processor configuration of a system node is defined by the selected processor feature. Each feature defines a set of four Power10 processors chips with the same core density (10, 12, or 15).

A 4-node Power E1080 server scales up to 16 processor sockets and 160, 192, or 240 cores, depending on the number of cores provided by the configured SCM type.

All system nodes with a Power E1080 server must be configured with the same processor feature.

Each system node can support up to a maximum of 16 TB of system memory by using the largest available memory DIMM density. A fully configured 4-node Power E1080 can support up to 64 TB of memory.

To support internal boot capability, each system node enables the use of up to four non-volatile memory express (NVMe) drive bays. More drive bays are provided through expansion drawers.

Each system node provides eight Peripheral Component Interconnect Express (PCIe) Gen 5 capable slots, with a maximum of 32 per Power E1080 server.

Any one-, two-, three-, or four-system node configuration requires the system control unit (SCU) to operate. The SCU provides system hardware, firmware, and virtualization control through redundant Flexible Service Processors (FSPs). Only one SCU is required and supported for every Power E1080 server. For more information about the system control unit, see 1.3, "System control unit" on page 10.

For more information about the environmental and physical aspects of the server, see 1.4, "Server specifications" on page 11.

1.1.2 Expansion drawers and storage enclosures

Capacity can be added to your system by using expansion drawers and storage enclosures.

An optional 19-inch PCIe Gen 3 4U I/O expansion drawer provides 12 PCIe Gen 3 slots. The I/O expansion drawer connects to the system node with a pair of PCIe x16 to CXP converter cards that are housed in the system node. Each system node can support up to four I/O expansion drawers, for a total of 48 PCIe Gen 3 slots. A fully configured Power E1080 can support a maximum of 16 I/O expansion drawers, which provides a total of 192 PCIe Gen 3 slots.

An optional EXP24SX SAS storage enclosure provides 24 2.5-inch small form factor (SFF) serial-attached SCSI (SAS) bays. It supports up to 24 hot-swap hard disk drives (HDDs) or solid-state drives (SSDs) in only 2U rack units of space in a 19-inch rack. The EXP24SX is connected to the Power E1080 server by using SAS adapters that are plugged into system node PCIe slots or I/O expansion drawer slots.

For more information about enclosures and drawers, see 1.6, "I/O drawers" on page 26.

For more information about IBM storage products, see this web page.

1.1.3 Hardware at-a-glance

The Power E1080 server provides the following hardware components and characteristics:

► 10-, 12-, or 15-core Power10 processor chips that ar packaged in a single chip module per socket
► One, two, three, or four system nodes with four Power10 processor sockets each
► Redundant clocking in each system node
► Up to 60 Power10 processor cores per system node and up to 240 per system
► Up to 16 TB of DDR4 memory per system node and up to 64 TB per system
► 8 PCIe Gen 5 slots per system node and a maximum of 32 PCIe Gen 5 slots per system
► PCIe Gen 1, Gen 2, Gen 3, Gen 4, and Gen 5 adapter cards supported in system nodes
► Up to 4 PCIe Gen 3 4U I/O expansion drawers per system node providing a maximum of 48 additional PCIe Gen 3 slots
► Up to 192 PCIe Gen 3 slots using 16 PCIe Gen 3 I/O expansion drawers per system
► Up to over 4,000 directly attached SAS HDDs or SSDs through EXP24SX SFF drawers
► System control unit, which provides redundant Flexible Service Processors and support for the operations panel, the system VPD, and external attached DVD

The massive computational power, exceptional system capacity, and the unprecedented scalability of the Power E1080 server hardware are unfolded by unique enterprise class firmware and system software capabilities and features. The following important characteristics and features are offered by the IBM Power enterprise platform:

► Support for IBM AIX, IBM i, and Linux operating system environments
► Innovative dense math engine that is integrated in each Power10 processor-core to accelerate AI inferencing workloads
► Optimized encryption units that are implemented in each Power10 processor-core
► Dedicated data compression engines that are provided by the Power10 processor technology

- ► Hardware and firmware assisted and enforced security provide trusted boot and pervasive memory encryption support
- ► Up to 1,000 virtual machines (VMs) or logical partitions (LPARs) per system
- ► Dynamic LPAR support to modify available processor and memory resources according to workload, without interruption of the business
- ► Capacity on demand (CoD) processor and memory options to help respond more rapidly and seamlessly to changing business requirements and growth
- ► IBM Power System Private Cloud Solution with Dynamic Capacity featuring Power Enterprise Pools 2.0 that support unsurpassed enterprise flexibility for real-time workload balancing, system maintenance and operational expenditure cost management.

Table 1-1 compares important technical characteristics of the Power E1080 server with those of the Power System E980 server, based on IBM POWER9™ processor-based technology.

Table 1-1   Comparison between the Power E980 and the Power E1080 server

Features	Power E980 server	Power E1080 server
Processor	POWER9	Power10
Processor package	Single Chip Module	Single Chip Module
Cores per single chip module	6, 8, 10, 11, 12	10, 12, 15
Number of cores per system	Up to 192 cores	Up to 240 cores
Sockets per node	4	4
System configuration options	1-, 2-, 3-, and 4-node systems	1-, 2-, 3-, and 4-node systems
Maximum memory per node	16 TB	16 TB
Maximum memory per system	64 TB	64 TB
Maximum memory bandwidth per node	920 GBps	1636 GBps
Aggregated maximum memory bandwidth per system	3680 GBps	6544 GBps
Pervasive memory encryption	No	Yes
PCIe slots per node	Eight PCIe Gen 4 slots	Eight PCIe Gen 5 slots
I/O drawer expansion option	Yes	Yes
Acceleration ports	Yes CAPI 2.0 & OpenCAPI 3.0[a]	Yes OpenCAPI 3.0, 3.1
PCIe hot-plug Support	Yes	Yes
I/O bandwidth per node	545 GBps	576 GBps
Integrated USB	USB 3.0	Not available
Internal storage bays per node	Four NVMe PCIe Gen 3 bays[b]	Four NVMe PCIe Gen 4 bays
Per lane bit rate between sockets	25 Gbps	32 Gbps

Features	Power E980 server	Power E1080 server
Reliability, availability, and serviceability (RAS)	SMP[c] cable concurrent repair	Non-active SMP cables with concurrent maintenance capability and TDR[d] fault isolation
Secure and trusted boot	Yes	Yes

a. CAPI designates the coherent accelerator processor interface technology and OpenCAPI designates the open coherent accelerator processor interface technology. For more information about architectural specifications and the surrounding system, see this web page.

b. NVMe designates the Non-Volatile Memory Express interface specification under supervision of the NVM Express consortium: https://nvmexpress.org.

c. SMP designates the symmetric multiprocessing architecture, which is used to build monolithic servers out of multiple processor entities.

d. Time domain reflectometry (TDR) allows the server to actively detect faults in cables and locate discontinuities in a connector.

Figure 1-1 shows a 4-node Power E1080 server that is mounted in an IBM rack. Each system node is cooled by a set of five fans, which are arranged side-by-side in one row. The cooling assemblies show through the front door of the rack.

Figure 1-1   Power E1080 4-node server mounted in S42 rack with #ECRT door

1.1.4 System capacities and features

With any initial orders, the Power E1080 supports up to four system nodes. The maximum memory capacity that is supported in each node is 4 TB.

The maximum number of supported PCIe Gen 3 I/O expansion drawer is four per system node. Each I/O expansion drawer can be populated with two Fanout Modules. Each Fanout Module in turn is connected to a system node through one PCIe x16 to CXP Converter Card.

Memory features #EMC1 128 GB, #EMC2 256 GB, #EMC3 512 GB, and #EMC4 1024 GB are available.

The following characteristics are available:

► Maximum of 4 #EDN1 5U system node drawers
► Maximum of 16 TB of system memory per node drawer
► Maximum of 16 #EMX0 PCIe Gen 3 I/O expansion drawers
► Maximum of 32 #EMXH PCIe Gen 3 6-slot Fanout Module for PCIe Gen 3 expansion drawers
► Maximum of 32 #EJ24 PCIe x16 to CXP Converter Cards

1.2 System nodes

The full operational Power E1080 includes one SCU and one, two, three, or four system nodes. A system node is also referred to as a *central electronic complex* (CEC), or *CEC drawer*.

Each system node is 5U rack units high and holds four air-cooled Power10 single-chip modules (SCMs) that are optimized for performance, scalability, and AI workloads. An SCM is constructed of one Power10 processor chip and more logic, pins, and connectors that enable plugging the SCM into the related socket on the system node planar.

The Power E1080 Power10 SCMs are available in 10-core, 12-core, or 15-core capacity. Each core can run in eight-way simultaneous multithreading (SMT) mode, which delivers eight independent hardware threads of parallel execution power.

The 10-core SCMs are ordered in a set of four per system node through processor feature #EDP2. In this way, feature #EDP2 provide 40 cores of processing power to one system node and 160 cores of total system capacity in a 4-node Power E1080 server. The maximum frequency of the 10-core SCM is specified with 3.9 GHz, which makes this SCM suitable as a building block for entry class Power E1080 servers.

The 12-core SCMs are ordered in a set of four per system node through processor feature #EDP3. In this way, feature #EDP3 provides 48 cores capacity per system node and a maximum of 192 cores per fully configured 4-node Power E1080 server. This SCM type offers the highest processor frequency at a maximum of 4.15 GHz, which makes it a perfect choice if highest thread performance is one of the most important sizing goals.

The 15-core SCMs are ordered in a set of four per system node through processor feature #EDP4. In this way, feature #EDP4 provides 60 cores per system node and an impressive 240 cores total system capacity for a 4-node Power E1080. The 15-core SCMs run with a maximum of 4.0 GHz and meet the needs of environments with demanding thread performance and high compute capacity density requirements.

Note: All Power10 SCMs within a system node must be of the same type: 10-core, 12-core, or 15-core. Also, all system nodes within a specific Power E1080 server must be configured consistently with identical processor features.

Three PowerAXON[1] 18-bit wide buses per Power10 processor chip are used to span a fully connected fabric within a CEC drawer. In this way, each SCM within a system node is directly connected to every other SCM of the same drawer at 32 Gbps speed. This on-planar interconnect provides 128 GBps chip-to-chip data bandwidth, which marks an increase of 33% relative to the previous POWER9 processor-based on-planar interconnect implementation in Power E980 systems. The throughput can be calculated as 16 b lane with * 32 gbps = 64 GBps per direction * 2 directions for an aggregated rate of 128 GBps.

Each of the four Power10 processor chips in a Power E1080 CEC drawer is connected directly to a Power10 processor chip at the same position in every other CEC drawer in a multi-node system This connection is made by using a symmetric multiprocessing (SMP) PowerAXON 18-bit wide bus per connection running at 32 Gbps speed.

The Power10 SCM provides eight PowerAXON connectors directly on the module of which six are used to route the SMP bus to the rear tailstock of the CEC chassis. This innovative implementation allows to use passive SMP cables, which in turn reduces the data transfer latency and enhances the robustness of the drawer to drawer SMP interconnect. As discussed in 1.2, "System nodes" on page 6, cable features #EFCH, #EFCE, #EFCF, and #EFCG are required to connect system node drawers to the system control unit. They also are required to facilitate the SMP interconnect among each drawer in a multi-node Power E1080 configuration.

To access main memory, the Power10 processor technology introduces the new open memory interface (OMI). The 16 available high-speed OMI links are driven by 8 on-chip memory controller units (MCUs) that provide a total aggregated bandwidth of up to 409 GBps per SCM. This design represents a memory bandwidth increase of 78% compared to the POWER9 processor-based technology capability.

Every Power10 OMI link is directly connected to one memory buffer-based differential DIMM (DDIMM) slot. Therefore, the four sockets of one system node offer a total of 64 DDIMM slots with an aggregated maximum memory bandwidth of 1636 GBps. The DDIMM densities supported in Power E1080 servers are 32 GB, 64 GB,128 GB, and 256 GB, all of which use Double Data Rate 4 (DDR4) technology.

The Power E1080 memory options are available as 128 GB (#EMC1), 256 GB (#EMC2), 512 GB (#EMC3), and 1024 GB (#EMC4) memory features. Each memory feature provides four DDIMMs.

---

[1] PowerAXON stands for A-bus/X-bus/OpenCAPI/Networking interfaces of the Power10 processor.

Each system node supports a maximum of 16 memory features that cover the 64 DDIMM slots. The use of 1024 GB DDIMM features yields a maximum of 16 TB per node. A 2-node system has a maximum of 32 TB capacity. A 4-node system has a maximum of 64 TB capacity. Minimum memory activations of 50% of the installed capacity are required.

> **Note:** The Power10 processor memory logic and the memory I/O subsystem of the Power E1080 server are designed to support DDR4 DIMMs, but also can use the next generation DDR5 dynamic random-access memory technology.

The Power10 processor I/O subsystem is driven by 32 GHz differential Peripheral Component Interconnect Express 5.0 (PCIe Gen 5) buses that provide 32 lanes that are grouped in two sets of 16 lanes. The 32 PCIe lanes deliver an aggregate bandwidth of 576 GBps per system node and are used to support 8 half-length, low-profile (half-height) adapter slots for external connectivity and 4 Non-Volatile Memory Express (NVMe) mainstream Solid State Drives (SSDs) of form factor U.2. for internal storage.

Six of the eight external PCIe slots can be used for PCIe Gen 4 x16 or PCIe Gen 5 x8 adapters and the remaining two offer PCIe Gen 5 x8 capability. All PCIe slots support earlier generations of the PCIe standard, such as PCIe Gen 1 (PCIe 1.0), PCIe Gen 2 (PCIe 2.0), PCIe Gen 3 (PCIe 3.0), and PCIe Gen 4 (PCIe 4.0).

For extra connectivity, up to four 19-inch PCIe Gen 3 4U high I/O expansion units (#EMX0) optionally can be attached to one system node. Each expansion drawer contains one or two PCIe Fanout Modules (#EMXH) with six PCIe Gen 3 full-length, full-height slots each.

A fully configured 4-node Power E1080 server offers a total of 32 internal PCIe slots and up to 192 PCIe slots through I/O expansion units.

Figure 1-2 shows the front view of a system node. The fans and power supply units (PSUs) are redundant and concurrently maintainable. Fans are n+1 redundant; therefore, the system continues to function when any one fan fails. Because the power supplies are n+2 redundant, the system continues to function, even if any two power supplies fail.
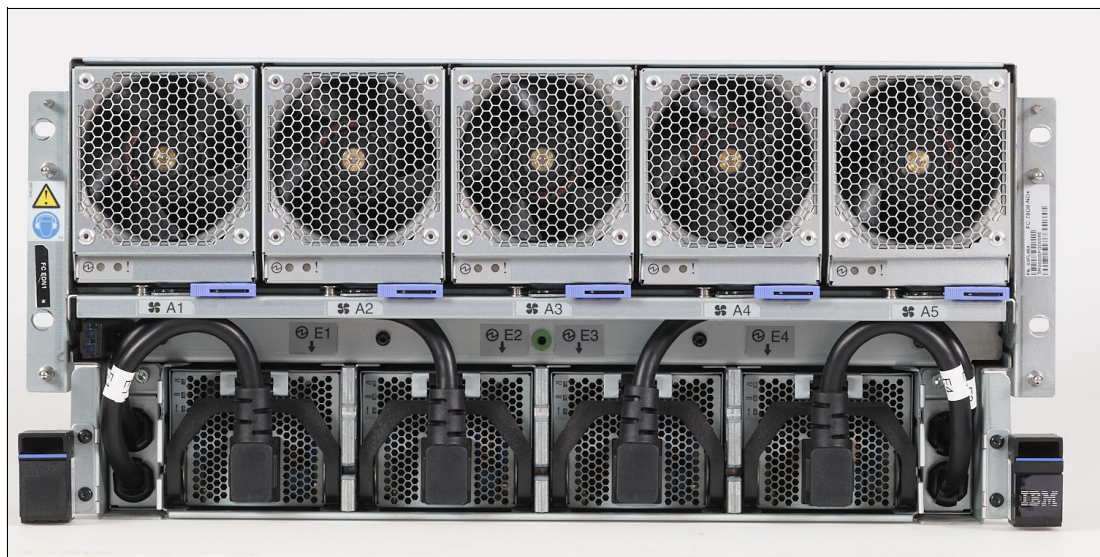
Figure 1-2   Front view of a Power E1080 server node

Figure 1-3 shows the rear view of a system node with the locations of the external ports and features.

Figure 1-3   Rear view of a Power E1080 server node

Figure 1-4 shows the internal view of a system node and some of the major components like heat sinks, processor voltage regulator modules (VRMs), VRMs of other miscellaneous components, differential DIMM (DDIMM) slots, DDIMMs, system clocks, trusted platform modules (TPMs), and internal SMP cables.
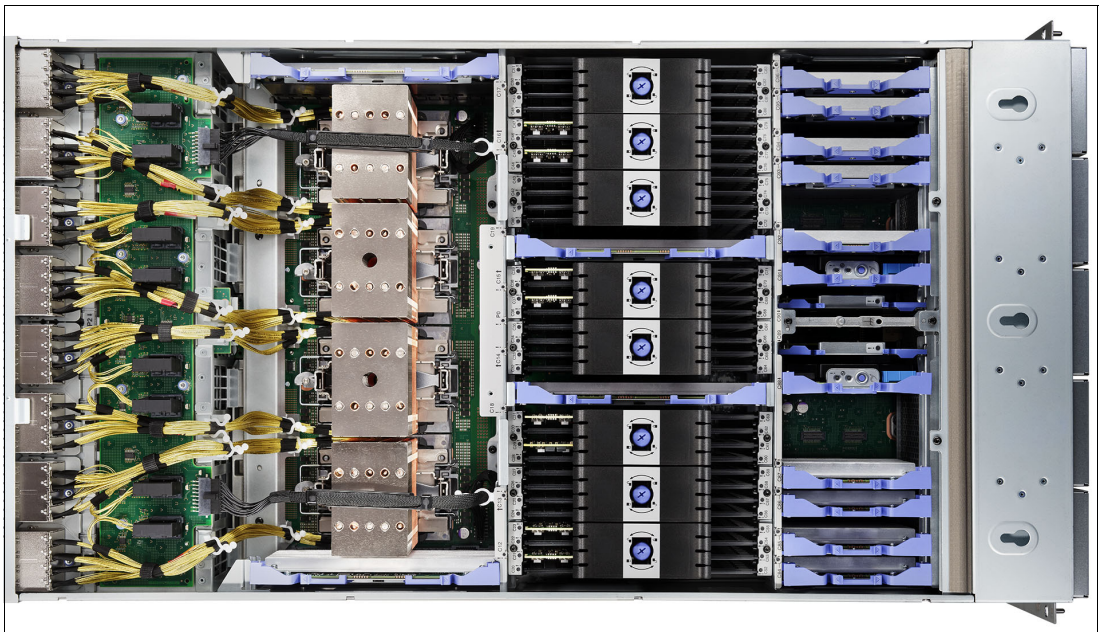
Figure 1-4   Top view of a Power E1080 server node with the top cover assembly removed

1.3 System control unit

The system control unit (SCU) is implemented in a 2U high chassis and provides system hardware, firmware, and virtualization control functions through a pair of redundant Flexible Service Processor (FSP) devices. It also contains the operator panel and the electronics module that stores the system vital product data (VPD). The SCU is also prepared to facilitate USB connectivity that can be used by the Power E1080 server.

One SCU is required and supported for each Power E1080 server (any number of system nodes) and, depending on the number of system nodes, the SCU is powered according to the following rules:

► Two universal power interconnect (UPIC) cables are used to provide redundant power to the SCU.

► In a Power E1080 single system node configuration, both UPIC cables are provided from the single system node to be connected to the SCU.

► For a two, three, or four system nodes configuration, one UPIC cable is provided from the first system node and the second UPIC cable is provided from second system node to be connected to the SCU.

The set of two cables facilitate a 1+1 redundant electric power supply. In case of a failure of one cable, the remaining UPIC cable is sufficient to feed the needed power to the SCU.

Two service processor cards in SCU are ordered by using two mandatory #EDFP features. Each one provides two 1 Gb Ethernet ports for the Hardware Management Console (HMC) system management connection. One port is used as primary connection and the second port can be used for redundancy. To enhance resiliency, it is recommended to implement a dual HMC configuration by attaching separate HMCs to each of the cards in the SCU.

Four FSP ports per FSP card provide redundant connection from the SCU to each system node. System nodes connect to the SCU by using the cable features #EFCH, #EFCE, #EFCF, and #EFCG. Feature #EFCH connects the first system node to the SCU and it is included by default in every system node configuration. It provides FSP, UBIC, and USB cables, but no symmetric multiprocessing (SMP) cables. All the other cable features are added depending on the number of extra system nodes that are configured and includes FSP and SMP cables.

The SCU implementation also includes the following highlights:

► Elimination of clock cabling since the introduction of POWER9 processor-based servers
► Front-accessible system node USB port
► Optimized UPIC power cabling
► Optional external DVD
► Concurrently maintainable time of day clock battery

Figure 1-5 shows the front and rear view of a SCU with the locations of the external ports and features.
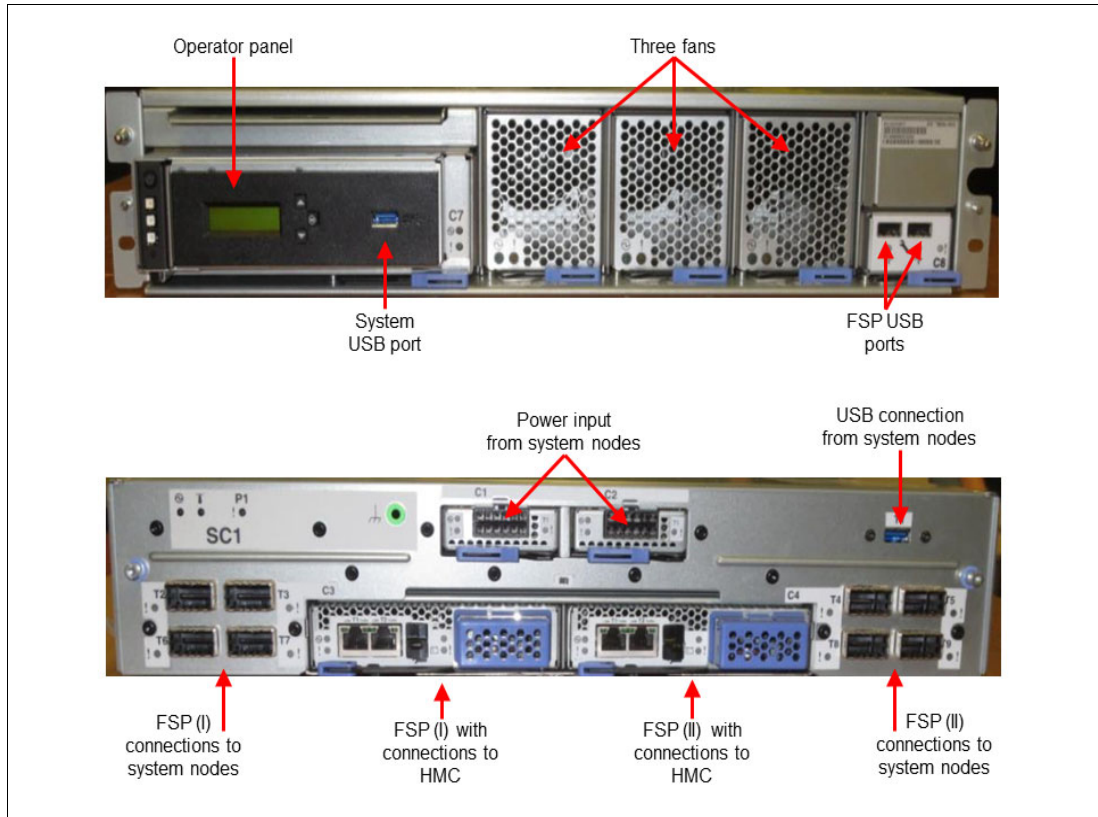
Figure 1-5   Front and rear view of the system control unit

1.4 Server specifications

The Power E1080 server specifications are essential to plan for your server. For a first assessment in the context of your planning effort, this section provides you with an overview related to the following topics:

► Physical dimensions
► Electrical characteristics
► Environment requirements and noise emission

For more information about the comprehensive Model 9080-HEX server specifications product documentation, see IBM Documentation.

1.4.1 Physical dimensions

The Power E1080 is a modular system that is build of a single SCU and one, two, three, or four system nodes.

Each system component must be mounted in a 19-inch industry standard rack. The SCU requires 2U rack units and each system node requires 5U rack units. Thus, a single-node system requires 7U, a two-node system requires 12U, a three-node system requires 17U, and a four-node system requires 22U rack units. More rack space must be allotted; for example, to PCIe I/O expansion drawers, a hardware management console, flat panel console kit, network switches, power distribution units, and cable egress space.

Table 1-2 lists the physical dimensions of the Power E1080 server control unit and a Power E1080 server node. The component height is also given in Electronic Industries Alliance (EIA) rack units. (One EIA unit corresponds to one rack unit (U) and is defined as 1 3/4 inch or 44.45 mm respectively).

Table 1-2   Physical dimensions of the Power E1080 server components,

Dimension	Power E1080 server control unit	Power E1080 server node
Width	445.6 mm (17.54 in)	445 mm (17.51 in)
Depth	779.7 mm (30.7 in)	866.95 mm (34.13 in)
Height	86 mm (3.39 in) / 2 EIA units	217.25 mm (8.55 in) / 5 EIA units
Weight	22.7 kg (50 lb)	81.6 kg (180 lb)

Lift tools

It is recommended to have a lift tool available at each site where one or more Power E1080 servers are located to avoid any delays when servicing systems. An optional lift tool #EB2Z is available for order with a Power E1080 server. One #EB2Z lift tool can be shared among many servers and I/O drawers. The #EB2Z lift tool provides a hand crank to lift and position up to 159 kg (350 lb). The #EB2Z lift tool is 1.12 meters x 0.62 meters (44 in x 24.5 in).

Note: A single system node can weigh up to 86.2 kg (190 lb). Also available are a lighter, lower-cost lift tool #EB3Z and wedge shelf tool kit for #EB3Z with feature code #EB4Z.

1.4.2 Electrical characteristics

Each Power E1080 server node has four 1950 W bulk power supplies. The hardware design provides N+2 redundancy for the system power supply, and any node can continue to operate at full function in nominal mode with any two of the power supplies functioning.

Depending on the specific Power E1080 configuration, the power for the SCU is provided through two UPIC cables connected to one or two system nodes, as described in 1.2, "System nodes" on page 6.

Table 1-3 lists the electrical characteristics per Power E1080 server node. For planning purposes, use the maximum values that are provided. However, the power draw and heat load depends on the specific processor, memory, adapter, and expansion drawer configuration and the workload characteristics.

Table 1-3   Electrical characteristics per Power E1080 server node

Electrical characteristics	Properties
Operating voltage	200 - 208 / 220 - 240 V AC
Operating frequency	50 or 60 Hz +/- 3 Hz AC
Maximum power consumption	4500 W
Maximum power source loading	4.6 kVA
Maximum thermal output	15355 BTU/h
Phase	Single

Note: The Power E1080 must be installed in a rack with a rear door and side panels for electromagnetic compatibility (EMC) compliance.

1.4.3 Environment requirements and noise emission

Environmental assessment: The IBM Systems Energy Estimator tool can provide more accurate information about power consumption and thermal output of systems based on a specific configuration.

The environment requirements for the Power E1080 servers are classified in operating and non-operating environments. The operating environments are further segmented regarding the recommended and allowable conditions.

The recommended operating environment designates the long-term operating environment that can result in the greatest reliability, energy efficiency, and reliability. The allowable operating environment represents where the equipment is tested to verify functionality. Because the stresses that operating in the allowable envelope can place on the equipment, these envelopes must be used for short-term operation, not continuous operation.

The condition of a non-operating environment pertains to the situation when equipment is removed from the original shipping container and is installed, but is powered down. The allowable non-operating environment is provided to define the environmental range that an unpowered system can experience short term without being damaged.

Table 1-4 on page 14 lists the environment requirements for the Power E1080 server regarding temperature, humidity, dew point, and altitude. It also lists the maximum noise emission level for a fully configured Power E1080 server.

Table 1-4   Power E1080 environment requirements and noise emission

Property	Environment		
	Operating		Non-operating
	Recommended	Allowable	
Temperature	18.0°C – 27.0°C (64.4°F – 80.6°F)	5.0°C – 40.0°C (41.0°F – 104.0°F)	5°C - 45°C (41°F - 113°F)
Low-end moisture	-9.0°C (15.8°F) dew point	-12.0°C (10.4°F) dew point and 8% relative humidity	N/A
High-end moisture	60% relative humidity and 15°C (59°F) dew point	85% relative humidity and 24.0°C (75.2°F) dew point	N/A
Relative humidity	N/A	N/A	8% to 85%
Maximum dew point	N/A	N/A	27.0°C (80.6°F)
Maximum altitude	N/A	3,050 m (10,000 ft)	N/A
Maximum noise level	10.0 B $L_{WA,m}$[a] (heavy workload on one fully configured 16-socket four-node system, 35°C (95°F) at 500 m (1640 ft))	N/A	N/A

a. Declared level $L_{WA,m}$ is the upper-limit A-weighted sound power level measured in bel (B).

Note: IBM does not recommend operation above 27C, however, one can expect full performance up to 35C for these systems. Above 35C, the system is capable of operating, but possible reductions in performance may occur to preserve the integrity of the system components. Above 40C there may be reliability concerns for components within the system.

A comprehensive list of noise emission values for various different Power E1080 server configurations is provided by the Power E1080 product documentation. For more information about noise emissions, search for "Model 9080-HEX server specifications" at IBM Documentation.

Note: Government regulations, such as those regulations that are issued by the Occupational Safety and Health Administration (OSHA) or European Community Directives, can govern noise level exposure in the workplace and might apply to you and your server installation. The Power E1080 server is available with an optional acoustical door feature that can help reduce the noise emitted from this system.

The sound pressure levels in your installation depend on various factors, including the number of racks in the installation; the size, materials, and configuration of the room where you designate the racks to be installed; the noise levels from other equipment; the room ambient temperature, and employees' location in relation to the equipment.

Also, compliance with such government regulations depends on various other factors, including the duration of employees' exposure and whether employees wear hearing protection. IBM recommends that you consult with qualified experts in this field to determine whether you are in compliance with the applicable regulations.

1.5 System features

This section lists and explains the available system features on a Power E1080 server. These features describe the resources that are available on the system by default or by virtue of procurement of configurable feature codes.

An overview of various feature codes and the essential information also is presented that can help users design their system configuration with suitable features that can fulfill the application compute requirement. This information also helps with building a highly available, scalable, reliable, and flexible system around the application.

1.5.1 Minimum configuration

A minimum configuration babels a user to order a fully qualified and tested hardware configuration of a Power system with a minimum set of offered technical features. The modular design of a Power E1080 server enables the user to start low with a minimum configuration and scale up vertically as and when needed.

Table 1-5 lists the Power E1080 server configuration with minimal features.

Table 1-5   Minimum configuration

Feature	FC	Feature code description	Min quantity
Primary operating system Feature Code	► 2145 ► 2146 ► 2147	► Primary OS - IBM i ► Primary OS - AIX ► Primary OS - Linux	1
System enclosure	EDN1	5U System node Indicator drawer	1
Flexible service processor	EDFP	Flexible service processor	2
Bezel	EBAB	IBM Rack-mount Drawer Bezel and Hardware	1
Processor	► EDP2 ► EDP3 ► EDP4	► 40-core (4x10) Typical 3.65 to 3.90 GHz (max) Power10 Processor with 5U system node drawer ► 48-core (4x12) Typical 3.60 to 4.15 GHZ (max) Power10 Processor with 5U system node drawer ► 60-core (4x15) Typical 3.55 to 4.00 GHz (max) Power10 Processor with 5U system node drawer	1 of either FC
Processor activation	► EDPB ► EDPC ► EDPD	► 1 core Processor Activation for #EDBB/#EDP2 ► 1 core Processor Activation for #EDBC/#EDP3 ► 1 core Processor Activation for #EDBD/#EDP4	► 16 ► 1 if in PEP 2.0
Memory	► EMC1 ► EMC2 ► EMC3 ► EMC4	► 128 GB (4x32 GB) DDIMMs, 3200 MHz, 16 GBIT DDR4 Memory ► 256 GB (4x64 GB) DDIMMs, 3200 MHz, 16 GBIT DDR4 Memory ► 512 GB (4x128 GB) DDIMMs, 2933 MHz, 16 GBIT DDR4 ► 1 TB (4x256 GB) DDIMMs, 2933 MHz, 16 GBIT DDR4	8 features of either FC

Feature	FC	Feature code description	Min quantity
Memory activation	EMAZ	1 GB Memory activation for HEX	50% or 256 GB in PEP 2.0
DASD backplanes	EJBC	4-NVMe U.2 (7mm) Flash drive bays	1
Data protection	0040	Mirrored System Disk Level, Specify Code	► 1 if IBM i ► 0 for other OS
UPIC cables	EFCH	System Node to System Control Unit Cable Set for Drawer 1	1

Note: The minimum configuration that is generated by IBM configurator includes more administrative and indicator features.

1.5.2 Processor features

Each system node in a Power E1080 server provides four sockets to accommodate Power10 single chip modules (SCMs). Each processor feature code represents four of these sockets, which are offered in 10-core, 12-core, and 15-core density.

Table 1-6 lists the available processor feature codes for a Power E1080 server. The system configuration requires one, two, three, or four quantity of same processor feature, according to the number of system nodes.

Table 1-6   Processor features.

Feature code	CCIN	Description	OS support
EDP2	5C6C	40-core (4x10) Typical 3.65 to 3.90 GHz (max) Power10 Processor with 5U system node drawer	AIX, IBM i, and Linux
EDP3	5C6D	48-core (4x12) Typical 3.60 to 4.15 GHZ (max) Power10 Processor with 5U system node drawer	AIX, IBM i, and Linux
EDP4	5C6E	60-core (4x15) Typical 3.55 to 4.00 GHz (max) Power10 Processor with 5U system node drawer	AIX, IBM i, and Linux

The system nodes connect to other system nodes and to the system control unit through cable connect features. Table 1-7 lists the set of cable features that are required for one-, two-, three-, and four-node configurations.

Table 1-7   Cable set features quantity

System configuration	Qty EFCH	Qty EFCG	Qty EFCF	Qty EFCE
1-node	1	0	0	0
2-node	1	0	0	1
3-node	1	0	1	1
4-node	1	1	1	1

Every feature code that is listed in Table 1-6 on page 16 provides the processor cores, not their activation. The processor core must be activated to be assigned as resource to a logical partition. The activations are offered through multiple permanent and temporary activation features. For more information about these options, see 2.4, "Capacity on-demand" on page 76.

Table 1-8 lists the processor feature codes and the associated permanent activation features. Any of these activation feature codes can permanently activate one core.

Table 1-8   Processor and activation features

Processor feature	Static activation feature	Static linux only activation feature
EDP2	EDPB	ELCL
EDP3	EDPC	ELCQ
EDP4	EDPD	ECLM

The following types of permanent activations are available:

Static   These features permanently activate cores or memory resources in a system. These activations cannot be shared among multiple systems and remain associated with the system for which they were ordered.

Regular   A regular static activation can run any supported operating system workload.

Linux-only   A Linux-only static activation can run only Linux workloads and are priced less than regular static activations.

Mobile   These features permanently activate cores or memory resources. They are priced more than static activation features because they can be shared among multiple eligible systems that are participating in a Power Enterprise Pool. They also can be moved dynamically among the systems without IBM involvement, which brings more value to the customer than static activations.

Base   In a Power Enterprise Pool 2.0 (PEP 2.0) environment, systems are ordered with some initial compute capacity. This initial capacity is procured by using base activation features. These activations do not move like mobile activations in PEP 1.0 environment, but can be shared among multiple eligible systems in a PEP 2.0 pool.

Any OS base   These base activations are supported on any operating system (AIX, IBM i, or Linux) in a PEP 2.0 environment.

Linux only base   These base activations are priced less than any OS base activation, but support only Linux workloads in a PEP 2.0 environment.

A minimum of 16 processor cores must always be activated with the static activation features, regardless of the Power E1080 configuration. Also, if the server is associated to a PEP 2.0, a minimum of one base activation is required.

For more information about other temporary activation offerings that are available for the Power E1080 server, see 2.4, "Capacity on-demand" on page 76.

Regular and PEP 2.0 associated activations for Power E1080 are listed in Table 1-9. The Order type table column includes the following designations:

Initial Denotes the orderability of a feature for *only* the new purchase of the system.

MES Denotes the orderability of a feature for *only* the MES upgrade purchases on the system.

Both Denotes the orderability of a feature for new *and* MES upgrade purchases.

Supported Denotes that feature is not orderable, but is supported. That is, the feature can be migrated only from existing systems.

Table 1-9   Processor activation features

Feature code	Description	Order type
EPS0	1 core Base Proc Act (Pools 2.0) for #EDP2 any OS (from Static)	MES
EPS1	1 core Base Proc Act (Pools 2.0) for #EDP3 any OS (from Static)	MES
EPS2	1 core Base Proc Act (Pools 2.0) for #EDP4 any OS (from Static)	MES
EPS5	1 core Base Proc Act (Pools 2.0) for #EDP2 Linux (from Static)	MES
EPS6	1 core Base Proc Act (Pools 2.0) for #EDP3 Linux (from Static)	MES
EPS7	1 core Base Proc Act (Pools 2.0) for #EDP4 Linux (from Static)	MES
EPSK	1 core Base Proc Act (Pools 2.0) for #EDP2 any OS (from Prev)	MES
EPSL	1 core Base Proc Act (Pools 2.0) for #EDP3 any OS (from Prev)	MES
EPSM	1 core Base Proc Act (Pools 2.0) for #EDP4 any OS (from Prev)	MES
EDP2	40-core (4x10) Typical 3.65 to 3.90 GHz (max) Power10 Processor with 5U system node drawer	Both
EDP3	48-core (4x12) Typical 3.60 to 4.15 GHZ (max) Power10 Processor with 5U system node drawer	Both
EDP4	60-core (4x15) Typical 3.55 to 4.00 GHz (max) Power10 Processor with 5U system node drawer	Both
ED2Z	Single 5250 Enterprise Enablement	Both
ED30	Full 5250 Enterprise Enablement	Both
EDPB	1 core Processor Activation for #EDBB/#EDP2	Both
EDPC	1 core Processor Activation for #EDBC/#EDP3	Both
EDPD	1 core Processor Activation for #EDBD/#EDP4	Both
EDPZ	Mobile processor activation for HEX/80H	Both
EPDC	1 core Base Processor Activation (Pools 2.0) for EDP2 any OS	Both
EPDD	1 core Base Processor Activation (Pools 2.0) for EDP3 any OS	Both
EPDS	1 core Base Processor Activation (Pools 2.0) for EDP4 any OS	Both
EPDU	1 core Base Processor Activation (Pools 2.0) for EDP2 Linux only	Both
EPDW	1 core Base Processor Activation (Pools 2.0) for EDP3 Linux only	Both

Feature code	Description	Order type
EPDX	1 core Base Processor Activation (Pools 2.0) for EDP4 Linux only	Both
ELCL	Power Linux processor activation for #EDP2	Both
ELCM	Power Linux processor activation for #EDP4	Both
ELCQ	Power Linux processor activation for #EDP3	Both

1.5.3 Memory features

This section describes the memory features that available on a Power E1080 server. Careful selection of these features helps the user to configure their system with the correct amount of memory that can meet the demands of memory intensive workloads. On a Power E1080 server, the memory features can be classified into the following feature categories:

► Physical memory
► Memory activation

These features are described next.

Physical memory features

Physical memory features that are supported on Power E1080 are the next generation differential dual inline memory modules, called DDIMM (see 2.3, "Memory subsystem" on page 72). DDIMMS that are used in the E1080 are Enterprise Class 4U DDIMMs.

The memory DDIMM features are available in 32-, 64-, 128-, and 256-GB capacity. Among these DDIMM features, 32 GB and 64 GB DDIMMs run at 3200 MHz frequency and 128 GB and 256 GB DDIMMs run at 2933 MHz frequency.

Each system node provides 64 DDIMM slots that support a maximum of 16 TB memory and a four system node E1080 can support a maximum of 64 TB memory. DDIMMs are ordered by using memory feature codes, which include a bundle of four DDIMMs with the same capacity.

Consider the following points regarding improved performance:

► Plugging DDIMMs of same density provides the highest performance.
► Filling all the memory slots provides maximum memory performance.
► System performance improves when more quads of memory DDIMMs match.
► System performance also improves as the amount of memory is spread across more DDIMM slots.

  For example, if 1TB of memory is required, 64 x 32 GB DDIMMs can provide better performance than 32 x 64 GB DDIMMs

Figure 1-6 shows a DDIMM memory feature.
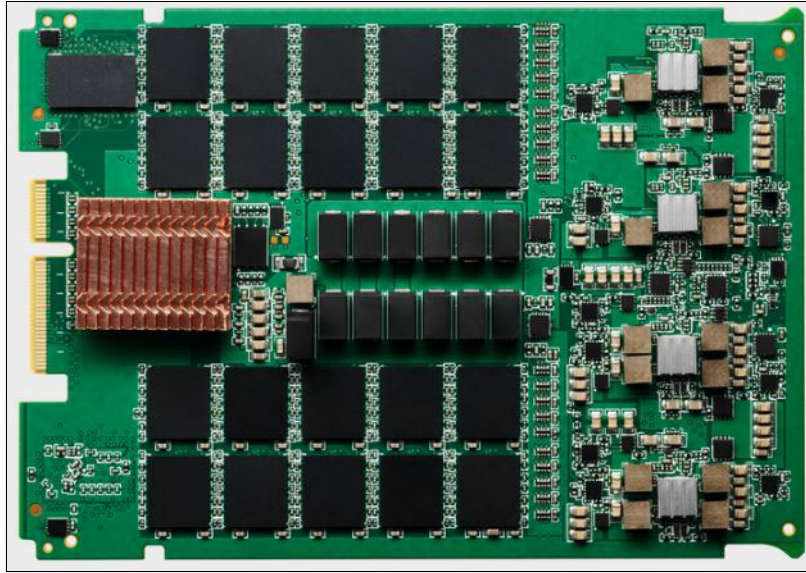
Figure 1-6   New DDIMM feature

Table 1-10 lists the available memory DDIMM feature codes for the Power E1080.

Table 1-10   E1080 memory feature codes

Feature code	CCIN	Description	OS support
EMC1	32AB	128 GB (4x32 GB) DDIMMs, 3200 MHz, 16 GBIT DDR4 Memory	AIX, IBM i, and Linux
EMC2	32AC	256 GB (4x64 GB) DDIMMs, 3200 MHz, 16 GBIT DDR4 Memory	AIX, IBM i, and Linux
EMC3	32AD	512 GB (4x128 GB) DDIMMs, 2933 MHz, 16 GBIT DDR4	AIX, IBM i, and Linux
EMC4	32AE	1 TB (4x256 GB) DDIMMs, 2933 MHz, 16 GBIT DDR4	AIX, IBM i, and Linux

Memory activation features

Software keys are required to activate part or all of the physical memory that is installed in the Power E1080 to be assigned to logical partitions (LPAR). Any software key is available when memory activation feature is ordered and can be ordered at any time during the life-cycle of the server to help the user scale up memory capacity without outages, unless an more physical memory upgrade and activation are required.

A server administrator or user cannot control which physical memory DDIMM features are activated when memory activations are used.

The amount of memory to activate depends on the feature code ordered; for example, if an order contains two of feature code EDAB (100 GB DDR4 Mobile Memory Activation for HEX), these feature codes activate 200 GB of the installed physical memory.

Different types of memory activation features that are available for the Power E1080 server are known to the PowerVM hypervisor as total quantity for each type. The PowerVM hypervisor determines the physical DDIMM memory to be activated and assigned to the LPARs.

Similar to processor core activation features, different types of permanent memory activation features are offered on the Power E1080 server. For more information about the available types of activations, 1.5.2, "Processor features" on page 16.

Orders for memory activation features must consider the following rules:

► The system must have a minimum of 50% activated physical memory. It can be activated by using static or static and mobile memory activation features.

► The system must have a minimum of 25% of physical memory activated by using static memory activation features.

► When a Power E1080 is part of a PEP 2.0 environment, the server must have a minimum of 256 GB of base memory activations.

Consider the following examples:

► For a system with 4 TB of physical memory, at least 2 TB (50% of 4 TB) must be activated.

► When a Power E1080 is part of a PEP 1.0 environment, a server with 4 TB of physical memory and 3.5 TB of activated memory requires a minimum of 896 GB (25% of 3.5 TB) of physical memory activated by using static activation features.

► When a Power E1080 is part of a PEP 2.0 environment, a server with 4 TB of physical memory requires a minimum of 256 GB of memory activated with base activation features.

Table 1-11 lists the available memory activation feature codes for Power E1080. The Order type column provides indicates whether the feature code is available for an initial order only, or also with a MES upgrade on a existing server only, or both.

Table 1-11   Memory activation features.

Feature code	Description	Order type
EDAB	100 GB DDR4 Mobile Memory Activation for HEX	Both
EDAG	256 GB Base Memory Activation (Pools 2.0)	Both
EDAH	512 GB Base Memory Activation (Pools 2.0)	Both
EDAL	256 GB Base Memory Activation Linux only	Both
EDAM	512 GB Base Memory Activation Linux only	Both
EDAP	1 GB Base Memory activation (Pools 2.0) from Static	Both
EDAQ	100 GB Base Memory activation (Pools 2.0) from Static	Both
EDAR	512 GB Base Memory activation (Pools 2.0) from Static	Both
EDAS	500 GB Base Memory activation (Pools 2.0) from Static	Both
EDAT	1 GB Base Memory activation (Pools 2.0) MES only	Both
EDAU	100 GB Base Memory activation (Pools 2.0) MES only	Both
EDAV	100 GB Base Memory Activation (Pools 2.0) from Mobile	Both
EDAW	500 GB Base Memory Activation (Pools 2.0) from Mobile	Both

Feature code	Description	Order type
EDAX	512 GB Base Memory Activation Linux only - Conversion	Both
ELME	512 GB Power Linux Memory Activations for HEX	Both
EMAZ	1 GB Memory activation for HEX	Both
EMBK	500 GB DDR4 Mobile Memory Activation for HEX/80H	Both
EMBZ	512 GB Memory Activations for HEX	Both
EMQZ	100 GB of #EMAZ Memory activation for HEX	Both

1.5.4 System node PCIe features

Each system node provides eight PCIe Gen 5 hot-plug enabled slots; therefore, a two-system nodes server provides 16 slots, a three-system nodes server provides 24 slots, and a four-system nodes system provides 32 slots.

Table 1-12 lists all the supported PCIe adapter feature codes inside the Power E1080 server node drawer.

Table 1-12   PCIe adapters supported on Power E1080 server node

Feature code	CCIN	Description	OS support
EN1A	578F	PCIe Gen 3 32 Gb 2-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN1B	578F	PCIe Gen 3 LP 32 Gb 2-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN1C	578E	PCIe Gen 3 16 Gb 4-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN1D	578E	PCIe Gen 3 LP 16 Gb 4-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN1F	579A	PCIe Gen 3 LP 16 Gb 4-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN1H	579B	PCIe Gen 3 LP 2-Port 16 Gb Fibre Channel Adapter	AIX, IBM i, and Linux
EN1K	579C	PCIe Gen 4 LP 32 Gb 2-port Optical Fibre Channel Adapter	AIX, IBM i, and Linux
EN2A	579D	PCIe Gen 3 16 Gb 2-port Fibre Channel Adapter	AIX, IBM i, and Linux
EN2B	579D	PCIe Gen 3 LP 16 Gb 2-port Fibre Channel Adapter	AIX, IBM i, and Linux
5260	576F	PCIe2 LP 4-port 1 GbE Adapter	AIX, IBM i, and Linux
5899	576F	PCIe2 4-port 1 GbE Adapter	AIX, IBM i, and Linux
EC2T	58FB	PCIe Gen 3 LP 2-Port 25/10 Gb NIC&ROCE SR/Cu Adapter[a]	AIX, IBM i, and Linux

Feature code	CCIN	Description	OS support
EC2U	58FB	PCIe Gen 3 2-Port 25/10 Gb NIC&ROCE SR/Cu Adapter[a]	AIX, IBM i, and Linux
EC67	2CF3	PCIe Gen 4 LP 2-port 100 Gb ROCE EN LP adapter	AIX, IBM i, and Linux
EN0S	2CC3	PCIe2 4-Port (10 Gb+1 GbE) SR+RJ45 Adapter	AIX, IBM i, and Linux
EN0T	2CC3	PCIe2 LP 4-Port (10 Gb+1 GbE) SR+RJ45 Adapter	AIX, IBM i, and Linux
EN0X	2CC4	PCIe2 LP 2-port 10/1 GbE BaseT RJ45 Adapter	AIX, IBM i, and Linux

a. Requires SFP to provide 10 Gb, 2 Gb, or 1 Gb BaseT connectivity

1.5.5 System node disk and media features

At the time of this writing, the Power E1080 server node supports up to four 7 mm NVMe U.2 drives that are plugged into the 4-bay NVMe carrier backplane (feature code EJBC). Each system node requires one backplane, even if no NVMe U.2 drives are selected.

Each NVMe U.2 drive can be independently assigned to different LPARs for hosting the operating system and to start from them. They also can be used for non-data intensive workloads. NVMe U.2 drives are concurrently replaceable.

Table 1-13 lists the available NVMe drive feature codes for the Power E1080 and the operating system support.

Table 1-13   NVMe features

Feature code	CCIN	Description	OS support
EC5J	59B4	Mainstream 800 GB SSD NVMe U.2 module	AIX and Linux
EC5K	59B5	Mainstream 1.6 TB SSD NVMe U.2 module	AIX and Linux
EC5L	59B6	Mainstream 3.2 TB SSD NVMe U.2 module	AIX and Linux
EC7Q	59B4	800 GB Mainstream NVMe U.2 SSD 4k for AIX/Linux	AIX and Linux

For more information, see 2.5.2, "Internal NVMe storage subsystem" on page 87.

For systems that are running IBM i, an expansion or storage drawer can meet the NVMe requirements.

1.5.6 System node USB features

The Power E1080 supports one stand-alone external USB drive that is associated to feature code EUA5. The feature code includes the cable that is used to the USB drive to the preferred front accessible USB port on the SCU.

The Power E1080 server node does not offer an integrated USB port. The USB 3.0 adapter feature code EC6J is required to provide connectivity to an optional external USB DVD drive and requires one system node or I/O expansion drawer PCIe slot. The adapter connects to

the USB port in the rear of the SCU with the cable associated to feature code EC6N. Because this cable is 1.5 m long, in case of a Power E1080 with more than one system node, the USB 3.0 adapter can be used in the first or the second system node only.

The USB 3.0 adapter feature code EC6J supports the assignment to an LPAR and can be migrated from an operating LPAR to another, including the connected DVD drive. This design allows it to assign the DVD drive feature to any LPAR according to the need.

Dynamic allocation of system resources such as processor, memory, and I/O is also referred to as *dynamic LPAR* or *DLPAR*.

For more information about the USB subsystem, see 1.5.6, "System node USB features" on page 23.

1.5.7 Power supply features

Each Power E1080 server node has four 1950 W bulk power supply units that are operating at 240 V. These power supply unit features are a default configuration on every Power E1080 system node. The four units per system node do not have an associated feature code and are always auto-selected by the IBM configurator when a new configuration task is started.

Four power cords from the power distribution units (PDU) drive these power supplies, which connect to four C13/C14 type receptacles on the linecord conduit in the rear of the system. The power linecord conduit source power from the rear and connects to the power supply units in the front of the system.

The system design provides N+2 redundancy for system bulk power, which allows the system to continue operation with any two of the power supply units functioning. The failed units must remain in the system until new power supply units are available for replacement.

The power supply units are hot-swappable, which allows replacement of a failed unit without system interruption. The power supply units are placed in front of the system, which makes any necessary service that much easier.

Figure 1-7 shows the power supply units and their physical locations marked as E1, E2, E3, and E4 in the system.
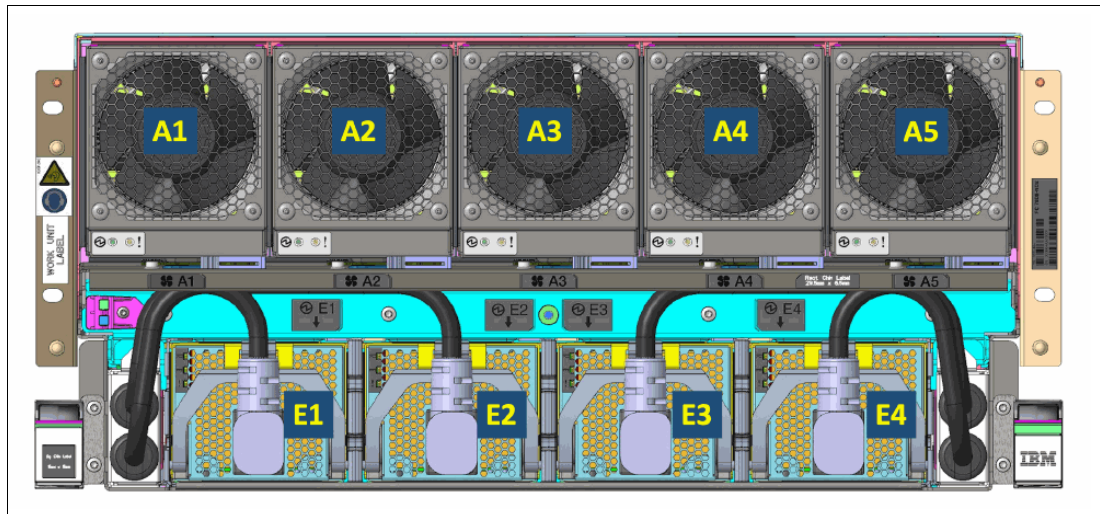
Figure 1-7   Power supply units

1.5.8 System node PCIe interconnect features

Each system node provides 8 PCIe Gen5 hot-plug enabled slots; therefore, a 2-node system provides 16 slots, a 3-node system provides 24 slots, and a 4-node system provides 32 slots.

Up to four I/O expansion drawer features #EMX0 can be connected per node to achieve the slot capacity that is listed in Table 1-14.

Table 1-14   PCIe slots availability for different system nodes configurations

System node #	I/O expansion #	LP slots	Full height slots
1	4	8	48
2	8	16	96
3	12	24	144
4	16	32	192

Each I/O expansion drawer consists of two Fanout Module feature #EMXH, each providing six PCIe slots. Each Fanout Module connects to the system by using a pair of CXP cable features. The CXP cable features are listed in Table 1-16 on page 27.

Table 1-15   Optical CXP cable feature

Feature code	CCIN	Description	OS support	Order type
ECCR		2 M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both
ECCY		10 M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both
ECCZ		20 M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both

The RPO-only cables in this list are not available for ordering new or MES upgrade, but for migrating from a source system. Select a longer length feature code for inter-rack connection between the system node and the expansion drawer.

The one pair of CXP optical cable connects to system node by using one 2-ports PCIe optical cable adapter feature EJ24, which is placed in the CEC.

Both CXP optical cable pair and the optical cable adapter features are concurrently maintainable. Therefore, careful balancing of I/O, assigning adapters through redundant EMX0 expansion drawers, and different system nodes can ensure high-availability for I/O resources that are assigned to partitions.

Note: At the time of this writing, feature EMXH is the only supported or orderable fanout module. Previous fanout module features EMXF and EMXG are not supported on Power E1080 system.

For more information abut internal buses and the architecture of internal and external I/O subsystems, see 2.5, "Internal I/O subsystem" on page 83.

1.6 I/O drawers

If more PCIe slots beyond the system node slots are required, the Power E1080 server supports adding I/O expansion drawers.

At initial availability zero, one, two, three, or four PCIe Gen 3 I/O Expansion Drawers per system node are supported. To connect an I/O expansion drawer, a PCIe slot is used to attach a 6-slot expansion module in the I/O drawer. A PCIe Gen 3 I/O Expansion Drawer (#EMX0) holds two expansion modules that are attached to any two PCIe slots in the same system node or in different system nodes.

For the connection of SAS disks, a disk-only I/O drawer is available. The EXP24SX is the only disk drawer that is supported.

1.6.1 PCIe Gen 3 I/O Expansion Drawer

The 19-inch 4 EIA (4U) PCIe Gen 3 I/O Expansion Drawer (#EMX0) and two PCIe Fanout Modules (#EMXH) provide 12 PCIe I/O full-length, full-height slots. One Fanout Module provides six PCIe slots that are labeled C1 - C6. C1 and C4 are x16 slots, and C2, IBM C3®, C5, and C6 are x8 slots. PCIe Gen1, Gen2, and Gen 3 full-high adapters are supported.

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

Concurrent repair and adding or removing PCIe adapters is done through HMC-guided menus or by operating system support utilities.

A PCIe CXP converter adapter and Active Optical Cables (AOCs) connect the system node to a PCIe Fanout Module in the I/O expansion drawer. Each PCIe Gen 3 I/O Expansion Drawer has two power supplies.

Drawers can be added to the server later, but system downtime must be scheduled for adding a PCIe Gen 3 Optical Cable Adapter or a PCIe Gen 3 I/O drawer (#EMX0) or Fanout Module.

Figure 1-8 shows a PCIe Gen 3 I/O Expansion Drawer.

Figure 1-8   PCIe Gen 3 I/O Expansion Drawer

The AOC cable feature codes are listed in Table 1-16. Also listed is the supported order type. The feature codes that are associated to cables that support RPO only are not available for new orders or MES upgrades. Instead, they are used to manage the migration of supported I/O expansion drawers from previous IBM POWER® technology-based servers to the Power E1080. Feature codes that are associated to cables with longer length are required to support inter-rack connection between the system node and I/O expansion drawer.

Table 1-16   Active Optical Cables feature codes

Feature code	Description	Order type
ECCR	2M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer	Both
ECCY	10M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer	Both
ECCZ	20M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer	Both

Careful balancing of I/O, assigning adapters through redundant EMX0 expansion drawers, and connectivity to different system nodes can ensure high-availability for I/O resources assigned to LPARs.

> **Note:** Consider the following points:
>
> ► Older PCIe Gen 3 Fanout Modules (#EMXG, #ELMF, and #ELMG) *cannot* be mixed with PCIe Gen 3 Fanout Module (feature #EMXH) in the same I/O Expansion Drawer.
>
> ► Older I/O Expansion Drawers *can* be mixed with PCIe Gen 3 Fanout Modules (#ELMF, #ELMG) and I/O Expansion Drawers with PCIe Gen 3 Fanout Modules #EMXH in same configuration.
>
> ► Older PCIe Gen 3 Fanout Modules (#ELMF, #ELMG) *can* be connect to older PCIe Gen 3 Optical Cable Adapters (#EJ05 or #EJ08).
>
> ► PCIe Gen 3 Optical Cable Adapters (#EJ20, or #EJ1R) requires the use of Optical Cables (#ECCR, #ECCX, ECCY, or #ECCZ, or copper cable #ECCS).

1.6.2 I/O drawers and usable PCIe slots

Figure 1-9 on page 28 shows the rear view of the PCIe Gen 3 I/O Expansion Drawer with the location codes for the PCIe adapter slots in the PCIe Gen 3 6-slot Fanout Module.
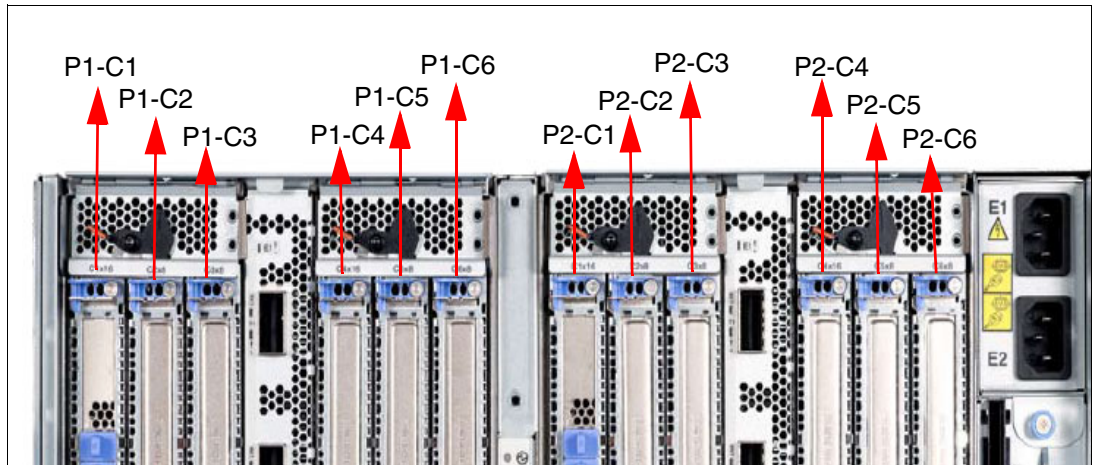
Figure 1-9   Rear view of a PCIe Gen 3 I/O Expansion Drawer with PCIe slots location codes

Table 1-17 lists the PCIe slots in the PCIe Gen 3 I/O Expansion Drawer.

Table 1-17   PCIe slot locations and descriptions for the PCIe Gen 3 I/O Expansion Drawer

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

Consider the following points regarding the information in Table 1-17:

► All slots support full-length, regular-height adapters or short (low-profile) adapters with a regular-height tailstock in single-wide, Gen 3 BSCs.

► Slots C1 and C4 in each PCIe Gen 3 6-slot Fanout Module are x16 PCIe Gen 3 buses; slots C2, C3, C5, and C6 are x8 PCIe buses.

► All slots support enhanced error handling (EEH).

► All PCIe slots are hot-swappable and support concurrent maintenance.

Table 1-18 lists the maximum number of I/O drawers that are supported and the total number of PCIe slots that are available when the expansion drawer consists of a single drawer type.

Table 1-18   Maximum number of supported I/O drawers and the total number of PCIe slots

System nodes	Maximum #EMX0 drawers	Total number of slots		
		PCIe Gen 3, x16	PCIe Gen 3, x8	Total PCIe Gen 3
One system node	4	16	32	48
Two system nodes	8	32	64	96
Three system nodes	12	48	96	144
Four system nodes	16	64	128	192

1.6.3 EXP24SX SAS Storage Enclosures

If you need more disks than are available with the internal disk bays, you can attach more external disk subsystems, such as an EXP24SX SAS Storage Enclosure (#ESLS).

The EXP24SX drawer is a storage expansion enclosure with 24 2.5-inch SFF SAS bays. It supports up to 24 hot-plug HDDs or SSDs in only 2 EIA of space in a 19-inch rack. The EXP24SX SFF bays use SFF Gen2 (SFF-2) carriers or trays.

With AIX/Linux/VIOS, the EXP24SX can be ordered with four sets of 6 bays (mode 4), two sets of 12 bays (mode 2) or one set of 24 bays (mode 1). With IBM i one set of 24 bays (mode 1) is supported. It is possible to change the mode setting in the field by using software commands along with a specifically documented procedure.

> **Important:** When changing modes, a skilled, technically qualified person must follow the special documented procedures. Improperly changing modes can destroy RAID sets, which prevent access to data, or allow other partitions to access another partition's data.

The attachment between the EXP24SX drawer and the PCIe Gen 3 SAS adapter is through SAS YO12 or X12 cables. The PCIe Gen 3 SAS adapters support 6 Gb throughput. The EXP24SX drawer can support up to 12 Gb throughput if future SAS adapters support that capability.

The EXP24SX drawer includes redundant AC power supplies and two power cords.

Figure 1-10 shows the EXP24SX drawer.

Figure 1-10   EXP24SX drawer

Note: For the EXP24SX drawer, a maximum of 24 2.5-inch SSDs or 2.5-inch HDDs are supported in the #ESLS 24 SAS bays. HDDs and SSDs cannot be mixed in the same mode-1 drawer. HDDs and SSDs can be mixed in a mode-2 or mode-4 drawer, but they cannot be mixed within a logical split of the drawer. For example, in a mode-2 drawer with two sets of 12 bays, one set can hold SSDs and one set can hold HDDs, but you cannot mix SSDs and HDDs in the same set of 12 bays.

For more information about SAS cabling and cabling configurations at IBM Documentation.

Table 1-19 lists the SFF-2 SSD and HDD feature codes that the Power E1080 supports in the expansion drawer at the time of this writing.

Table 1-19   Supported SFF-2 and HDD feature codes in the expansion drawer

Feature code	CCIN	Description	OS support
ES94	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ES95	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESB2	5B16	387 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ESB6	5B17	775 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ESBA	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESBB	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESBG	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESBH	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESBL	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESBM	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESGV	5B16	387 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ESGZ	5B17	775 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux

Feature code	CCIN	Description	OS support
ESJ0	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJ1	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJ2	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJ3	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJ4	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJ5	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJ6	5B2F	7.45 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJ7	5B2F	7.45 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJJ	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJK	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJL	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJM	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJN	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJP	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESJQ	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESJR	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESK1	5B16	387 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ESK3	5B17	775 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ESK8	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESK9	5B10	387 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESKC	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKD	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESKG	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKH	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESKK	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKM	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESKP	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKR	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESKT	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKV	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESKX	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESKZ	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESMB	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux

Feature code	CCIN	Description	OS support
ESMD	5B29	931 GB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESMF	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESMH	5B21	1.86 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESMK	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESMS	5B2D	3.72 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESMV	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESMX	5B2F	7.44 TB Mainstream SAS 4k SFF-2 SSD for IBM i	IBM i
ESNA	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESNB	5B11	775 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ESNE	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ESNF	5B12	1.55 TB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ETK1		387 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX and Linux
ETK3		775 GB Enterprise SAS 5xx SFF-2 SSD for AIX/Linux	AIX, and Linux
ETK8		387 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX, and Linux
ETK9		387 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ETKC		775 GB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ETKD		775 GB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
ETKG		1.55 TB Enterprise SAS 4k SFF-2 SSD for AIX/Linux	AIX and Linux
ETKH		1.55 TB Enterprise SAS 4k SFF-2 SSD for IBM i	IBM i
1953	19B1	300 GB 15k RPM SAS SFF-2 Disk Drive (AIX/Linux)	AIX and Linux
1964	19B3	600 GB 10k RPM SAS SFF-2 Disk Drive (AIX/Linux)	AIX and Linux
ESEU	59D2	571 GB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4224	IBM i
ESEV	59D2	600 GB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4096	AIX and Linux
ESF2	59DA	1.1 TB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4224	IBM i
ESF3	59DA	1.2 TB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4096	AIX and Linux
ESFS	59DD	1.7 TB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4224	IBM i
ESFT	59DD	1.8 TB 10K RPM SAS SFF-2 Disk Drive 4K Block - 4096	AIX and Linux
ESNL	5B43	283 GB 15K RPM SAS SFF-2 4k Block Cached Disk Drive (IBM i)	IBM i
ESNM	5B43	300 GB 15K RPM SAS SFF-2 4k Block Cached Disk Drive (AIX/Linux)	AIX and Linux
ESNQ	5B47	571 GB 15K RPM SAS SFF-2 4k Block Cached Disk Drive (IBM i)	IBM i
ESNR	5B47	600 GB 15K RPM SAS SFF-2 4k Block Cached Disk Drive (AIX/Linux)	AIX and Linux

1.6.4 IBM System Storage

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

IBM Storage simplifies data infrastructure by using an underlying software foundation to strengthen and streamline the storage in the hybrid cloud environment, which uses a simplified approach to containerization, management, and data protection. For more information about the various offerings, see this web page.

The following section highlights a few of the offerings.

IBM FlashSystem Family

The IBM FlashSystem® family is a portfolio of cloud-enabled storage systems designed to be easily deployed and quickly scaled to help optimize storage configurations, streamline issue resolution, and lower storage costs.

IBM FlashSystem is built with IBM Spectrum® Virtualize software to help deploy sophisticated hybrid cloud storage solutions, accelerate infrastructure modernization, address security needs, and maximize value by using the power of AI. The products are designed to provide enterprise-grade functions without compromising affordability or performance. They also offer the advantages of end-to-end NVMe, the innovation of IBM FlashCore® technology, and SCM for ultra-low latency. For more information, see this web page.

IBM System Storage DS8000

IBM DS8900F is the next generation of enterprise data systems that are built with the most advanced Power processor technology and feature ultra-low application response times.

Designed for data-intensive and mission-critical workloads, DS8900F adds next-level performance, data protection, resiliency, and availability across hybrid cloud solutions through ultra-low latency, better than seven 9's (99.99999) availability, transparent cloud, tiering, and advanced data protection against malware and ransomware. This enterprise class storage solution provides superior performance and higher capacity, which enables the consolidation of all mission critical workloads in one place.

IBM DS8900F can provide 100% data encryption at-rest, in-flight and in the cloud. This flexible storage supports IBM Power, IBM Z®, and IBM LinuxONE. For more information, see this web page.

IBM SAN Volume Controller

IBM SAN Volume Controller is an enterprise-class system that consolidates storage from over 500 IBM and third-party storage systems to improve efficiency, simplify management and operations, modernize storage with new capabilities, and enable a common approach to hybrid cloud regardless of storage system type.

IBM SAN Volume Controller provides a complete set of data resilience capabilities with high availability, business continuance, and data security features. Storage supports automated tiering with AI-based IBM Easy Tier® that can help improve performance at a lower cost. For more information, see this web page.

1.7 System racks

The Power E1080 server fits a standard 19-inch rack. The server is certified and tested in the IBM Enterprise racks (7965-S42, 7014-T42, 7014-T00, or 7965-94Y). Customers can choose to place the server in other racks if they are confident that those racks have the strength, rigidity, depth, and hole pattern characteristics that are needed. Contact IBM Support to determine whether other racks are suitable.

> **Order information:** It is highly recommended that you order the Power E1080 server with an IBM 42U enterprise rack 7965-S42 feature. This rack provides a more complete and higher quality environment for IBM Manufacturing system assembly and testing, and provides a complete package.

If a system is installed in a rack or cabinet that is not from IBM, ensure that the rack meets the requirements that are described in 1.7.7, "Original equipment manufacturer racks" on page 41.

> **Responsibility:** The customer is responsible for ensuring the installation of the drawer in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the drawer requirements for power, cooling, cable management, weight, and rail security.

1.7.1 New rack considerations

Consider the following points when racks are ordered:

► The new IBM Enterprise 42U Slim Rack 7965-S42 offers 42 EIA units (U) of space in a slim footprint.

► The 7014-T42, 7014-T00, and 7965-94Y racks are no longer available to purchase with a Power E1080 server. Installing a Power E1080 server in these racks is still supported.

> **Vertical PDUs:** All PDUs that are installed in a rack that contains a Power E1080 server must be installed horizontally to allow for cable routing in the sides of the rack.

1.7.2 IBM Enterprise 42U Slim Rack 7965-S42

The 2.0-meter (79-inch) Model 7965-S42 is compatible with past and present IBM Power servers and provides an excellent 19-inch rack enclosure for your data center. Its 600 mm (23.6 in.) width combined with its 1100 mm (43.3 in.) depth plus its 42 EIA enclosure capacity provides great footprint efficiency for your systems. It can be placed easily on standard 24-inch floor tiles.

Compared to the 7965-94Y Slim Rack, the Enterprise Slim Rack provides extra strength and shipping and installation flexibility.

The 7965-S42 rack includes space for up to four PDUs in side pockets. Extra PDUs beyond four are mounted horizontally and each uses 1U of rack space.

The Enterprise Slim Rack front door, which can be Basic Black/Flat (#ECRM) or High-End appearance (#ECRT) has perforated steel, which provides ventilation, physical security, and visibility of indicator lights in the installed equipment within.

Standard is a lock that is identical to the locks in the rear doors. The door (#ECRG) can be hinged on the left or right side.

In addition to the #ECRT door, you can order the #ECRF high-end appearance door (Figure 1-11)

Figure 1-11   The #ECRF and the #ECRT rack doors.

Orientation: #ECRT must not be flipped because the IBM logo would be upside down.

1.7.3 AC power distribution unit and rack content

The Power E1080 servers that are integrated into a rack at the factory feature PDUs that are mounted horizontally in the rack. Each PDU takes 1U of space in the rack. Mounting the PDUs vertically in the side of the rack can cause cable routing issues and interfere with optimal service access.

Two possible PDU ratings are supported: 60A/63A (orderable in most countries) and 30A/32A. Consider the following points:

► The 60A/63A PDU supports four system node power supplies and one I/O expansion drawer or eight I/O expansion drawers.

► The 30A/32A PDU supports two system node power supplies and one I/O expansion drawer or four I/O expansion drawers.

Rack-integrated system orders require at least two of #7109, #7188, or #7196.

High-function PDUs provide more electrical power per PDU and offer better "PDU footprint" efficiency. In addition, they are intelligent PDUs that provide insight to power usage by receptacle and remote power on and off capability for easier support by individual receptacle. The new PDUs are orderable as #ECJJ, #ECJL, #ECJN, and #ECJQ.

High-function PDU FCs are listed in Table 1-20.

Table 1-20   Available high-function PDUs

PDUs	1-phase or 3-phase depending on country wiring standards	3-phase 208 V depending on country wiring standards
Nine C19 receptacles	ECJJ	ECJL
Twelve C13 receptacles	ECJN	ECJQ

In addition, the following high-function PDUs are available:

► High Function 9xC19 PDU plus (#ECJJ)

   This intelligent, switched 200-240 volt AC PDU includes nine C19 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, which makes the nine C19 receptacles easily accessible.

► High Function 9xC19 PDU plus 3-Phase (#ECJL)

   This intelligent, switched 208 volt 3-phase AC PDU includes nine C19 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, which makes the nine C19 receptacles easily accessible.

► High Function 12xC13 PDU plus (#ECJN)

   This intelligent, switched 200-240 volt AC PDU includes 12 C13 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, which makes the 12 C13 receptacles easily accessible.

► High Function 12xC13 PDU plus 3-Phase (#ECJQ)

   This intelligent, switched 208 volt 3-phase AC PDU includes 12 C13 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, which makes the 12 C13 receptacles easily accessible.

Two or more additional PDUs can be installed horizontally in the rear of the rack. Mounting PDUs horizontally uses 1U per PDU and reduces the space that is available for other racked components. When mounting PDUs horizontally, the preferred approach is to use fillers in the EIA units that are occupied by these PDUs to facilitate proper air-flow and ventilation in the rack.

Each PDU requires one PDU-to-wall power cord. Various power cord features are available for various countries and applications by varying the PDU-to-wall power cord, which must be ordered separately.

Each power cord provides the unique design characteristics for the specific power requirements. To match new power requirements and save previous investments, these power cords can be requested with an initial order of the rack or with a later upgrade of the rack features.

Table 1-21 lists the available wall power cord options for the 7188 and High Function PDUs and iPDU features, which must be ordered separately.

Table 1-21   Wall power cord options for the PDU and iPDU features

Feature code	Wall plug	Rated voltage (V AC)	Phase	Rated amperage	Geography
6653	IEC 309, 3P+N+G, 16 A	230	3	16 amps/phase	Internationally available
6489	IEC309 3P+N+G, 32 A	230	3	32 amps/phase	EMEA
6654	NEMA L6-30	200 - 208, 240	1	24 amps	US, Canada, LA, and Japan
6655	RS 3750DP (watertight)	200 - 208, 240	1	24 amps	US, Canada, LA, and Japan
6656	IEC 309, P+N+G, 32 A	230	1	24 amps	EMEA
6657	PDL	230-240	1	32 amps	Australia, New Zealand
6667	PDL	380-415	3	32 amps	Austraiia, New Zealand
6658	Korean plug	220	1	30 amps	North and South Korea
6492	IEC 309, 2P+G, 60 A	200 - 208, 240	1	48 amps	US, Canada, LA, and Japan
6491	IEC 309, P+N+G, 63 A	230	1	63 amps	EMEA

Notes: Ensure that the suitable power cord feature is configured to support the power that is being supplied. Based on the power cord that is used, the PDU can supply 4.8 - 19.2 kVA. The power of all the drawers that are plugged into the PDU must not exceed the power cord limitation.

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

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

For more information about power requirements of and the power cord for the 7965-94Y rack, see IBM Documentation.

1.7.4 PDU connection limits

Two possible PDU ratings are supported: 60/63 amps and 30/32 amps. The PDU rating is determined by the power cord that is used to connect the PDU to the electrical supply. The number of system nodes and I/O expansion drawers that are supported by each power cord are listed in Table 1-22.

Table 1-22   Maximum supported enclosures by power cord

Feature code	Wall plug	PDU Rating	Maximum supported system nodes per PDU pair	Maximum supported I/O drawers with no system nodes
6653	IEC 309, 3P+N+G, 16 A	60 Amps	Two system nodes and 1 I/O expansion drawer	8
6489	IEC309 3P+N+G, 32 A	60 Amps	Two system nodes and 1 I/O expansion drawer	8
6654	NEMA L6-30	30 Amps	One system node and 1 I/O expansion drawer	4
6655	RS 3750DP (watertight)	30 Amps	One system node and 1 I/O expansion drawer	4
6656	IEC 309, P+N+G, 32 A	30 Amps	One system node and 1 I/O expansion drawer	4
6657	PDL	30 Amps	One system node and 1 I/O expansion drawer	4
6658	Korean plug	30 Amps	One system node and 1 I/O expansion drawer	4
6492	IEC 309, 2P+G, 60 A	60 Amps	Two system node and 1 I/O expansion drawer	4
6491	IEC 309, P+N+G, 63 A	60 Amps	Two system nodes and 1 I/O expansion drawer	8

1.7.5 Rack-mounting rules

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

► For rack stability, start filling the rack from the bottom.

► IBM recommends the use of an IBM approved lift tool for installation of systems into any IBM or non-IBM rack.

► IBM does not support installation of the server nodes higher than the 29U position.

► Any remaining space in the rack can be used to install other systems or peripheral devices. Ensure that the maximum permissible weight of the rack is not exceeded and the installation rules for these devices are followed.

► Before placing the system into the service position, follow the rack manufacturer's safety instructions regarding rack stability.

1.7.6 Useful rack additions

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

IBM System Storage 7226 Model 1U3 Multi-Media Enclosure

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

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

- ► IBM POWER6 processor-based systems
- ► IBM POWER7 processor-based systems
- ► IBM POWER8® processor-based systems
- ► IBM POWER9 processor-based systems
- ► IBM Power10 processor-based systems

The IBM System Storage 7226 Multi-Media Enclosure offers an expansive list of drive feature options, as listed in Table 1-23.

Table 1-23   Supported drive features for the 7226-1U3

Feature code	Description	Status
1420	DVD-RAM SAS Optical Drive	Available
1422	DVD-RAM Slim SAS Optical Drive	Available
5762	DVD-RAM USB Optical Drive	Available
5763	DVD Front USB Port Sled with DVD-RAM USB Drive	Available
5757	DVD RAM Slim USB Optical Drive	Available
8348	LTO Ultrium 6 Half High Fibre Tape Drive	Available
8341	LTO Ultrium 6 Half High SAS Tape Drive	Available
8441	LTO Ultrium 7 Half High SAS Tape Drive	Available
8546	LTO Ultrium 8 Half High Fibre Tape Drive	Available
EU03	RDX 3.0 Removable Disk Docking Station	Available

The following options are available:

- ► LTO Ultrium 6 Half-High 2.5 TB SAS and FC Tape Drive: With a data transfer rate up to 320 MBps (assuming a 2.5:1 compression), the LTO Ultrium 6 drive is read/write compatible with LTO Ultrium 6 and 5 media, and read-only compatibility with LTO Ultrium 4. By using data compression, an LTO-6 cartridge can store up to 6.25 TB of data.

- ► The LTO Ultrium 7 drive offers a data rate of up to 300 MBps with compression. It also provides read/write compatibility with Ultrium 7 and Ultrium 6 media formats, and read-only compatibility with Ultrium 5 media formats. By using data compression, an LTO-7 cartridge can store up to 15TB of data.

- ► The LTO Ultrium 8 drive offers a data rate of up to 300 MBps with compression. It also provides read/write compatibility with Ultrium 8 and Ultrium 7 media formats. It is not read or write compatible with other Ultrium media formats. By using data compression, an LTO-8 cartridge can store up to 30 TB of data.

► DVD-RAM: The 9.4 GB SAS Slim Optical Drive with an SAS and USB interface option is compatible with most standard DVD disks.

► RDX removable disk drives: The RDX USB docking station is compatible with most RDX removable disk drive cartridges when it is used in the same OS. The 7226 offers the following RDX removable drive capacity options:

  – 500 GB (#1107)
  – 1.0 TB (#EU01)
  – 2.0 TB (#EU2T)

Removable RDX drives are in a rugged cartridge that inserts in to an RDX removable (USB) disk docking station (#1103 or #EU03). RDX drives are compatible with docking stations, which are installed internally in POWER8, POWER9, and Power10 processor-based servers, where applicable.

Figure 1-12 shows the IBM System Storage 7226 Multi-Media Enclosure.

Figure 1-12   IBM System Storage 7226 Multi-Media Enclosure

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

The IBM System Storage 7226 Multi-Media Enclosure is compatible with most POWER8, POWER9, and Power10 processor-based systems that offer current level AIX, IBM i, and Linux operating systems.

> **Unsupported:** IBM i does not support 7226 USB devices.

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

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

Flat panel display options

The IBM 7316 Model TF5 is a rack-mountable flat panel console kit that can also be configured with the tray pulled forward and the monitor folded up, which provides full viewing and keying capability for the HMC operator.

The Model TF5 is a follow-on product to the Model TF4 and offers the following features:

► A slim, sleek, and lightweight monitor design that occupies only 1U (1.75 in.) in a 19-inch standard rack

► A 18.5-inch (409.8 mm x 230.4 mm) flat panel TFT monitor with truly accurate images and virtually no distortion

▶ The ability to mount the IBM Travel Keyboard in the 7316-TF5 rack keyboard tray

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

The #4283 is a 1x8 Console Switch that fits in the 1U space behind the TF5 It is a CAT5-based switch. It contains eight analog rack interface (ARI) ports for connecting PS/2 or USB console switch cables. It supports chaining of servers that use an IBM Conversion Options switch cable (#4269). This feature provides four cables that connect a KVM switch to a system, or can be used in a daisy-chain scenario to connect up to 128 systems to a single KVM switch. It also supports server-side USB attachments

1.7.7 Original equipment manufacturer racks

The system can be installed in a suitable OEM rack if that the rack conforms to the EIA-310-D standard for 19-inch racks. This standard is published by the Electrical Industries Alliance. For more information, see IBM Documentation.

The IBM Documentation provides the general rack specifications, including the following information:

▶ The rack or cabinet must meet the EIA Standard EIA-310-D for 19-inch racks that was published August 24, 1992. The EIA-310-D standard specifies internal dimensions, for example, the width of the rack opening (width of the chassis), the width of the module mounting flanges, and the mounting hole spacing.

▶ The front rack opening must be a minimum of 450 mm (17.72 in.) wide, and the rail-mounting holes must be 465 mm plus or minus 1.6 mm (18.3 in. plus or minus 0.06 in.) apart on center (horizontal width between vertical columns of holes on the two front-mounting flanges and on the two rear-mounting flanges).

Figure 1-13 is a top view showing the rack specification dimensions.
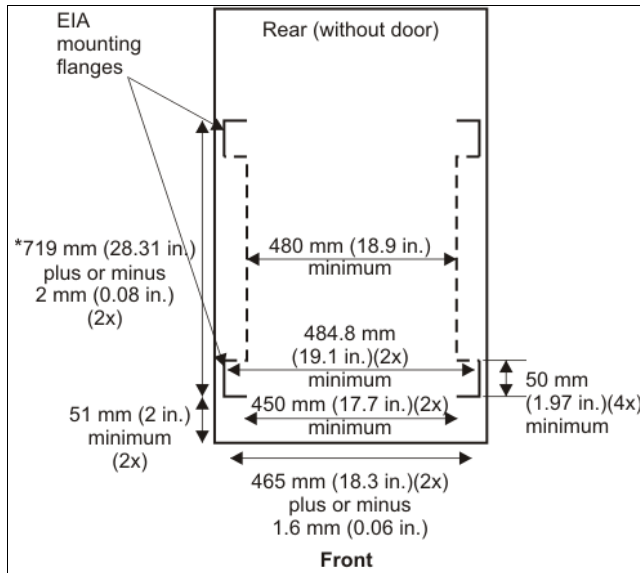
Figure 1-13   Rack specifications (top-down view)

▶ The vertical distance between mounting holes must consist of sets of three holes that are spaced (from bottom to top) 15.9 mm (0.625 in.), 15.9 mm (0.625 in.), and 12.7 mm (0.5 in.) on center, which makes each three-hole set of vertical hole spacing 44.45 mm (1.75 in.) apart on center.

Figure 1-14 shows the vertical distances between the mounting holes.
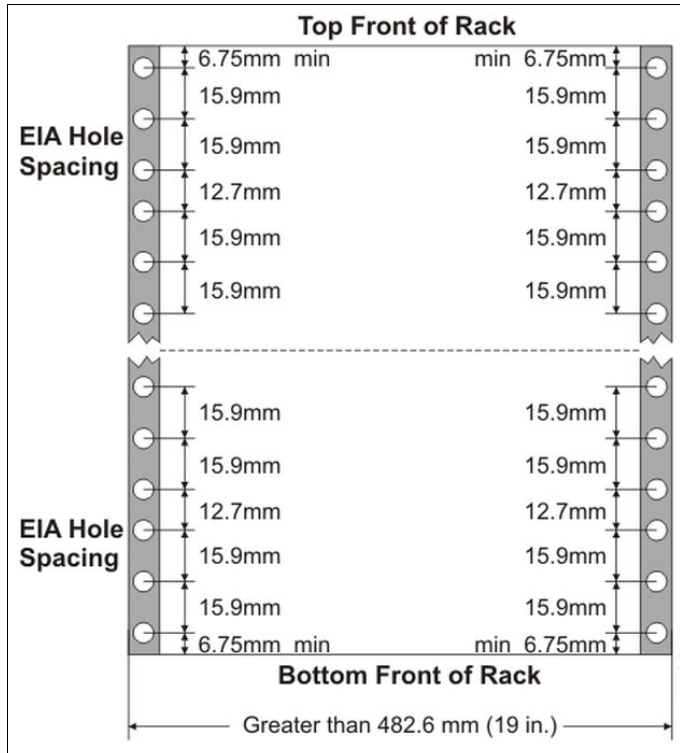
Figure 1-14   Vertical distances between mounting holes

► The following rack hole sizes are supported for racks where IBM hardware is mounted:

   – 7.1 mm (0.28 in.) plus or minus 0.1 mm (round)
   – 9.5 mm (0.37 in.) plus or minus 0.1 mm (square)

The rack or cabinet must be capable of supporting an average load of 20 kg (44 lb.) of product weight per EIA unit. For example, a four EIA drawer has a maximum drawer weight of 80 kg (176 lb.).

1.8 Hardware management console overview

The Hardware Management Console (HMC) can be a hardware appliance or virtual appliance that can be used to configure and manage your systems. The HMC connects to one or more managed systems and provides capabilities for following primary functions:

► Systems Management functions, such as Power Off, Power on, system settings, Capacity on Demand, Enterprise Pools, Shared processor Pools, Performance and Capacity Monitoring, and starting Advanced System Management Interface (ASMI) for managed systems.

► Deliver virtualization management through support for creating, managing, and deleting Logical Partitions, Live Partition Mobility, Remote Restart, configuring SRIOV, managing Virtual IO Servers, dynamic resource allocation, and operating system terminals.

- Acts as the service focal point for systems and supports service functions, including call home, dump management, guided repair and verify, concurrent firmware updates for managed systems, and around-the-clock error reporting with Electronic Service Agent for faster support.
- Provides appliance management capabilities for configuring network, users on the HMC, updating and upgrading the HMC

We discuss the available HMC offerings next.

1.8.1 HMC 7063-CR2

The 7063-CR2 IBM Power Systems HMC (see Figure 1-15) is a second-generation Power processor-based HMC.

The CR2 model includes the following features:

- 6-core POWER9 130W processor chip
- 64 GB (4x16 GB) or 128 GB (4x32 GB) of memory RAM
- 1.8 TB with RAID1 protection of internal disk capacity
- 4-ports 1 Gb Ethernet (RH-45), 2-ports 10 Gb Ethernet (RJ-45), two USB 3.0 ports (front side) and two USB 3.0 ports (rear side), and 1 Gb IPMI Ethernet (RJ-45)
- Two 900W power supply units
- Remote Management Service: IPMI port (OpenBMC) and Redfish application programming interface (API)
- Base Warranty is 1 year 9x5 with available optional upgrades

A USB Smart Drive not included.

> **Note:** The recovery media for V10R1 is the same for 7063-CR2 and 7063-CR1.

Figure 1-15   HMC 7063-CR2

The 7063-CR2 is compatible with flat panel console kits 7316-TF3, TF4, and TF5.

> **Note:** The 7316-TF3 and TF4 were withdrawn from marketing.

1.8.2 Virtual HMC

Initially, the HMC was sold only as a hardware appliance, including the HMC firmware installed. However, IBM extended this offering to allow the purchase of the hardware appliance and a virtual appliance that can be deployed on ppc64le architectures and deployed on x86 platforms.

Any customer with valid contract can download it from ESS site or it can be included within an initial Power E1080 order.

The virtual HMC supports the following hypervisors:

► On x86 processor-based servers:
  – KVM
  – Xen
  – VMware

► On Power processor-based servers: PowerVM

The following minimum requirements must be met to install the virtual HMC:

► 16 GB of Memory
► 4 virtual processors
► 2 network interfaces (maximum 4 allowed)
► 1 disk drive (500 GB available disk drive)

For an initial Power E1080 order with the IBM configurator (e-config), HMC virtual appliance can be found by selecting **add software** → **Other System Offering**s (as product selections) and then:

► 5765-VHP for IBM HMC Virtual Appliance for Power V10
► 5765-VHX for IBM HMC Virtual Appliance x86 V10

For more information about an overview of the Virtual HMC, see this web page.

For more information about how to install the virtual HMC appliance and all requirements, see IBM Documentation.

1.8.3 Baseboard management controller network connectivity rules for 7063-CR2

The 7063-CR2 HMC features a baseboard management controller (BMC), which is a specialized service processor that monitors the physical state of the system by using sensors. OpenBMC that is used on 7063-CR2 provides a graphical user interface (GUI) that can be accessed from a workstation that has network connectivity to the BMC. This connection requires an Ethernet port to be configured for use by the BMC.

The 7063-CR2 provides two network interfaces (eth0 and eth1) for configuring network connectivity for BMC on the appliance.

Each interface maps to a different physical port on the system. Different management tools name the interfaces differently. The HMC task **Console Management** → **Console Settings** → **Change BMC/IPMI Network Settings** modifies only the Dedicated interface.

The BMC ports are listed see Table 1-24.

Table 1-24   BMC ports

Management Tool	Logical Port	Shared/Dedicated	CR2 Physical Port
OpenBMC UI	eth0	Shared	eth0
OpenBMC UI	eth1	Dedicated	Management port only
ipmitool	lan1	Shared	eth0
ipmitool	lan2	Dedicated	Management port only
HMC task (change BMC/IMPI Network settings	lan2	Dedicated	Management port only

Figure 1-16 shows the BMC interfaces of the HMC.

BMC interface: **eth0**
ipmitool name: **lan 1**
Type: **Shared**

BMC interface: **eth1**
ipmitool name: **lan 2**
Type: **Dedicated**
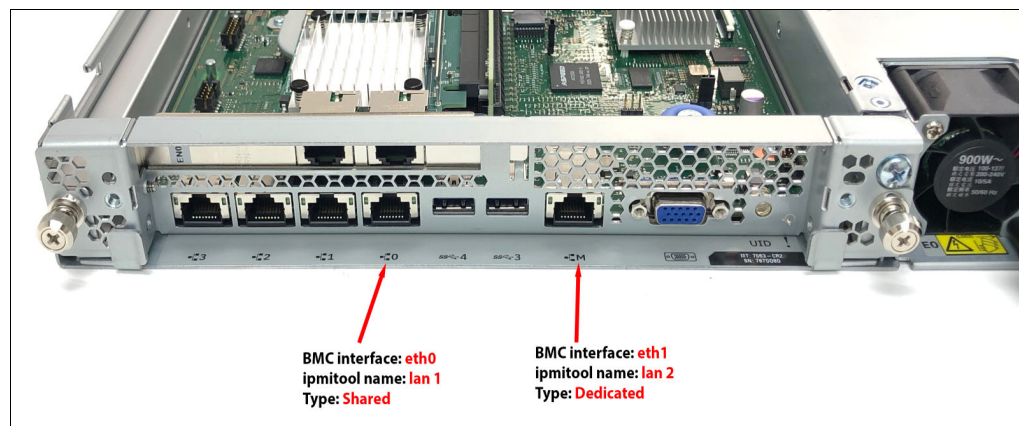
Figure 1-16   BMC interfaces

The main difference is the shared and dedicated interface to the BMC can coexist. Each has its own LAN number and physical port. Ideally, the customer configures one port, but both can be configured. Connecting power systems to the HMC rules remain the same as previous versions.

1.8.4 High availability HMC configuration

For the best manageabiltiy and redundancy, a dual HMC configuration is suggested. This configuration can be two hardware appliances, but also one hardware appliance and one virtual appliance or two virtual appliances.

The following requirements must be met:

► Two HMCs are at the same version.

► The HMCs use different subnets to connect to the FSPs.

► The HMCs can communicate with the servers' partitions over a public network to allow for full synchronization and function.

1.8.5 HMC code level requirements for the Power E1080

The minimum required HMC version for the Power E1080 is V10R1. V10R1 is supported only on 7063-CR1, 7063-CR2, and Virtual HMC appliances. It is *not* supported on 7042 Machine types. HMC with V10R1 cannot manage POWER7 processor-based systems.

An HMC that is running V10R1 M1010 includes the following features:

► HMC OS Secure Boot support for the 7063-CR2 appliance
► Ability to configure login retries and suspended time and support for inactivity expiration in password policy
► Ability to specify HMC location and data replication for groups
► VIOS Management Enhancements:
  – Prepare for VIOS Maintenance:
    • Validation for redundancy for the storage and network that is provided by VIOS to customer partitions
    • Switch path of redundant storage and network to start failover
    • Rollback to original configuration on failure of prepare
    • Audit various validation and prepare steps performed
    • Report any failure seen during prepare
  – Command Line and Scheduled operations support for VIOS backup or restore VIOS Configuration and SSP Configuration
  – Option to backup or restore Shared Storage Pool configuration in HMC
  – Options to import or export the backup files to external storage
  – Option to failover all Virtual NIC's from one VIOS to another
► Support 128 and 256 MB LMB sizes
► Automatically choose fastest network for LPM memory transfer
► HMC user experience enhancements:
  – Usability and performance improvements
  – Enhancements to help connect global search
  – Quick view of serviceable events
  – More progress information for UI Wizards
► Allow LPM/Remote Restart when virtual optical device is assigned to partition
► Update Access Key Support

► Scheduled operation function: In the Electronic Service Agent, a new feature that allows customers to receive message alerts only if scheduled operations fail (see Figure 1-17).
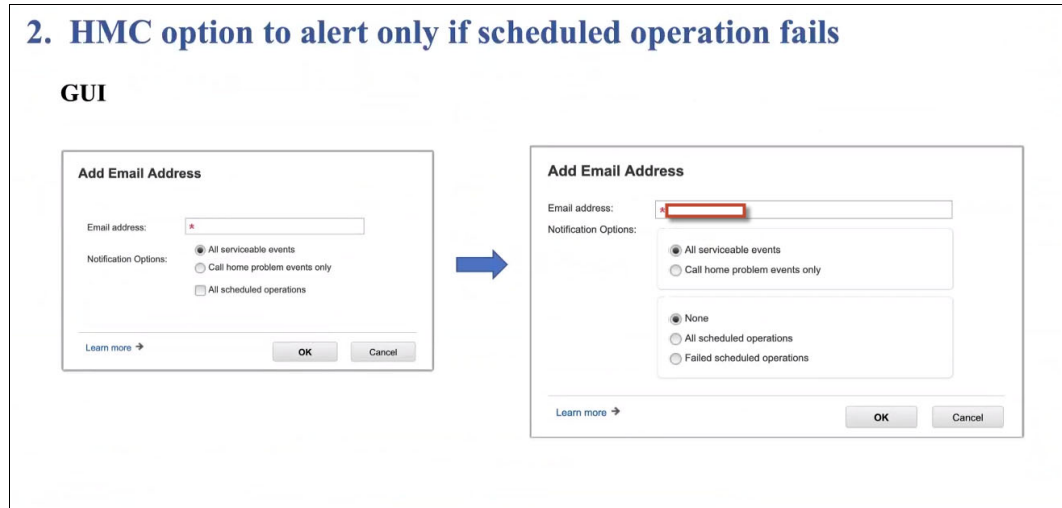
Figure 1-17   HMC alert feature

Log retention of the HMC audit trail also is increased.

1.8.6 HMC currency

In recent years, cybersecurity emerged as a national security issue and an increasingly critical concern for CIOs and enterprise IT managers.

The IBM Power processor-based architecture always ranked highly in terms of end-to-end security, which is why it remains a platform of choice for mission-critical enterprise workloads.

A key aspect of maintaining a secure Power environment is ensuring that the HMC (or virtual HMC) is current and fully supported (including hardware, software, and Power firmware updates).

Outdated or unsupported HMCs represent a technology risk that can quickly and easily be mitigated by upgrading to a current release.

2

Architecture and technical overview

This chapter describes the overall system architecture for the IBM Power E1080 (9080-HEX) server. The bandwidths that are provided throughout are theoretical maximums that are used for reference. The speeds that are shown are at an individual component level. Multiple components and application implementations are key to achieving the best performance.

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

This chapter includes the following topics:

2.1 IBM Power10 processor

Figure 2-1 shows the logical system architecture of the Power E1080 server.
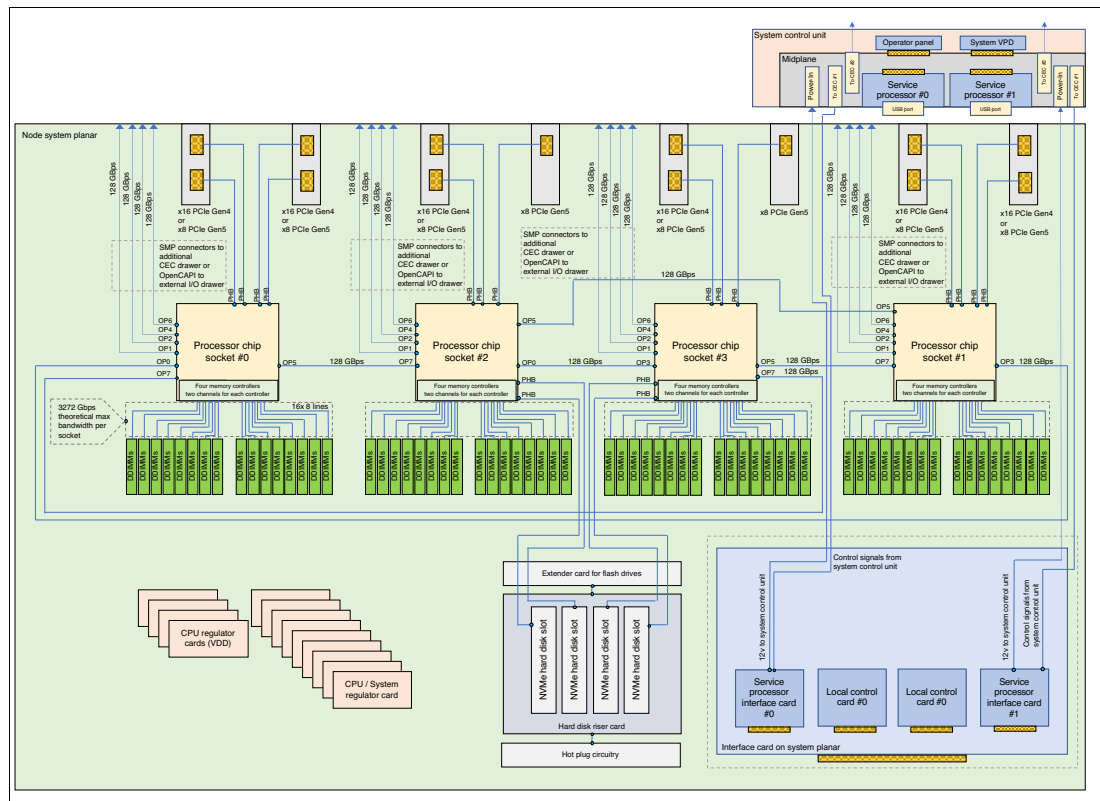
Figure 2-1   Power E1080 logical system

The IBM Power10 processor was introduced to the general public on August 17, 2020 at the 32nd HOT CHIPS[1] semiconductor conference. At that meeting, the new capabilities and features of the latest Power processor microarchitecture and the Power Instruction Set Architecture (ISA) v3.1B were revealed and categorized according to the following Power10 processor design priority focus areas:

► Data plane bandwidth focus area

Terabyte per second signaling bandwidth on processor functional interfaces, petabyte system memory capacities, 16-socket symmetric multiprocessing (SMP) scalability, and memory clustering and memory inception capability.

► Powerful enterprise core focus area

New core micro-architecture, flexibility, larger caches, and reduced latencies.

► End-to-end security focus area

Hardware enabled security features that are co-optimized with PowerVM hypervisor support.

► Energy efficiency focus area

Up to threefold energy efficiency improvement in comparison to POWER9 processor technology.

---

[1] https://hotchips.org/

▶ Artificial intelligence (AI)-infused core focus area

A 10-20x matrix-math performance improvement per socket compared to the POWER9 processor technology capability.

The remainder of this section provides more specific information about the Power10 processor technology as it is used in the Power E1080 scale-up enterprise class server.

The *IBM's Power10 Processor* session material as presented at the 32nd HOT CHIPS conference is available through the HC32 conference proceedings archive at this web page.

2.1.1 Power10 processor overview

The Power10 processor is a superscalar symmetric multiprocessor that is manufactured in complimentary metal-oxide-semiconductor (CMOS) 7 nm lithography with 18 layers of metal. The processor contains up to 15 cores that support eight simultaneous multithreading (SMT8) independent execution contexts.

Each core has private access to 2 MB L2 cache and local access to 8 MB of L3 cache capacity. The local L3 cache region of a specific core also is accessible from all other cores on the processor chip. The cores of one Power10 processor share up to 120 MB of latency optimized non-uniform cache access (NUCA) L3 cache.

The processor supports the following three distinct functional interfaces which all are capable to run with a signaling rate of up to 32 GTps[2]:

▶ Open memory interface

The Power10 processor has eight memory controller unit (MCU) channels that support one open memory interface (OMI) port with two OMI links each[3]. One OMI link aggregates 8 lanes running at 32 GTps speed and connects to one memory buffer based differential DIMM (DDIMM) slot to access main memory. Physically, the OMI interface is implemented in two separate die areas of 8 OMI links each. The maximum theoretical full-duplex bandwidth aggregated over all 128 OMI lanes is 1 TBps.

▶ SMP fabric interconnect (PowerAXON)

A total of 144 lanes are available in the Power10 processor to facilitate the connectivity to other processors in a symmetric multiprocessing (SMP) architecture configuration. Each SMP connection requires 18 lanes, eight data lanes plus one spare lane per direction (2 x(8+1)). In this way the processor can support a maximum of eight SMP connections with at total of 128 data lanes per processor. This configuration yields a maximum theoretical full-duplex bandwidth aggregated over all SMP connections of 1 TBps.

The generic nature of the interface implementation also allows the use of 128 data lanes to potentially connect accelerator or memory devices through the OpenCAPI protocols. Also, it can support memory cluster and memory interception architectures.

Because of the versatile characteristic of the technology, it is also referred to as *PowerAXON* interface (Power A-bus/X-bus/OpenCAPI/Networking[4]). The OpenCAPI and the memory clustering and memory interception use cases can be pursued in the future and are currently not use by available technology products.

---

[2] Giga transfers per second (GTps)

[3] The OMI links are also referred to as OMI sub-channels.

[4] A-busses and X-busses provide SMP fabric ports used between CEC drawers or within CEC drawers respectively.

► PCIe Version 5.0 interface

To support external I/O connectivity and access to internal storage devices, the Power10 processor provides differential Peripheral Component Interconnect Express version 5.0 interface busses (PCIe Gen 5) with a total of 32 lanes. The lanes are grouped in two sets of 16 lanes that can be used in one of the following configurations:

– 1 x16 PCIe Gen 4
– 2 x8 PCIe Gen 4
– 1 x8, 2 x4 PCIe Gen 4
– 1 x8 PCIe Gen 5, 1 x8 PCIe Gen 4
– 1 x8 PCIe Gen 5, 2 x4 PCIe Gen 4

Figure 2-2 shows the Power10 processor die with several functional units labeled. Note, 16 SMT8 processor cores are shown, but only 10-, 12-, or 15-core processor options are available for Power E1080 server configurations.
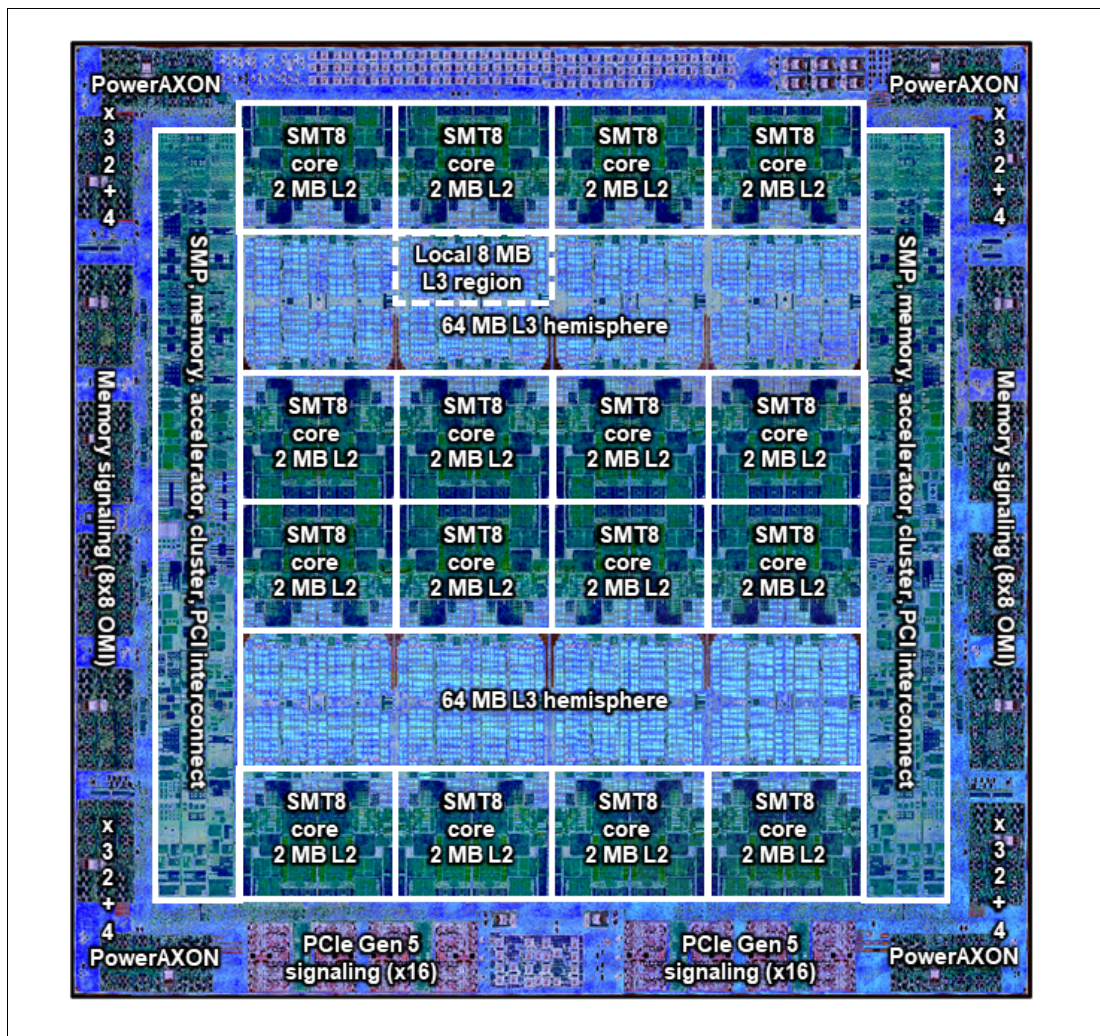
Figure 2-2   The Power10 processor chip (Die photo courtesy of Samsung Foundry)

Important Power10 processor characteristics are listed in [2-1].

Table 2-1   Summary of the Power10 processor chip and processor core technology

Technology	Power10 processor
Processor die size	602 mm^2
Fabrication technology	► CMOS[a] 7-nm lithography ► 18 layers of metal
Maximum processor cores per chip	15
Maximum execution threads per core / chip	8 / 120
Maximum L2 cache core	2 MB
Maximum On-chip L3 cache per core / chip	8 MB / 120 MB
Number of transistors	18 billion
Processor compatibility modes	Support for Power ISA[b] of POWER8 and POWER9

a. Complimentary metal-oxide-semiconductor (CMOS)
b. Power instruction set architecture (Power ISA)

The Power10 processor is packaged as single-chip module (SCM) for exclusive use in the Power E1080 servers. The SCM contains the Power10 processor plus more logic that is needed to facilitate power supply and external connectivity to the chip. It also holds the connectors to plug SMP cables directly onto the socket to build 2-, 3-, and 4-node Power E1080 servers.

Figure 2-3 shows the logical diagram of the Power10 SCM.
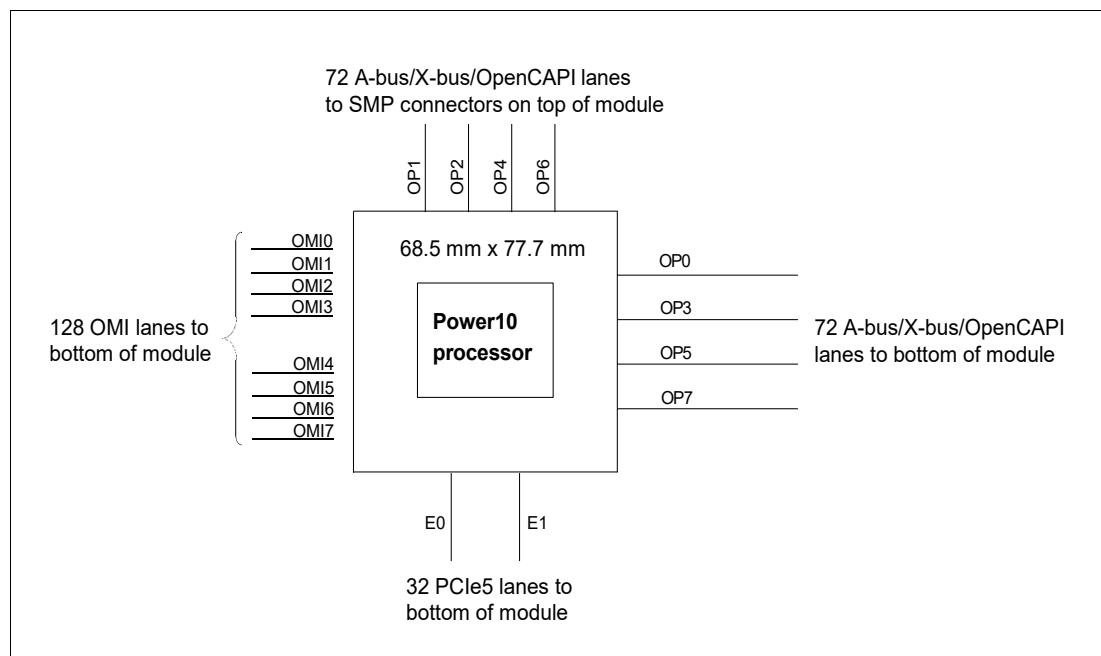
Figure 2-3   Power10 single chip module

As indicated in Figure 2-3 on page 53, the PowerAXON interface lanes are grouped in two sets of 72 lanes each. One set provides four interface ports (OP1, OP2, OP4, and OP6), which are accessible to SMP connectors that are physically placed on the top of the SCM module.

The second set of ports (OP0, OP3, OP5, and OP7) are used to implement the fully connected SMP fabric between the four sockets within a system node. Eight open memory interface ports (OMI0 to OMI7) with two OMI links each provide access to the buffered main memory differential DIMMs (DDIMMs). The 32 PCIe Gen 5 lanes are grouped in two PCIe host bridges (E0, E1).

Figure 2-4 shows a physical diagram of the Power10 SCM. The eight SMP connectors (OP1A, OP1B, OP2A, OP2B, OP4A, OP4B, OP6A, and OP6B) externalize 4 SMP busses, which are used to connect system node drawers in 2-, 3-, and 4-node Power E1080 configurations. The OpenCAPI connectivity options are also indicated, although they are not used by any commercially available product.
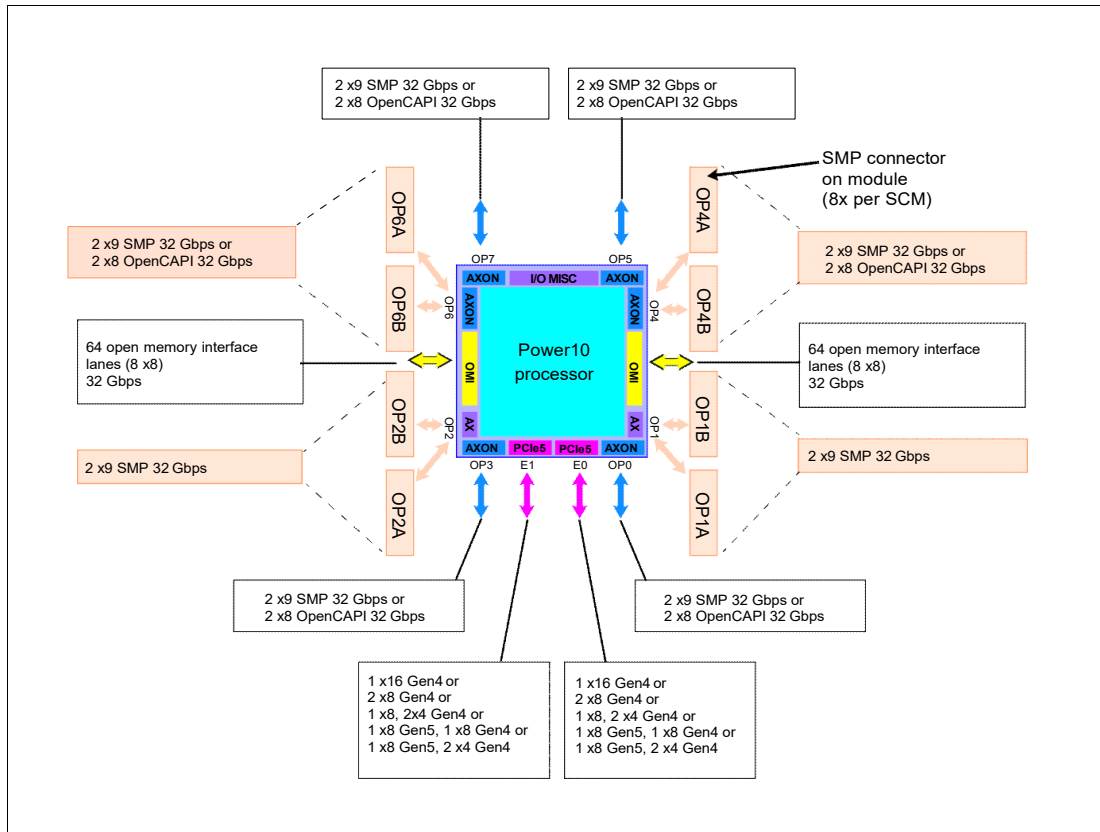
Figure 2-4   Power10 single chip module physical diagram

2.1.2 Power10 processor core

The Power10 processor core inherits the modular architecture of the POWER9 processor core, but the re-designed and enhanced micro-architecture significantly increases the processor core performance and processing efficiency. The peak computational throughput is markedly improved by new execution capabilities and optimized cache bandwidth characteristics. Extra matrix math acceleration engines can deliver significant performance gains for machine learning, particularly for AI inferencing workloads.

The Power E1080 server uses the Power10 enterprise-class processor variant in which each core can run with up to eight essentially independent hardware threads. If all threads are active, the mode of operation is referred to as 8-way simultaneous multithreading (SMT8) mode. A Power10 core with SMT8 capability is named Power10 SMT8 core or SMT8 core for short. The Power10 core also supports modes with four active threads (SMT4); that is, two active threads (SMT2) and one single active thread (ST).

The SMT8 core includes two execution resource domains. Each domain provides the functional units to service up to four hardware threads.

Figure 2-5 shows the functional units of a SMT8 core where all 8 threads are active. The two execution resource domains are highlighted with colored backgrounds in two different shades of blue.
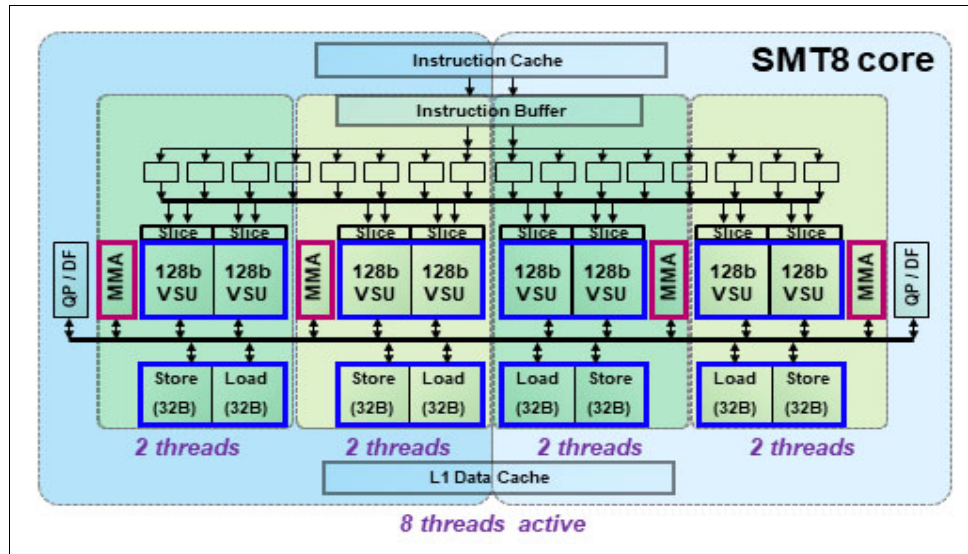
Figure 2-5   Power10 SMT8 core

Each of the two execution resource domains supports between one and four threads and includes 4 vector scalar units (VSU) of 128-bit width, two matrix-multiply assist (MMA) accelerators, and one quad-precision floating-point (QP) and decimal floating-point (DF) unit.

One VSU and the directly associated logic is called an execution *slice*. Two neighboring slices can also be used as a combined execution resource which is then named *super-slice*. When operating in SMT8 mode, eight SMT threads are subdivided in pairs that collectively run on two adjacent slices as indicated through colored backgrounds in different shades of green.

In SMT4 or lower thread modes, one to two threads each share a four slices resource domain. Figure 2-5 also indicates other essential resources that are shared among the SMT threads, such as instruction cache, instruction buffer and L1 data cache.

The SMT8 core supports automatic workload balancing to change the operational SMT thread level. Depending on the workload characteristics the number of threads that is running on one chiplet can be reduced from four to two and even further to only one active thread. An individual thread can benefit in terms of performance if fewer threads run against the core's executions resources.

Micro-architecture performance and efficiency optimization lead to a significant improvement of the performance per watt signature compared with the previous POWER9 core implementation. The overall energy efficiency is better by a factor of approximately 2.6, which demonstrates the advancement in processor design that is manifested by Power10.

The Power10 processor core includes the following key features and improvements that affect performance:

► Enhanced load and store bandwidth
► Deeper and wider instruction windows
► Enhanced data prefetch
► Branch execution and prediction enhancements
► Instruction fusion

Enhancements in the area of computation resources, working set size, and data access latency are described next. The change in relation to the POWER9 processor core implementation is provided in parentheses.

Enhanced computation resources

The following are major computational resource enhancements:

► Eight vector scalar unit (VSU) execution slices, each supporting 64-bit scalar or 128-bit single instructions multiple data (SIMD) +100% for permute, fixed-point, floating-point, and crypto (Advanced Encryption Standard (AES)/SHA) +400% operations.

► Four units for matrix-math assist (MMA) acceleration each capable of producing a 512-bit result per cycle (new), +400% Single and Double precision FLOPS plus support for reduced precision AI acceleration).

► Two units for quad-precision floating-point and decimal floating-point operations additional instruction types

Larger working sets

The following major changes were implemented in working set sizes:

► L1 instruction cache: 2 x 48 KB 6-way (96 KB total). (+50%)
► L2 cache: 2 MB 8-way. (+400%)
► L2 translation lookaside buffer (TLB): 2 x 4K entries (8K total). (+400%)

Data access with reduced latencies

The following major changes reduce latency for load data:

► L1 data cache access at four cycles nominal with zero penalty for store-forwarding. (- 2 cycles) for store forwarding

► L2 data access at 13.5 cycles nominal. (-2 cycles)

► L3 data access at 27.5 cycles nominal. (-8 cycles)

► Translation lookaside buffer (TLB) access at 8.5 cycles nominal for effective-to-real address translation (ERAT) miss including for nested translation. (-7 cycles)

Micro-architectural innovations that complement physical and logic design techniques and specifically address energy efficiency include the following examples:

- ► Improved clock-gating
- ► Reduced flush rates with improved branch prediction accuracy
- ► Fusion and gather operating merging
- ► Reduced number of ports and reduced access to selected structures
- ► Effective address (EA)-tagged L1 data and instruction cache yield ERAT access only on a cache miss

In addition to significant improvements in performance and energy efficiency, security represents a major architectural focus area. The Power10 processor core supports the following security features:

- ► Enhanced hardware support that provides improved performance while mitigating for speculation-based attacks
- ► Dynamic Execution Control Register (DEXCR) support
- ► Return oriented programming (ROP) protection

2.1.3 Simultaneous multithreading

Each core of the Power10 processor supports multiple hardware threads that represent independent execution contexts. If only one hardware thread is used, the processor core runs in single-threaded (ST) mode.

If more than one hardware thread is active, the processor runs in simultaneous multi-threading (SMT) mode. In addition to the ST mode, the Power10 processor supports the following different SMT modes:

- ► SMT2: Two hardware threads active
- ► SMT4: Four hardware threads active
- ► SMT8: Eight hardware threads active

SMT enables a single physical processor core to simultaneously dispatch instructions from more than one hardware thread context. Computational workloads can use the processor core's execution units with a higher degree of parallelism. This ability significantly enhances the throughput and scalability of multi-threaded applications and optimizes the compute density for single-threaded workloads.

SMT is primarily beneficial in commercial environments where the speed of an individual transaction is not as critical as the total number of transactions that are performed. SMT typically increases the throughput of most workloads, especially those workloads with large or frequently changing working sets, such as database servers and web servers.

Table 2-2 lists a historic account of the SMT capabilities that are supported by each implementation of the IBM Power Architecture® since POWER4.

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

Technology	Cores/system	Supported hardware threading modes	Maximum hardware threads per partition
IBM POWER4	32	ST	32
IBM POWER5	64	ST, SMT2	128
IBM POWER6	64	ST, SMT2	128
IBM POWER7	256	ST, SMT2, SMT4	1024
IBM POWER8	192	ST, SMT2, SMT4, SMT8	1536
IBM POWER9	192	ST, SMT2, SMT4, SMT8	1536
IBM Power10	240	ST, SMT2, SMT4, SMT8	1920[a]

a. PHYP supports a maximum 240 x SMT8 = 1920. AIX support up to 1920 (240xSMT8) total threads in a single partition, starting with AIX v7.3 + Power10.

2.1.4 Matrix-multiply assist AI workload acceleration

The matrix-multiply assist (MMA) facility was introduced by the Power Instruction Set Architecture (ISA) v3.1. The related instructions implement numerical linear algebra operations on small matrices and are meant to accelerate computation-intensive kernels, such as matrix multiplication, convolution, and discrete Fourier transform.

To efficiently accelerate MMA operations, the Power10 processor core implements a *dense math engine* (DME) microarchitecture that effectively provides an accelerator for cognitive computing, machine learning, and AI inferencing workloads.

The DME encapsulates compute efficient pipelines, a physical register file, and associated data-flow that keeps resulting accumulator data local to the compute units. Each MMA pipeline performs outer-product matrix operations, reading from and writing back a 512-bit accumulator register.

Power10 implements the MMA accumulator architecture without adding an architected state. Each architected 512-bit accumulator register is backed by four 128-bit Vector Scalar eXtension (VSX) registers.

Code that uses the MMA instructions is included in OpenBLAS and Eigen libraries. This library can be built by using the most recent versions of GNU Compiler Collection (GCC) compiler. The latest version of OpenBLAS is available at this web page.

OpenBLAS is used by Python-NumPy library, PyTorch, and other frameworks, which makes it easy to use the performance benefit of the Power10 MMA accelerator for AI workloads.

For more information about the implementation of the Power10 processor's high throughput math engine, see the white paper *A matrix math facility for Power ISA processors*.

For more information about fundamental MMA architecture principles with detailed instruction set usage, register file management concepts, and various supporting facilities, see *Matrix-Multiply Assist Best Practices Guide*, REDP-5612.

2.1.5 Power10 compatibility modes

The Power10 core implements the Processor Compatibility Register (PCR) as described in the Power instruction set architecture (ISA) version 3.1, primarily to facilitate live partition mobility (LPM) to and from previous generations of IBM Power hardware.

Depending on the specific settings of the PCR, the Power10 core runs in a compatibility mode that pertains to POWER9 (Power ISA version 3.0) or POWER8 (Power ISA version 2.07) processors. The support for processor compatibility modes also enables older operating systems versions of AIX, IBM i, Linux, or Virtual I/O server environments to run on Power10 processor-based systems.

The Power10 processors of the Power E1080 server support the following compatibility modes:

► POWER8: Enabled by firmware level FW810
► POWER9 Base: Enabled by firmware level FW910
► POWER9: Enabled by firmware level FW940
► Power10: Enabled by firmware level FW1010

2.1.6 Processor feature codes

Each system node in a Power E1080 server provides four sockets to accommodate Power10 single chip modules (SCMs). Three different SCM types with core density of 10-, 12-, and 15-cores are offered. The processor configuration of a system node is determined by one of three available processor features. Each feature represents a set of four Power10 SCMs of identical core density. All system nodes in a Power E1080 server must be configured with the same processor feature.

Table 2-3 lists the processor feature codes that are available for Power E1080 servers. Power10 processors automatically optimize their core frequencies based on workload requirements, thermal conditions, and power consumption characteristics. Each processor feature code that is listed is associated with a processor core frequency range within which the SCM cores typically operate. Also, the maximum frequency that is supported by the relevant SCM type is indicated.

Table 2-3   Power E1080 processor features

Feature code	Description
EDP2	40-core (4x10) Typical 3.65 to 3.90 GHz (max) Power10 Processor with 5U system node drawer
EDP3	48-core (4x12) Typical 3.60 to 4.15 GHz (max) Power10 Processor with 5U system node drawer
EDP4	60-core (4x15) Typical 3.55 to 4.00 GHz (max) Power10 Processor with 5U system node drawer

2.1.7 On-chip L3 cache and intelligent caching

The Power10 processor includes a large on-chip L3 cache of up to 120 MB with a non-uniform cache access (NUCA) architecture that provides mechanisms to distribute and share cache footprints across a set of L3 cache regions. Each processor core can access an associated local 8MB of L3 cache. It also can access the data in the other L3 cache regions on the chip and throughout the system.

Each L3 region serves as a victim cache for its associated L2 cache and also can provide aggregate storage for the on-chip cache footprint.

Intelligent L3 cache management enables the Power10 processor to optimize the access to L3 cache lines and minimize cache latencies. The L3 includes a replacement algorithm with data type and reuse awareness. It also supports an array of prefetch requests from the core, including instruction and data, and works cooperatively with the core, memory controller, and SMP interconnection fabric to manage prefetch traffic, which optimizes system throughput and data latency.

The L3 cache supports the following key features:

► Enhanced bandwidth that supports up to 64 bytes per core processor cycle to each SMT8 core.

► Enhanced data prefetch that is enabled by 96 L3 prefetch request machines that service prefetch requests to memory for each SMT8 core.

► Plus-one prefetching at the memory controller for enhanced effective prefetch depth and rate.

► Power10 software prefetch modes that support fetching blocks of data into the L3 cache.

► Data access with reduced latencies.

2.1.8 Open memory interface

The Power10 processor introduces a new and innovative open memory interface (OMI). The OMI is driven by 8 on-chip memory controller units (MCUs) and is implemented in two separate physical building blocks that lie in opposite areas at the outer edge of the Power10 die. Each area supports 64 OMI lanes that are grouped in four ports. One port in turn consists of two links with 8 lanes each, which operate in a latency-optimized manner with unprecedented bandwidth and scale at 32 Gbps speed.

The aggregated maximum theoretical full-duplex bandwidth of the OMI interface culminates at 2 x 512 GBps = 1 TBps per Power10 single chip module (SCM).

The OMI physical interface enables low latency, high-bandwidth, technology-agnostic host memory semantics to the processor and allows attaching established and emerging memory elements. With the Power E1080 server. OMI initially supports one main tier, low-latency, enterprise-grade Double Data Rate 4 (DDR4) differential DIMM (DDIMM) per OMI link. This configuration yields a total memory capacity of 16 DDIMMs per SCM and 64 DDIMMs per Power E1080 server node. The memory bandwidth depends on the DDIMM density configured for a specific Power E1080 server.

The maximum theoretical duplex memory bandwidth is 409 GBps per SCM if 32 GB or 64 GB DDIMMs running at 3200 MHz are used. The maximum memory bandwidth is slightly reduced to 375 GBps per SCM if 128 GB or 256 GB DDIMMs running at 2933 MHz are used.

In summary, the Power10 SCM supports 128 OMI lanes with the following characteristics:

► 32 Gbps signaling rate
► Eight lanes per OMI link
► Two OMI links per OMI port (2 x 8 lanes)
► Eight OMI ports per single chip module (16 x 8 lanes)

2.1.9 Pervasive memory encryption

The Power10 MCU provides the system memory interface between the on-chip symmetric multiprocessing (SMP) interconnect fabric and the OMI links. This design qualifies the MCU as ideal functional unit to implement memory encryption logic. The Power10 on-chip MCU encrypts and decrypts all traffic to and from system memory that is based on the AES technology.

The Power10 processor supports the following modes of operation:

► AES XTS mode

   XTS is an abbreviation for the xor–encrypt–xor based tweaked-codebook mode with ciphertext stealing. AES XTS provides a block cipher with strong encryption, which is particularly useful to encrypt persistent memory.

   Persistent DIMM technology retains the data that is stored inside the memory DIMMs, even if the power is turned off. A malicious attacker who gains physical access to the DIMMs can steal memory cards. The data that is stored in the DIMMs can leave the data center in the clear if not encrypted.

   Also, memory cards that leave the data center for repair or replacement can be a potential security breach. Because the attacker might have arbitrary access to the persistent DIMM data. the stronger encryption of the AES XTS mode is required for persistent memory. The AES XTS mode of the Power10 processor is supported for future use if persistent memory solutions become available for IBM Power servers.

► AES CTR mode

   CTR stands for *Counter* mode of operation and designates a low-latency AES bock cipher. Although the level of encrypting is not as strong as with the XTS mode, the low-latency characteristics make it the preferred mode for memory encryption of volatile memory. AES CTR makes it more difficult to physically gain access to data through the memory card interfaces. The goal is to protect against physical attacks, which becomes increasingly important in the context of cloud deployments.

The Power E1080 servers support the AES CTR mode for pervasive memory encryption. Each Power10 processor holds a 128-bit encryption key that is used by the processor's MCU to encrypt the data of the differential DIMMs that are attached to the OMI links.

The MCU crypto engine is transparently integrated into the data path, which ensures that the data fetch and store bandwidth are not compromised by the AES CTR encryption mode. Because the encryption has no noticeable performance effect and because of the obvious security benefit, the pervasive memory encryption is enabled by default and cannot be switched off through any administrative interface.

Note: The pervasive memory encryption of the Power10 processor does not affect the encryption status of a system dump content. All data that is coming from the DDIMMs is decrypted by the memory controller unit before it is passed onto the dump devices under the control of the dump program code. This statement applies to the traditional system dump under the operating system control and the firmware assist dump utility.

Note: The PowerVM Live Partition Mobility (LPM) data encryption does not interfere with the pervasive memory encryption. Data transfer during a LPM operation uses the following general flow:

1. On the source server, the Mover Server Partition (MSP) provides the hypervisor with a buffer.

2. The hypervisor of the source system copies the partition memory into the buffer.

3. The MSP transmits the data over the network.

4. The data is received by the MSP on the target server and copied in to the related buffer.

5. The hypervisor of the target system copies the data from the buffer into the memory space of the target partition.

To facilitate LPM data compression and encryption, the hypervisor on the source system presents the LPM buffer to the on-chip nest accelerator (NX) unit as part of process in Step 2. The reverse decryption and decompress operation is applied on the target server as part of process in Step 4.

The pervasive memory encryption logic of the MCU decrypts the memory data before it is compressed and encrypted by the NX unit on the source server. It also encrypts the data before it is written to memory but after it is decrypted and decompressed by the NX unit of the target server.

2.1.10 Nest accelerator

The Power10 processor has an on-chip accelerator called nest accelerator unit or NX unit. The coprocessor features that are available on the Power10 processor are similar to the features on the POWER9 processor. These coprocessors provide specialized functions, such as the following examples:

► IBM proprietary data compression and decompression
► Industry standard gzip compression and decompression
► AES and Secure Hash Algorithm (SHA) cryptography
► Random number generation

Figure 2-6 on page 63 shows a block diagram of the NX unit.
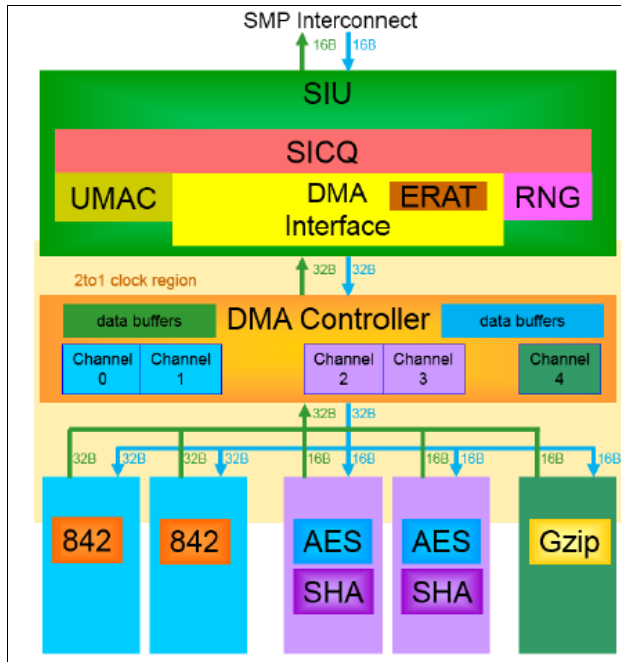
Figure 2-6   Block diagram of the NX unit

Each one of the AES/SHA engines, data compression, and Gzip units consist of a coprocessor type and the NX unit features three coprocessor types. The NX unit also includes more support hardware to support coprocessor invocation by user code, use of effective addresses, high-bandwidth storage accesses, and interrupt notification of job completion.

The direct memory access (DMA) controller of the NX unit helps to start the coprocessors and move data on behalf of coprocessors. SMP interconnect unit (SIU) provides the interface between the Power10 SMP interconnect and the DMA controller.

The NX coprocessors can be started transparently through library or operating system kernel calls to speed up operations that are related to data compression, live partition mobility migration, IPSec, JFS2 encrypted file systems, PKCS11 encryption, random number generation, and the most recently announced logical volume encryption.

In effect, this on-chip NX unit on Power10 systems implements a high throughput engine that can perform the equivalent work of multiple cores. The system performance can benefit by off-loading these expensive operations to on-chip accelerators, which in turn can greatly reduce the CPU usage and improve the performance of applications.

The accelerators are shared among the logical partitions (LPARs) under the control of the PowerVM hypervisor and accessed by way of hypervisor call. The operating system, along with the PowerVM hypervisor, provides a send address space that is unique per process requesting the coprocessor access. This configuration allows the user process to directly post entries to the first in - first out (FIFO) queues that are associated with the NX accelerators. Each NX coprocessor type has a unique receive address space corresponding to a unique FIFO for each of the accelerators.

For more information about the use of the xgzip tool that uses the gzip accelerator engine, see the following resources:

► IBM support article *Using the POWER9 NX (gzip) accelerator in AIX*

► IBM Power Systems community article *Power9 GZIP Data Acceleration with IBM AIX*

► AIX community article *Performance improvement in openssh with on-chip data compression accelerator in power9*

► IBM Documentation: *nxstat Command*

2.1.11 SMP interconnect and accelerator interface

The Power10 processor provides a highly-optimized, 32 Gbps differential signaling technology interface that is structured in 16 entities. Each entity consists of eight data lanes and one spare lane. This interface can facilitate the following functional purposes:

► First- or second-tier, symmetric multiprocessing link interface, enabling up to 16 Power10 processors to be combined into a large, robustly scalable, single-system image.

► Open Coherent Accelerator Processor Interface (OpenCAPI), to attach cache coherent and I/O-coherent computational accelerators, load/store addressable host memory devices, low latency network controllers, and intelligent storage controllers.

► Host-to-host integrated memory clustering interconnect, enabling multiple Power10 systems to directly use memory throughout the cluster.

Note: The OpenCAPI interface and the memory clustering interconnect are Power10 technology option for future use.

Because of the versatile nature of signaling technology, the 32 Gbps interface is also referred to as Power/A-bus/X-bus/OpenCAPI/Networking ($PowerAXON$) interface. The IBM proprietary X-bus links connect two processors on a board with a common reference clock. The IBM proprietary A-bus links connect two processors in different drawers on different reference clocks by using a cable.

OpenCAPI is an open interface architecture that allows any microprocessor to attach to the following items:

► Coherent user-level accelerators and I/O devices
► Advanced memories accessible through read/write or user-level DMA semantics

The OpenCAPI technology is developed, enabled, and standardized by the OpenCAPI Consortium. For more information about the consortium's mission and the OpenCAPI protocol specification, see OpenCAPI Consortium.

The PowerAXON interface is implemented on dedicated areas that are at each corner of the Power10 processor die. The Power E1080 server makes use of this interface to implement single-drawer chip-to-chip and drawer-to-drawer chip interconnects.

The Power E1080 single-drawer chip-to-chip SMP interconnect features the following properties:

► Three (2 x 9)-bit on planar buses per Power10 single chip module (SCM)

► Eight data lanes, plus one spare lane in each direction per chip-to-chip connection

► 32 Gbps signaling rate providing 128 GBps per chip-to-chip SMP connection bandwidth, an increase of 33% compared to the Power E980 single-drawer implementation

► 4-way SMP architecture implementations build out of four Power10 SCMs per drawer in 1-hop topology

The Power E1080 drawer-to-drawer SMP interconnect features the following properties:

► Three (2 x 9)-bit buses per Power10 SCM

► Eight data lanes plus one spare lane in each direction per chip-to-chip connection

► Each of the four SCMs in a drawer is connected directly to an SCM at the same position in every other drawer in a multi-node system

► 32 Gbps signaling rate, which provides 128 GBps per chip-to-chip inter node SMP connection bandwidth

► 8-socket, 12-socket and 16-socket SMP configuration options in 2-hop topology

Figure 2-7 shows the SMP connections for a fully configured 4-node 16-socket Power E1080 system. The blue lines represent the chip-to-chip connections within one system node. The green lines represent the drawer-to-drawer SMP connections.
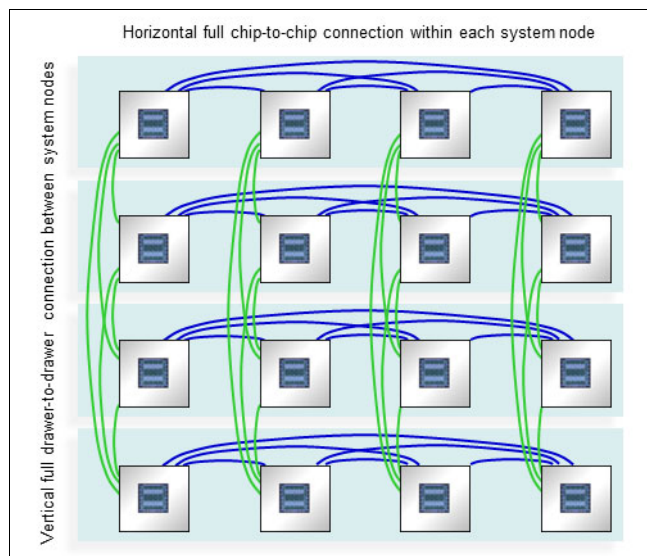

Figure 2-7   SMP interconnect in a 4-node 16-socket Power E1080 system

From the drawing that is shown in Figure 2-7, it is easy to deduce that each socket is directly connected to any other socket within one system node and only one intermediary socket is required to get from a chip to any other chip in another CEC drawer.

2.1.12 Power and performance management

Power10 processor-based servers implement an enhanced version of the power management EnergyScale technology. As in the previous POWER9 EngeryScale implementation, the Power10 EnergyScale technology supports dynamic processor frequency changes that are dependent on several factors, such as workload characteristics, the number of active cores, and environmental conditions.

Based on the extensive experience that was gained over the past few years, the Power10 EnergyScale technology evolved to use the following effective and simplified set of operational modes:

► Power saving mode
► Static mode (nominal frequency)
► Maximum performance mode (MPM)

The POWER9 dynamic performance mode (DPM) has many features in common with the POWER9 maximum performance mode (MPM). Because of this redundant nature of characteristics, the DPM for Power10 processor-based systems was removed in favor of an enhanced MPM. For example, the maximum frequency is now achieveable in the Power10 enhanced maximum performance mode (regardless of the number of active cores), which was not always the case with POWER9 processor-based servers.

The Power10 processor-based Power E1080 system features MPM enabled by default. This mode dynamically adjusts processor frequency to maximize performance and enable a much higher processor frequency range. Each of the power saver modes deliver consistent system performance without any variation if the nominal operating environment limits are met.

For Power10 processor-based systems that are under control of the PowerVM hypervisor, the MPM is a system-wide configuration setting, but each processor module frequency is optimized separately.

The following factors determine the maximum frequency that a processor module can run at:

► Processor utilization: Lighter workloads run at higher frequencies.

► Number of active cores: Fewer active cores run at higher frequencies.

► Environmental conditions: At lower ambient temperatures, cores are enabled to run at higher frequencies.

The following Power10 EnergyScale modes are available:

► Power saving mode

    The frequency is set to the minimum frequency to reduce energy consumption. Enabling this feature reduces power consumption by lowering the processor clock frequency and voltage to fixed values. This configuration reduces power consumption of the system while delivering predictable performance.

► Static mode

    The frequency is set to a fixed point that can be maintained with all normal workloads and in all normal environmental conditions. This frequency is also referred to as *nominal frequency*.

► Maximum performance mode

    Workloads run at the highest frequency possible, depending on workload, active core count, and environmental conditions. The frequency does not go below the static frequency for all normal workloads and in all normal environmental conditions.

In MPM, the workload is run at the highest frequency possible. The higher power draw enables the processor modules to run in an MPM typical frequency range (MTFR), where the lower limit is well above the nominal frequency and the upper limit is given by the system's maximum frequency.

The MTFR is published as part of the system specifications of a specific Power10 system if it is running by default in MPM. The higher power draw potentially increases the fan speed of the respective system node to meet the higher cooling requirements, which in turn causes a higher noise emission level of up to 15 decibels.

The processor frequency typically stays within the limits that are set by the MTFR, but can be lowered to frequencies between the MTFR lower limit and the nominal frequency at high ambient temperatures above 27 °C (80.6 °F). If the data center ambient environment is less than 27 °C, the frequency in MPM consistently is in the upper range of the MTFR (roughly 10% - 20% better than nominal). At lower ambient temperatures (below 27 °C, or 80.6 °F), MPM mode also provides deterministic performance. As the ambient temperature increases above 27 °C, determinism can no longer be ensured. This mode is the default mode in the Power E1080.

► Idle power saver mode (IPS)

IPS mode lowers the frequency to the minimum if the entire system (all cores of all sockets) is idle. It can be enabled or disabled separately from all other modes. The Power E1080 does not support this mode.

Figure 2-8 shows the comparative frequency ranges for the Power10 power saving mode, static or nominal mode, and the maximum performance mode. The frequency adjustments for different workload characteristics, ambient conditions, and idle states are also indicated.
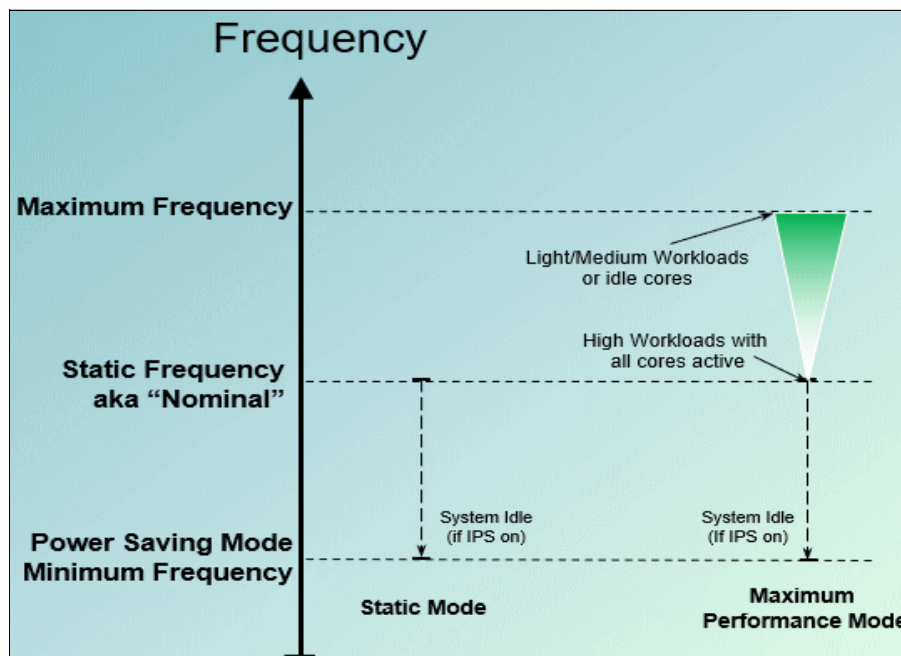
Figure 2-8   Power10 power management modes and related frequency ranges

Table 2-4 shows the power saving mode and the static mode frequencies, and the frequency ranges of the MPM for all three processor module types that are available for the Power E1080 server.

Table 2-4   Characteristic frequencies and frequency ranges for Power E1080 server

Feature code	Cores per single-chip module	Power saving mode frequency [GHz]	Static mode frequency [GHz]	Maximum performance mode frequency range [GHz]
EDP2	10	3.25	3.65	3.65 to 3.90 GHz (max)
EDP3	12	3.40	3.60	3.60 to 4.15 GHz (max)
EDP4	15	3.25	3.55	3.55 to 4.00 GHz (max)

For Power E1080 servers, the MPM is enabled by default.

The controls for all power saver modes are available on the Advanced System Management Interface (ASMI) and can be dynamically modified. A system administrator can also use the Hardware Management Console (HMC) to set power saver mode or to enable static mode or MPM.

Figure 2-9 shows the ASM interface menu for Power and Performance Mode Setup on a Power E1080 server.
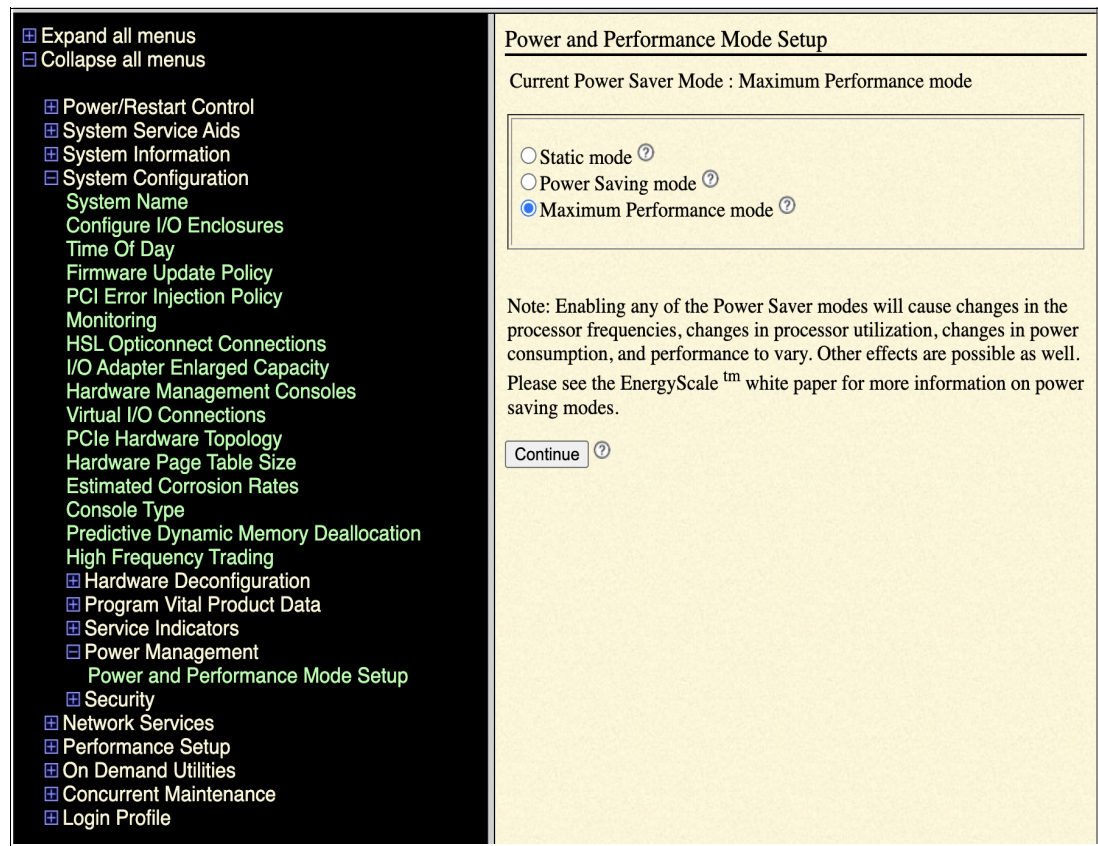
Figure 2-9   Power E1080 ASMI menu for Power and Performance Mode setup

Figure 2-10 shows the HMC menu for power and performance mode setup.

Power and Performance Mode Setup

Enabling any of the Power Saver Modes causes changes in the processor frequencies, processor utilization, power consumption, and performance. Before you turn on any power management modes, it is advisable to understand these effects as described in the Learn More.

Power Saver

Enabling the power saver mode on this managed system reduces the power consumption of the processors.

✔ Power Saver Mode: Enabled Maximum Performance Mode

　　⦿ Maximum Performance
　　◯ Static
　　◯ Power Saver

Idle Power Saver

Reduced energy consumption when the system is considered idle.

🚫 Idle Power Saver Mode: Disabled Idle Power Saver

Idle Power Management is not supported.

Learn more ➜　　　　　　　　　　　OK　　　　Cancel

Figure 2-10　Power E1080 HMC menu for Power and Performance Mode setup

2.1.13 Comparing Power10, POWER9, and POWER8 processors

The Power E1080 enterprise class systems use exclusively single chip modules (SCM) with up to 15 active SMT8 capable cores. These SCM processor modules are structural and performance optimized for usage in *scale-up multi-socket* systems.

Table 2-5 compares key features and characteristics of the Power10, POWER9, and POWER8 processor implementations as used in enterprise class scale-up servers.

Table 2-5　Comparison of the Power10 processor technology to prior processor generations

Characteristics	Power10	POWER9	POWER8
Technology	7 nm	14 nm	22 nm
Die size	602 mm^2	693 mm^2	649 mm^2
Processor module size	68.5 mm x 77.5 mm	68.5 mm x 68.5 mm	71.5 mm x 71.5 mm
Number of transistors	18 billion	8 billion	4.2 billion
Maximum cores	15	12	12
Maximum hardware threads per core	8	8	8
Maximum static frequency / high performance frequency range	3.75 - 4.15 GHZ	3.9 - 4.0 GHz	4.15 GHz
L2 Cache per core	2048 KB	512 KB	512 KB

Characteristics	Power10	POWER9	POWER8
L3 Cache	8 MB of L3 cache per core with each core having access to the full 120 MB of L3 cache, on-chip high-efficiency SRAM[a]	10 MB of L3 cache per core with each core having access to the full 120 MB of L3 cache, on-chip eDRAM[b]	8 MB of L3 cache per core with each core having access to the full 96 MB of L3 cache, on-chip eDRAM
Support memory technology	DDR4[c]	DDR4 and DDR3[d]	DDR3 and DDR4
I/O bus	PCIe Gen 5	PCIe Gen 4	PCIe Gen 3

a. Static random-access memory
b. Embedded dynamic random-access memory
c. Power10 processor memory logic and the Power E1080 memory subsystem can use DDR5 technology DIMMs.
d. Only DDR3 memory CDIMMs, which are transferred in the context of a model upgrade from Power E870, Power E870C, Power E880, or Power E880C systems to a Power E980 server, are supported.

2.2 SMP interconnection

Each of the four Power10 processor chips in the system node drawer are connected directly to a Power10 processor chip at the same position in every other system node drawer by using the SMP OP-bus. Each OP-bus connection between any two system node drawers uses a pair of cables for the connection, with each cable carrying one half of the data lanes.

Power System E980 SMP cables cannot be used on Power E1080 because the Power10 processor-based server uses a different set of cables.

Figure 2-11 shows how each Power10 processor socket in a system node drawer has the bus routed to the rear tailstock of the chassis. The Power10 processor chip in the first socket (see processor chip socket #0 in Figure 2-11) uses ports T0 and T1 for the OP6 bus connection (see top left corner in Figure 2-11).
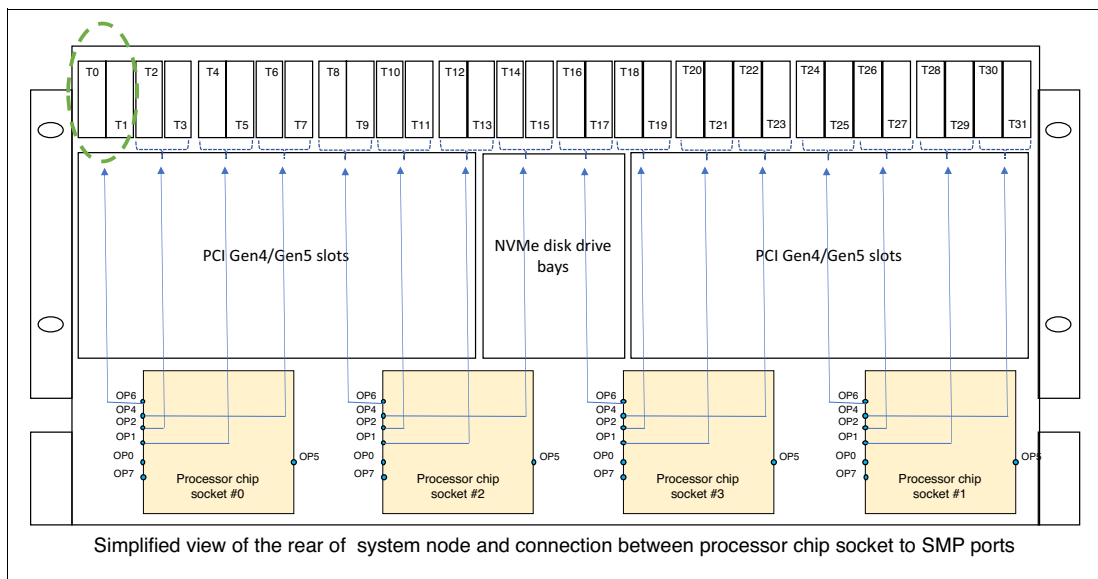
Figure 2-11   Logical connection from processor chip sockets to SMP external ports

Each Power10 processor chip socket externalizes four SMP/OpenCAPI buses:

► OP6: SMP or OpenCAPI
► OP2: SMP only
► OP1: SMP only
► OP4: SMP or OpenCAPI

The same cables are used for two different modes:

► SMP for interconnecting the system node drawers
► OpenCAPI

Each cable contains 9 bit lanes:

► SMP mode uses nine of nine lanes available
► OpenCAPI mode uses eight of the nine lanes available

2.2.1 Two-system node drawers OP-bus connection

In the two system drawer configuration, two OP-buses interconnect the processors between the two drawers. Figure 2-12 shows all the connections between the SMP ports of the first system node drawer with the SMP ports of the second system node drawer that are accomplished by using the proper set of cables.
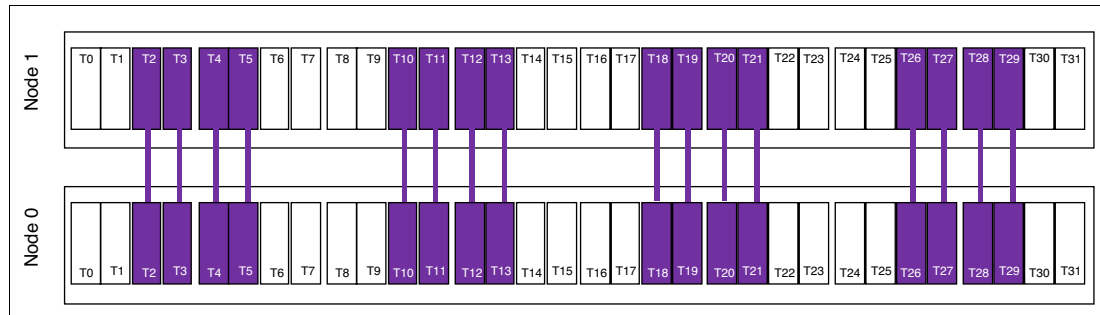
Figure 2-12   Two system node drawers SMP connection

The length of every cable is 775 mm (2.5 ft.) and its tape color is violet. As Figure 2-12 shows, 16 cables are needed for this configuration.

2.2.2 SMP cable RAS attribute

Each SMP bus has 18 lanes and it is split between two separate cables, where each cable has nine lanes. An SMP spare lane is used on each SMP cable (eight lanes for data and one lane as spare). If one of the two SMP cables fail, all the traffic is routed to the one remaining cable, which reduces the bandwidth by half.

The Power E1080 supports concurrent repair of the external SMP cables; therefore, the entire server does not need to be shut down to replace an external SMP cable.

The concurrent repair process enables the system to restore the OP-bus to the full function without a system outage.

2.3 Memory subsystem

The Power E1080 server uses the next-generation differential dual inline memory modules (DIMMs), which are high-performance, high-reliability, high-function memory cards that contain a buffer chip, intelligence, and 2933/3200 MHz DRAM memory.

DDR4 technology is used and provides the 2933/3200 MHz DRAM memory. DDIMMs are placed in the DDIMM slots in the system node.

Every Power10 chip memory controller unit (MCU) provides the system memory interface between the on-chip SMP interconnect fabric and the OpenCAPI memory interface (OMI) links. Logically eight, essentially-independent MCU channels are on every chip that interface to a total of 16 high-speed OMI links.

Each memory channel supports two OMI links (referred to as subchannels A and B) as shown in Figure 2-13 on page 73, which also shows the memory slot locations that are associated to the related subchannels, channels, and memory controller units.

All 16 of the OMI subchannels come off the Power10 SCM and are wired to 16 memory buffer based DDIMM slots, with one DDIMM connected to each OMI subchannel.

Figure 2-13   Memory subsystem logical diagram

Each Power10 processor chip supports 16 DDIMMs; therefore, a maximum of 64 DDIMMs in each Power E1080 CEC drawer can support up to 16 TB of memory RAM capacity.
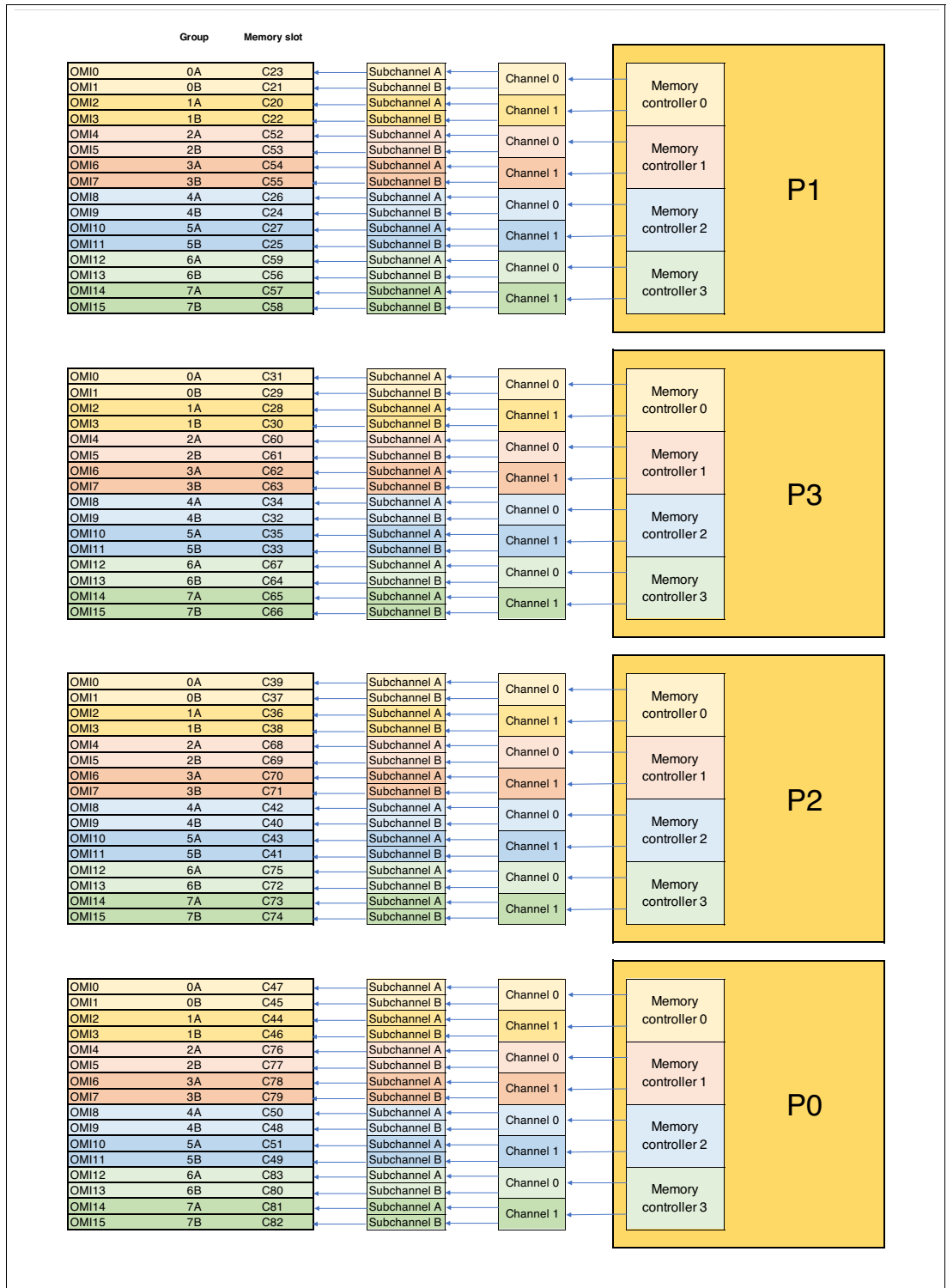
The DDIMM densities that are supported in the Power E1080 are 32 GB, 64 GB,128 GB and 256 GB and any memory feature code provides four DDIMMs. The DDIMMs have their own voltage regulators and are n+1. redundant. Two PMICs plus two spares are used.

It takes two Power Management IC (PMICs) to supply all the voltage levels that are required by the DDIMM. On the DDIMM, four PMICs are used, which consist of two redundant pairs. If one PMIC in each of the redundant pairs is still functional, the DDIMM is not called for replacement.

As described in 1.5.3, "Memory features" on page 19, the Power E1080 requires the activation of 50% or more of the installed physical memory.

PowerVM hypervisor provides Active Memory Mirroring, which is designed to ensure that system operation continues (even in the unlikely event of an uncorrectable error occurring in main memory).

2.3.1 Memory bandwidth

The Power10 single chip module (SCM) supports 16 open memory interface (OMI) links and one differential DIMM (DDIMM) is driven by each OMI link in turn. One OMI link represents a bundle of 8 lanes and is as such capable to transfer 8 byte with one transaction.

The Power E1080 offers four different DDIMM sizes: 32 GB, 64 GB, 128 GB, and 256 GB. The 32 GB and 64 GB DDIMMs run at a data rate of 3200 Mbps; the 128 GB and 256 GB DDIMMs at a data rate of 2933 Mbps.

Table 2-6 lists the available DDIMM capacities and their related maximum theoretical bandwidth figures per OMI link and per Power10 SCM.

Table 2-6   Memory bandwidth of supported DDIMM sizes

DDIMM capacity	Data rate	Maximum theoretical bandwidth per OMI link	Maximum theoretical bandwidth per SCM[a]
32 GB / 64 GB	3200 Mbps	25.6 GBps	409.6 GBps
128 GB / 256 GB	2933 Mbps	23.5 GBps	375.4 GBps

a. DDIMM modules attached to one SCM must be all of the same size

Important: For the best possible performance, it is generally recommended that memory be installed in all memory slots and evenly across all system node drawers and all SCMs in the system.

Balancing memory across the installed system planar cards enables memory access in a consistent manner and typically results in better performance for your configuration.

The Active Memory Expansion (AME) feature is an option that can increase the effective memory capacity of the system. AME uses the on-chip nest accelerator (NX) unit to compress or decompress memory content. This separately priced hardware feature is enabled at partition (LPAR) level. The related LPAR profile parameter allows it to configure the compression factor in accordance with memory capacity needs and performance trade-off consideration.

2.3.2 Memory placement rules

In a Power E1080 single node system, each SCM must have eight of the 16 memory DDIMMs slots populated. Only increments of four DDIMMs are supported, all of which have the same capacity. These four DDIMMs must all be plugged into four of the eight remaining slots that are connected to an SCM.

DDIMMs cannot be installed in such a way that an odd number of DDIMM slots are populated behind an SCM. DDIMMs also cannot be installed in a configuration that results in 10 or 14 DDIMM slots being populated behind an SCM. The only valid number of DDIMMs plugged into slots connected to a specific SCM is 8, 12, or 16.

The Power E1080 system must adhere to the following DDIM plug rules:

► Minimum memory that is allowed is 1 TB (see 1.5.1, "Minimum configuration" on page 15).

► All DDIMMs under each SCM must have the same capacity.

► DDIMMs must be plugged in groups of four (quads).

► Each DDIMM quad must all be the same size and type (identical to each other) and must be the same with other DDIMM quads under each SCM, but can be different from other DDIMM quads from other SCMs.

  For example, in a one system node configuration:

  – First SCM can be connected to 8x32 GB DDIMMS
  – Second SCM can be connected to 8x64 GB DDIMMS
  – Third SCM can be connected to 8x128 GB DDIMMS
  – Forth SCM can be connected to 8x256 GB DDIMMS

Overall system performance improves when all the quads match each other. Physical memory features from a previous generation of servers cannot be migrated to Power E1080.

For the best possible performance, it is generally recommended that memory be installed in all memory slots and evenly across all system node drawers and all SCMs in the system.

Balancing memory across the installed system planar cards enables memory access in a consistent manner and typically results in better performance for your configuration.

You must account for any plans for future memory upgrades when you decide which memory feature size to use at the time of the initial system order.

Plug sequence for minimum memory

A Power E1080 drawer requires a minimum of 32 DDIMMs. Two quads of DDIMMs per SCM are required.

Spreading the DDIMMs evenly between the two memory controllers provides maximum performance. The DDIMMs within each quad must be the same size and type and all the DDIMMs under each SCM have to be same capacity.

A DDIMM quad can be different from the DDIMM quads plugged elsewhere.

Adhere to the following order to plug the 32 DDIMMs and use Figure 2-13 on page 73 as reference:

► C20, C21, C22, C23, C24, C25, C26, C27 (slots connected to SCM P1)
► C28, C29, C30, C31, C32, C33, C34, C35 (slots connected to SCM P3)
► C36, C37, C38, C39, C40, C41, C42, C43 (slots connected to SCM P2)
► C44, C45, C46, C47, C48, C49, C50, C51 (slots connected to SCM P0)

Plug Sequence for the Remaining DDIMMs

For a Power E1080 drawer, the next quad of DDIMMs sequence and rules are described in this section.

For the next quad of DDIMMs to be installed in slots that are connected to one of the four pSCMs, consider the following rules:

► The DDIMMs within each quad must be the same size and type.

► All DDIMMs under each SCM must have the same capacity.

► The DDIMM quad can be different from the DDIMM quads plugged elsewhere, including the other quad connected to the same SCM.

► Plug the remaining quad of DDIMMs in any of the remaining open quad of DIMM slots and all DDIMMs under each SCM must have the same capacity

Adhere to the following order and use Figure 2-13 on page 73 as reference:

► C52, C53, C54, C55 (slots connected to SCM P1) or
► C60, C61, C62, C63 (slots connected to SCM P3) or
► C68, C69, C70, C71 (slots connected to SCM P2) or
► C76, C77, C78, C79 (slots connected to SCM P0) or
► C56, C57, C58, C59 (slots connected to SCM P1) or
► C64, C65, C66, C67 (slots connected to SCM P3) or
► C72, C73, C74, C75 (slots connected to SCM P2) or
► C80, C81, C82, C83 (slots connected to SCM P0)

> **Note:** In a multi-system node configuration, a drawer can be fully populated. The other drawers can be partially populated or have all the drawers symmetrically populated.

2.4 Capacity on-demand

Starting with IBM Power Systems Private Cloud with Shared Utility Capacity (Power Enterprise Pools 2.0), multiple types of offerings are available on the Power E1080 server to help meet changing resource requirements in an on-demand environment by using resources that are installed on the system but that are not activated.

For more information about CoD Activation codes, see Power Systems Capacity on Demand.

2.4.1 New Capacity on Demand features

The Power E1080 server features similar capabilities that are offered for the Power E980 server with a few changes:

► With IBM Power Systems Private Cloud with Shared Utility Capacity (Power Enterprise Pools 2.0) configuration, it is possible to order a Power E1080 server with as few as one Base Processor Activation and 256 GB Base Memory activations. When the Power Enterprise Pools 2.0 configuration is used, all of the resources are activated; therefore, other CoD features are not supported.

► Without IBM Power Systems Private Cloud with Shared Utility Capacity configuration:

  – Each Power System E1080 server requires a minimum of 16 permanent processor core activations that use static activations or Linux on Power activations. The remaining cores can be permanently or temporarily activated by using extra features.

- A minimum of 50% of installed memory capacity must have activations. These activations can be static, mobile, or Linux on Power.

- At least 25% of memory capacity must have static activations or Linux on Power activations.

- The Power E1080 server can participate in the same IBM Power Enterprise Pool 1.0 with other Power E1080 servers and with previous generation Power E980 servers, but not with Power E870C, Power E880, and Power E880C servers.

- CUoD, Pre-paid Elastic Capacity on Demand (4586-COD), and Trial Capacity on Demand (Trial CoD) are available with Power E1080 server.

- Utility Capacity on Demand (Utility CoD) is not supported on Power E1080 server.

- Post-Pay Elastic CoD offering is not available on Power E1080 server.

- Enterprise Capacity Backup (CBU) offering is not available on Power E1080 server.

- Mobile-Enabled activations are not available on Power E1080 server.

2.4.2 IBM Power Systems Private Cloud with Shared Utility Capacity

IBM Power Systems Private Cloud with Shared Utility Capacity solution (Power Enterprise Pools 2.0) is an infrastructure offering model that enables cloud agility and cost optimization with pay-for-use pricing.

All installed processors and memory on systems in a pool are activated and made available for immediate use when a pool is started. Processor and memory usage on each server are tracked by the minute and aggregated across the pool.

The capacity in this model consists of Base Activations and Capacity Credits, which are shared across the pool without having to move them from server to server. The unpurchased capacity in the pool can be used on a pay-as-you-go basis.

Resource usage that exceeds the pool's aggregated base resources is charged as metered capacity by the minute. It also is debited against purchased Capacity Credits on a real-time basis, as shown in Figure 2-14.
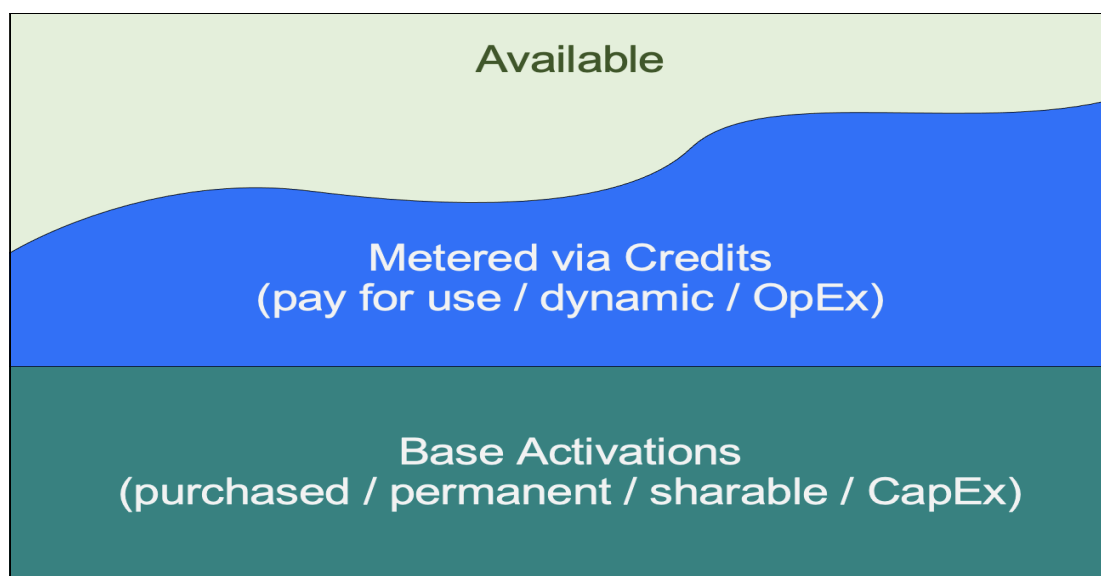

Figure 2-14   IBM Power Systems Private Cloud with Shared Utility Capacity

IBM Power Systems Private Cloud with Shared Utility Capacity solution is supported only on specific POWER9 and Power10 processor-based systems. Power E1080 servers can co-exist with Power E980 systems in the same pool.

A single Power Enterprise Pool 2.0 supports up to 2000 VMs; up to 1000 VMs are supported per HMC. At the time of this writing, up to 200 VMs are supported per Power E1080 server, which is planned to be increased to 750 VMs per Power E1080.

Besides processor and memory metering, Power Enterprise Pools 2.0 enables metering AIX and IBM i software entitlements and SLES and RHEL subscriptions in the pool.

For more information and requirements, see IBM Power Systems Private Cloud with Shared Utility Capacity, SG24-8478.

2.4.3 Static, Mobile, and Base activations

Static processor and Memory activations are restricted to a single system. They are enabled by entering a code into the HMC that manages the system. The extra cores or memory are immediately available for use by workloads on the system.

Static activations are available in two versions:

► Static activations: This standard processor or memory activation can run VIOS, AIX, IBM i, and Linux workloads

► Static activations for Linux: This processor or memory activation can run only VIOS and Linux workloads

Mobile activations are ordered against a specific server, but can be moved to any server within the Power Enterprise Pool 1.0 and support any type of application. Mobile activations can be purchased in the initial order or with an MES upgrade.

Base Processor and Memory activations are associated with a single system and initially act as Static activations. When a system is added to a Power Enterprise Pools 2.0 configuration, all of its Base activations are added to the pool's base capacity.

Base activations for processor and memory resources are available in the following types:

► Base activations: This processor or memory activation can run VIOS, AIX, IBM i and Linux workloads. Initially, Base activations act like Static activations until the system is added to a Power Enterprise Pool 2.0. Static and Mobile activations can be converted to Base activations with an MES order.

► Base Linux activations: This Base Processor activation can run only VIOS and Linux workloads.

Table 2-7 lists the Static, Mobile, and Base processor activation features that are available for initial order on the Power E1080 server.

Table 2-7   Static, Mobile, and Base processor activation features

Processor feature	EDP2 (3.65 - 3.90 GHz 40-core)	EDP3 (3.60 - 4.15 GHz 48-core)	EDP4 (3.55 - 4.00 GHz 60-core)
Static activation	EDPB	EDPC	EDPD
Static activation for Linux	ELCL	ELCQ	ELCM
Base activation	EPDC	EPDD	EPDS
Base activation for Linux	EPDU	EPDW	EPDX

Processor feature	EDP2 (3.65 - 3.90 GHz 40-core)	EDP3 (3.60 - 4.15 GHz 48-core)	EDP4 (3.55 - 4.00 GHz 60-core)
Mobile activation	EDPZ	EDPZ	EDPZ

Table 2-8 lists the Static, Mobile, and Base memory activations that are available for initial order on the Power E1080 server.

Table 2-8   Static, Mobile, and Base Memory activation features

Feature code	Description
EMAZ	1 GB Memory activation for HEX - Static
EMQZ	100 GB Memory activation for HEX - Static
EMBZ	512 GB Memory activation for HEX - Static
ELME	512 GB Memory activations for Linux on Power HEX - Static
EDAG	256 GB Base Memory activation for Pools 2.0
EDAH	512 GB Base Memory activation for Pools 2.0
EDAB	100 GB DDR4 Mobile Memory Activation for HEX
EMBK	500 GB DDR4 Mobile Memory Activation for HEX

2.4.4 Capacity Upgrade on-Demand

The Power E1080 server includes several active processor cores and memory units. It can also include inactive processor cores and memory units.

Active processor cores or memory units are processor cores or memory units that are available for use on your server when it comes from the manufacturer. Inactive cores or units are processor cores or memory units that are included with your server, but not available for use until you activate them.

Inactive processor cores and memory units can be permanently activated by purchasing an activation feature that is called CUoD and entering the provided activation code on the HMC for the server.

With the CUoD offering, you can purchase more static processors or memory capacity and dynamically activate them without restarting your server or interrupting your business. All the static processor or memory activations are restricted to a single server.

CUoD features several benefits that enable a more flexible environment. One benefits is reducing the initial investment in a system. Traditional projects that use other technologies means that a system must be acquired with all the resources available to support the entire lifecycle of the project. As a result, you pay up front for capacity that you do not need until the later stages of the project or possibly at all, which affects software licensing costs and software maintenance.

By using CUoD, a company starts with a system with enough *installed* resources to support the entire project lifecycle, but uses only *active* resources that are necessary for the initial project phases. Resources can be added as the project proceeds by activating resources as needed. Therefore, a company can reduce the initial investment in hardware and acquire software licenses only when they are needed for each project phase, which reduces the total cost of ownership (TCO) and total cost of acquisition (TCA) of the solution.

Figure 2-15 shows a comparison between two scenarios: a fully activated system versus a system with CUoD resources being activated along the project timeline.
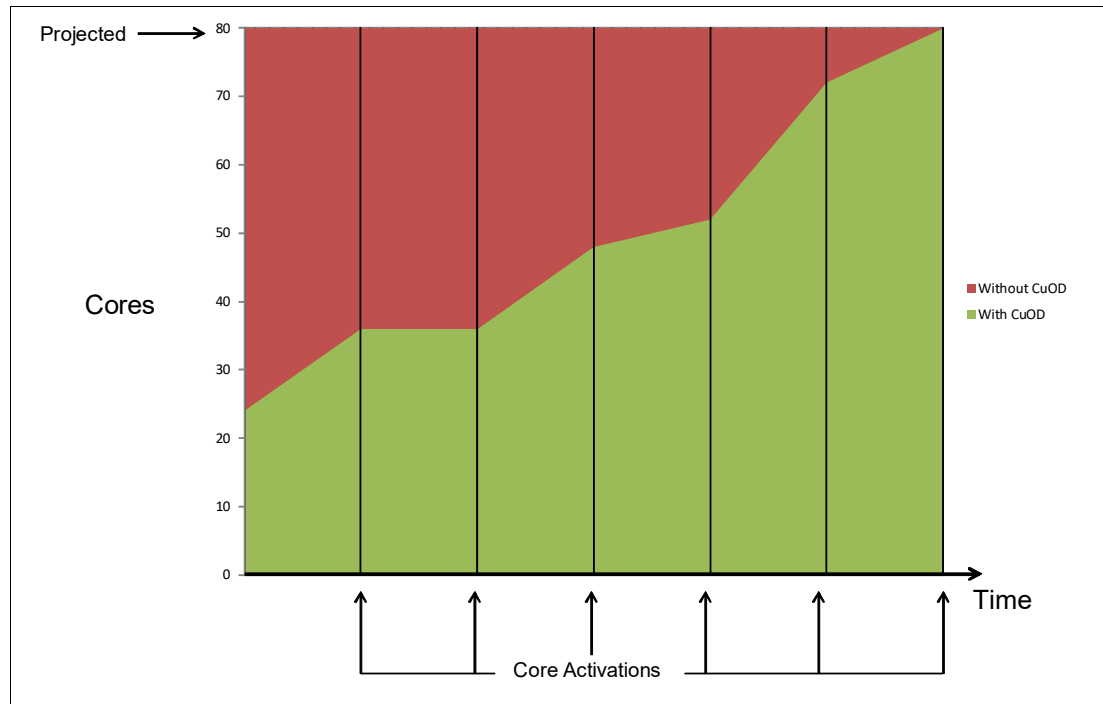
Figure 2-15 Active cores scenarios comparison during a project lifecycle

2.4.5 Elastic Capacity on-Demand (Temporary)

With the Elastic CoD offering, you can temporarily activate and deactivate processor cores and memory units to help meet the demands of business peaks, such as seasonal activity, period-end, or special promotions. When you order an Elastic CoD feature, you receive an enablement code that allows a system uses to make requests for more processor and memory capacity in increments of one processor day or 1 GB memory day. The system monitors the amount and duration of the activations. With Power E1080 server, only the prepaid Elastic Capacity on-Demand option is available with feature code 4586-COD in econfig or directly from ESS.

Processors and memory can be activated and turned off an unlimited number of times when more resources are needed.

This offering provides a system administrator an interface at the HMC to manage the activation and deactivation of resources. Before temporary capacity is used on your server, you must enable your server. To enable your server, an enablement feature (MES only) must be ordered and the required contracts must be in place. The 90-day enablement feature is *not* available for the Power E1080 processors.

For more information about enablement and usage of Elastic CoD, see Power Systems Capacity on Demand.

2.4.6 IBM Power Enterprise Pools 1.0 and Mobile Capacity on-Demand

Although static activations are valid for a single system, some customers might benefit from moving processor and memory activations to different servers because of workload rebalance or disaster recovery.

IBM Power Enterprise Pools 1.0 is a technology for dynamically sharing processor and memory activations among a group (or pool) of IBM Power Systems servers. By using Mobile Capacity on-Demand (CoD) activation codes, the systems administrator can perform tasks without contacting IBM.

With this capability, you can move resources between Power E1080 and Power E980 systems and have unsurpassed flexibility for workload balancing and system maintenance.

Note: Only two consecutive generation of Power servers are supported in the same Power Enterprise Pools 1.0; therefore, Power E1080 can be mixed with Power E980 servers only.

The supported Power Enterprise Pool 1.0 members are listed by pool type in Table 2-9.

Table 2-9   Supported Power Enterprise Pool 1.0 members by pool type

Power Enterprise Pool type	Pool members
Midrange Power Enterprise Pool 1.0	770+, E870, E870C, E880C
High-end Power Enterprise Pool 1.0	780+, 795, E880, E870C, E880C
Power Enterprise Pool 1.0	E870, E880, E870C, E880C, E980
Power Enterprise Pool 1.0 (with Power E1080)	E980, E1080

A pool can support systems with different clock speeds or processor generations.

Mobile Capacity on Demand (Mobile CoD) features the following basic rules:

► The Power E1080 server requires a minimum of 16 static processor activations.

► The Power E980 server requires a minimum of eight static processor activations.

► For all systems, 50% of the installed memory must have activations, and a minimum of 25% of those must be static activations.

An HMC can manage multiple IBM Power Enterprise Pools and systems that are not part of an IBM Power Enterprise Pool. Systems can belong to only one IBM Power Enterprise Pool at a time. Powering down an HMC does not limit the assigned resources of participating systems in a pool, but does limit the ability to perform pool change operations.

After an IBM Power Enterprise Pool is created, the HMC can be used to perform the following tasks:

► Mobile CoD processor and memory resources can be assigned to systems with inactive resources. Mobile CoD resources remain on the system to which they are assigned until they are removed from the system.

► New systems can be added to the pool and existing systems can be removed from the pool.

► New Mobile CoD processor and memory resources can be added to the pool.

► Pool information can be viewed, including pool resource assignments, compliance, and history logs.

For the mobile activation features to be configured, an IBM Power Enterprise Pool and the systems that are going to be included as members of the pool must be registered with IBM. Also, the systems must have #EB35 for mobile enablement configured, and the required contracts must be in place.

> **Note:** The following convention is used in the Order type column in the tables in this section:
>
> ► Initial: Only available when ordered as part of a new system
>
> ► MES: Only available as a Miscellaneous Equipment Specification (MES) upgrade
>
> ► Both: Available with a new system or as part of an upgrade
>
> ► Supported: Unavailable as a new purchase, but supported when migrated from another system or as part of a model conversion

Table 2-10 lists the mobile processor and memory activation features that are available for the Power E1080 server.

Table 2-10   Mobile activation features

Feature code	Description	Maximum	Order[a] type
EDAB	100 GB Mobile Memory activation for HEX	480	Both
EMBK	500 GB Mobile Memory activation for HEX	130	Both
EDPZ	Mobile processor activation for HEX	224	Both

  a. For more information about order types, see 2.4, "Capacity on-demand" on page 76.

For more information about IBM Power Enterprise Pools, see *Power Enterprise Pools on IBM Power Systems*, REDP-5101.

2.4.7 Utility Capacity on-Demand

Utility CoD is *not* supported on Power E1080 server.

2.4.8 Trial Capacity on-Demand

A *standard request* for Trial CoD requires you to complete a form that includes contact information and vital product data (VPD) from your Power E1080 server with inactive CoD resources.

A standard request activates eight processors or 64 GB of memory for 30 days. Subsequent standard requests can be made after each purchase of a permanent processor activation. An HMC is required to manage Trial CoD activations.

An *exception request* for Trial CoD requires you to complete a form that includes contact information and VPD from your Power E1080 server with inactive CoD resources. An exception request activates all inactive processors or all inactive memory (or all inactive

processor and memory) for 30 days. An exception request can be made only once over the life of the machine. An HMC is required to manage Trial CoD activations.

To request either a Standard or an Exception Trial, see Power Systems Capacity on Demand: Trial Capacity on Demand.

2.4.9 Software licensing and CoD

For software licensing considerations for the various CoD offerings, see the most recent revision of the *Power Systems Capacity on Demand User's Guide*.

2.5 Internal I/O subsystem

This section provides more information about the internal PCIe architecture of the Power E1080 system.

2.5.1 Internal PCIe Gen 5 subsystem and slot properties

The internal I/O subsystem on the Power E1080 server is connected to the PCIe Gen 5 controllers on a Power10 chip in the system. Each Power10 processor module has two PCIe host bridges (PHBs). The PHBs (PHB0 and PHB1) have 16 PCIe lanes each.

The PHBs provide different type of end point connections. Both the PHBs from Power10 sockets P0 and P1 and one PHB from sockets P2 and P3 connect directly to a PCIe Gen 4 x16 or PCIe Gen 5 x8 slot to provide six PCIe Gen 4 x16 or PCIe Gen 5 x8 slots per system node.

The second PHB on each Power10 socket P2 and P3 is split between a PCIe Gen 5 x8 slot and two Internal Non-Volatile Memory Express (NVMe) SSD PCIe Gen 4 x4 slots. The design overall provides each system node of Power E1080 with:

► Six PCIe Gen 4 x16 or PCIe Gen 5 x8
► Two PCIe Gen 5 x8
► Four Internal Non-Volatile Memory Express (NVMe) SSDs, each using a PCIe Gen 4 x4 connection

Bandwidths for the connections are listed in Table 2-11.

Table 2-11   Internal I/O connection speeds

Connection	Type	Speed (duplex)
PCIe adapter slot	PCIe Gen4 x16	64 GBps
PCIe adapter slot	PCIe Gen5 x8	64 GBps
NVMe slot	PCIe Gen4 x4	16 GBps

The following theoretical maximum I/O bandwidths in each system node are available:

► Six PCIe Gen 4 x16 / PCIe Gen 5 x8 slots at 64 GBps = 384 GBps
► Two PCIe Gen 5 x8 Slots at 64 GBps = 128 GBps
► Four NVMe slots at 16 GBps = 64 GBps

Total of 576 GBps = 384 GBps + 128 GBps + 64 GBps.

The rear view of a Power E1080 system node is as shown in Figure 2-16 and Figure 2-17.
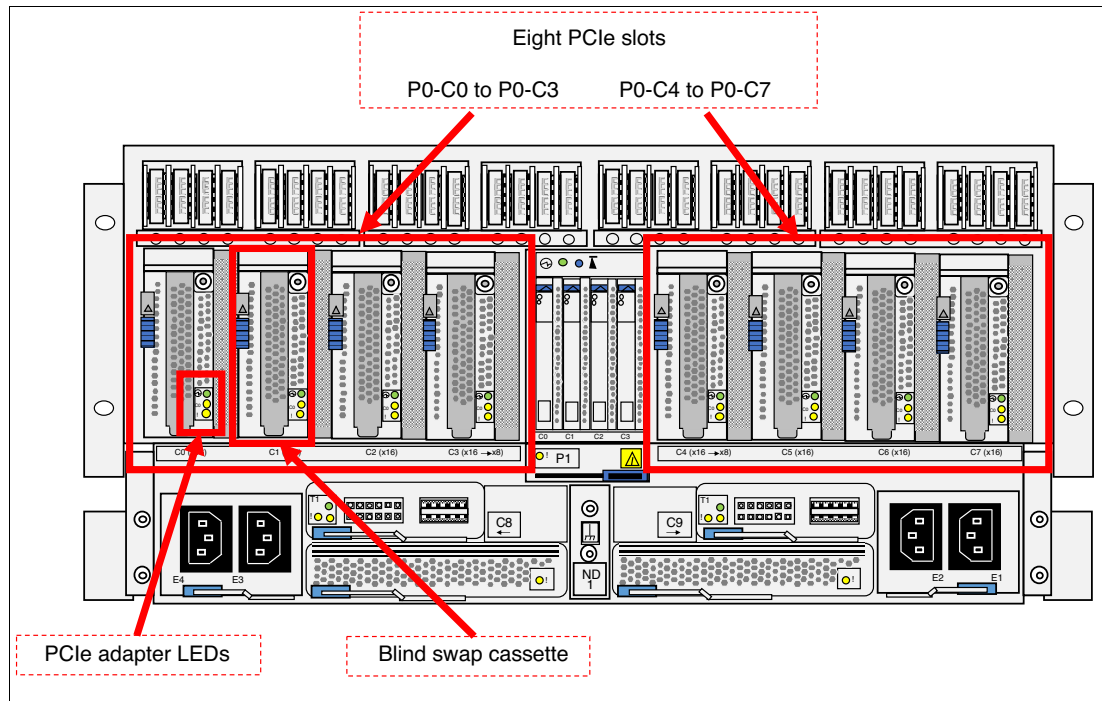

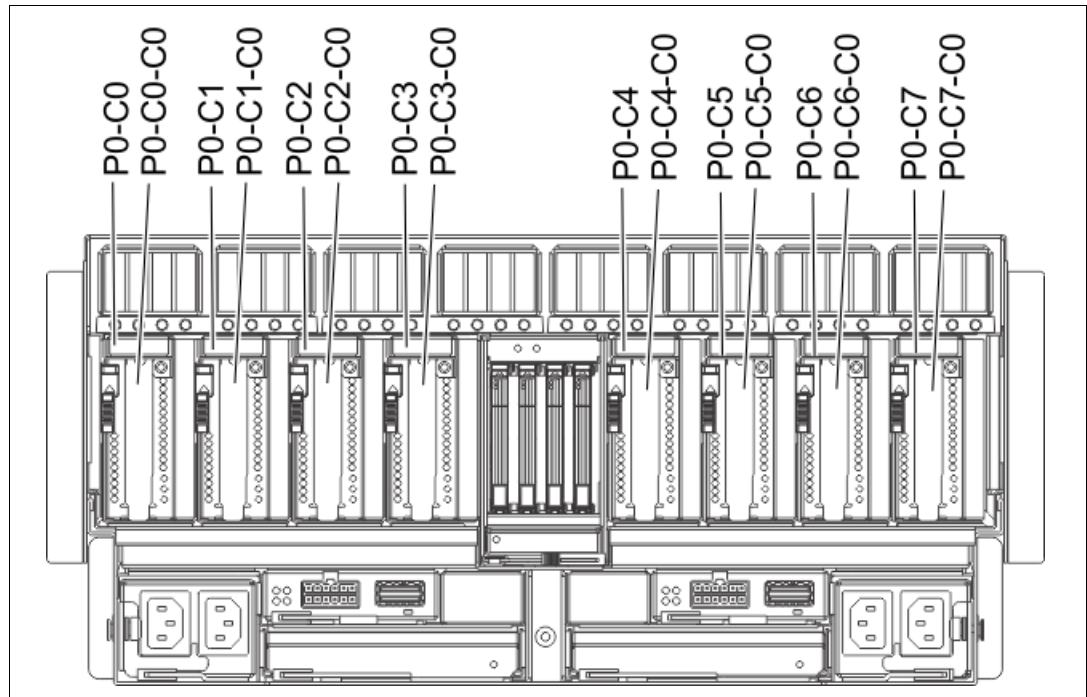*Figure 2-16   Rear view of a Power E1080 server node highlighting PCIe slots*

Figure 2-17   Rear view of a E1080 server node with PCIe slot location codes

Adapters that are used in the Power E1080 server are enclosed in a specially designed adapter cassette. The cassette includes a special mechanism for holding the adapter within a

socket and helps for easy insertion and removal of the adapter into a PCIe slot at the rear side of the server.

The cassette locations are P0-C0 - P0-C7. The adapter locations are P0-C0-C0 - P0-C7-C0.

All slots are of low-profile (LP) type; that is, half-height, half-length (HHHL). They support Enhanced Error Handling (EEH) and can be serviced with the system power turned on. The slots also support Gen 1 to Gen 3 adapters.

All the PCIe slots support Single Root I/O Virtualization (SR-IOV) adapters.

Currently, no adapter can dissipate more than 25 watts. Special considerations are needed to exceed this limit.

The Power E1080 server supports concurrent maintenance (hot plugging) of PCIe adapters. The adapter cassette is provided with three LEDs that can indicate power and activity (green), identify function (amber) for the adapter (labeled C0), and fault function LED (amber) for the cassette. The server can be located by using the blue identify LED on the enclosure.

The internal connections are shown in Figure 2-18.
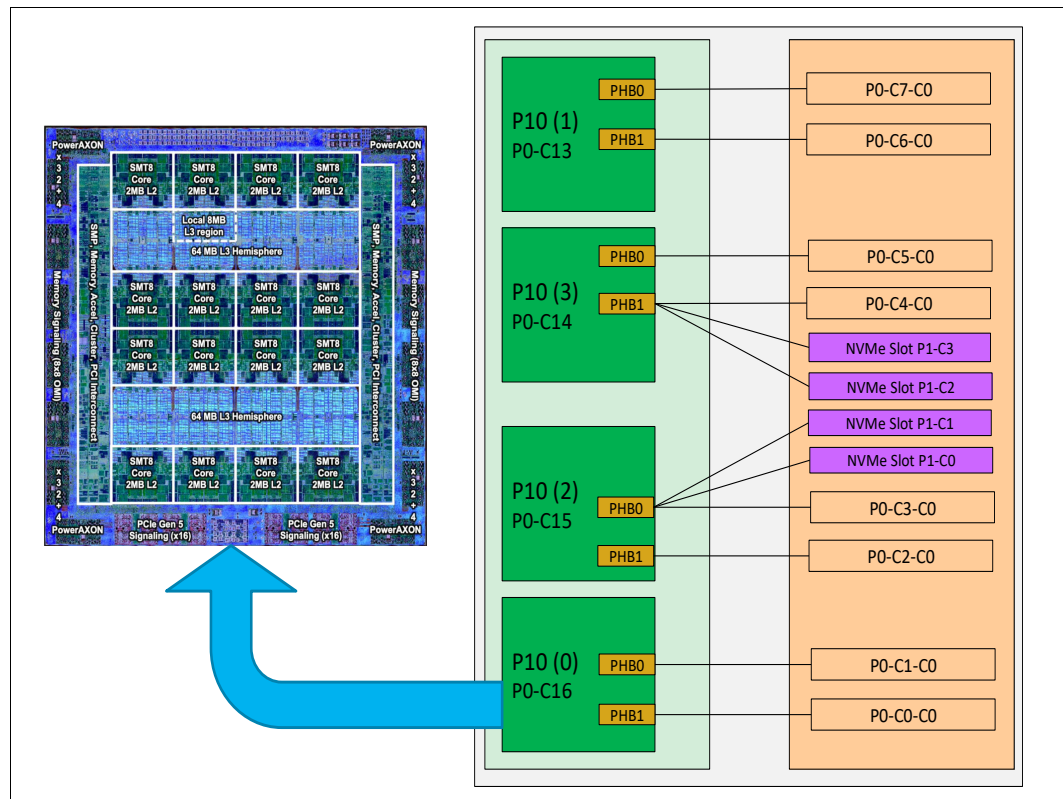

Figure 2-18   Power E1080 server node to PCIe slot internal connection schematics

The PCIe slots are numbered from P0-C0 - P0-C7.

Slot locations and descriptions for the Power E1080 servers are listed in Table 2-12 on page 86.

Table 2-12   Internal PCIe Slot number and Location codes

Location Code	Description	Socket / PHB	Slot Number	SRIOV
P0-C0	PCIe Gen 4 x16 or PCIe Gen 5 x8	P0/1	1	Yes
P0-C1	PCIe Gen 4 x16 or PCIe Gen 5 x8	P0/0	2	Yes
P0-C2	PCIe Gen 4 x16 or PCIe Gen 5 x8	P2/1	3	Yes
P0-C3	PCIe Gen 5 x8	P2/0	4	Yes
P1-C0	PCIe Gen 4 x4	P2/0	NVMe Slot 1	NA
P1-C1	PCIe Gen 4 x4	P2/0	NVMe Slot 2	NA
P1-C2	PCIe Gen 4 x4	P3/1	NVMe Slot 3	NA
P1-C3	PCIe Gen 4 x4	P3/1	NVMe Slot 4	NA
P0-C4	PCIe Gen 5 x8	P3/1	5	Yes
P0-C5	PCIe Gen 4 x16 or PCIe Gen 5 x8	P3/0	6	Yes
P0-C6	PCIe Gen 4 x16 or PCIe Gen 5 x8	P1/1	7	Yes
P0-C7	PCIe Gen 4 x16 or PCIe Gen 5 x8	P1/0	8	Yes

PCIe x16 cards are supported only in PCIe x16 slots and the slot priority is (1, 3, 7, 2, 6, 8).

PCIe adapters with x8 and lower lanes are supported in all the PCIe slots and the slot priority is (1, 7, 3, 6, 2, 8, 4, 5).

Place 8x and 16x adapters in same size slots first before mixing connector size with slot size. Adapters with smaller connectors are allowed in larger PCIe slot sizes, but larger connectors are not compatible in smaller PCIe slots sizes.

The system nodes allow for eight PCIe slots of which six are PCIe Gen4 x16 or PCIe Gen5 x8 slots and two are PCIe Gen5 x8 slots. Slots can be added by attaching PCIe expansion drawers; SAS disks can be attached to EXP24S SFF Gen2 expansion drawers. At the time of this writing, only the following expansion drawers are supported:

► EMX0 PCIe Gen3 I/O expansion drawer (#EMX0)
► EXP24SX I/O DASD drawer (#ESLS)

The PCIe expansion drawer is connected by using an #EJ24 adapter. The EXP24S storage enclosure can be attached to SAS adapters on the system nodes or on the PCIe expansion drawer.

For a list of adapters and their supported slots, see 2.6, "Supported PCIe adapters" on page 92.

> **Disk support:** SAS disks that are directly installed on the system nodes and PCIe Expansion Drawers are not supported. If directly attached SAS disks are required, they must be installed in a SAS disk drawer and connected to a supported SAS controller in one of the PCIe slots.

The adapters that do not require high bandwidth are recommended be placed in an EMX0 PCIe Gen 3 expansion drawer. Populating a low latency, high-bandwidth slot with a low-bandwidth adapter is not the best use of system resources.

IBM recommends the use of the high-profile version of all adapters whenever the system includes an EMX0 PCIe Gen 3 expansion drawer. This configuration allows the user to place most of the adapters within the EMX0 PCIe Gen 3 expansion drawer. Typically, it is advised to use node slots for the cable adapters (#EJ24) or high bandwidth PCIe Gen 4 or PCIe Gen 5 adapters.

2.5.2 Internal NVMe storage subsystem

The Power E1080 supports NVMe storage technology. The NVMe subsystem is available in each of the E1080 server node drawers. Four internal NVMe U.2 drive slots are present in each of the system nodes in the center of the rear of the system. The slots can be independently assigned to any of the partitions (see Figure 2-19).

4 x 2.5" 7mm NVMe SSDs or 2 x 2.5" 15mm NVMe SSDs

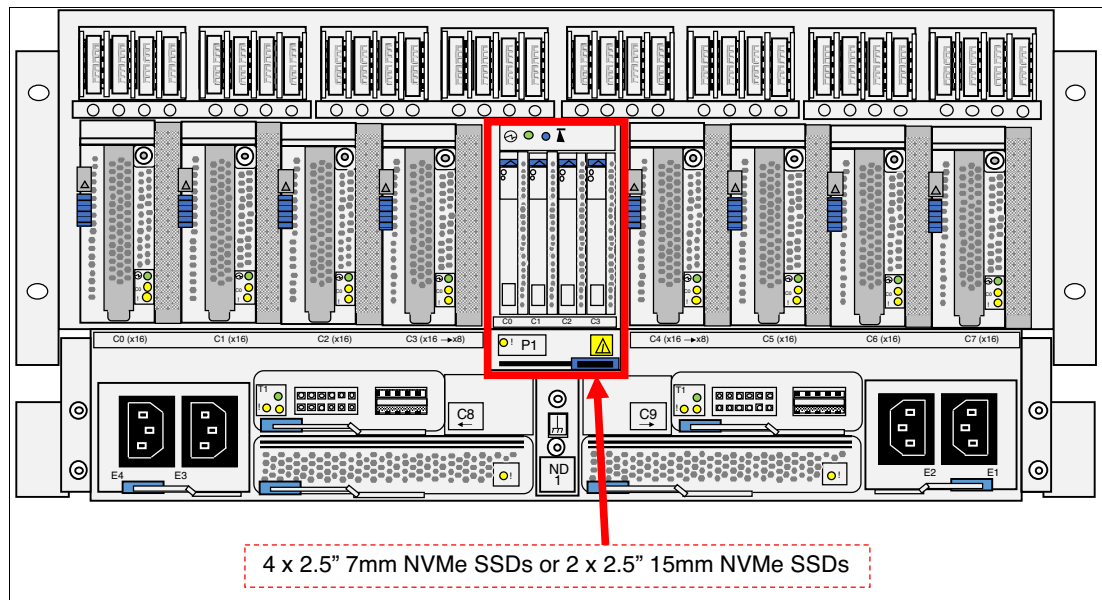*Figure 2-19   Rear view of a E1080 server node highlighting NVMe slots*

As of this writing, no feature codes support IBM i for the internal NVMe subsystem. An external drawer is needed in this case.

The internal 7mm NVMe devices support only one namespace. The primary use of the devices is to host Virtual I/O Servers.

The location of the NVMe U.2 drives in the system node is shown in Figure 2-20 on page 88.
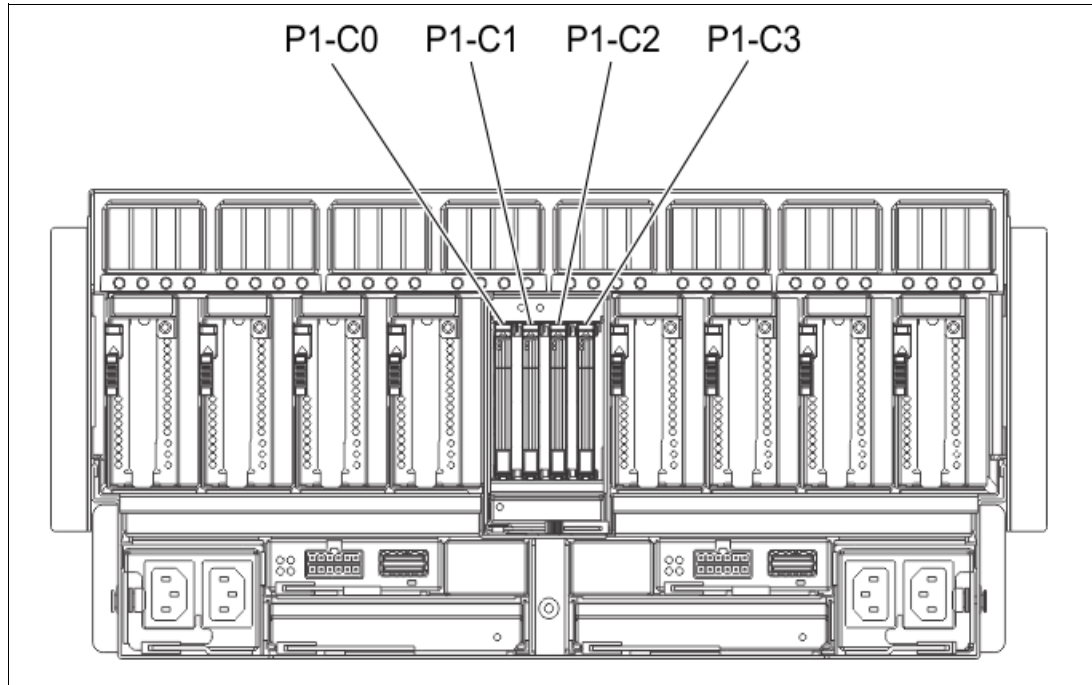
Figure 2-20   Rear view of a Power E1080 server node with NVMe slot location codes

Each slot is driven from a x4 PCIe connection, with 2 SSDs connected to the PHB on the P2 CPU socket and 2 SSDs connected to the PHB of the P3 CPU socket as shown in Figure 2-18 on page 85. The busses are routed through system planar from the CPU to the NVMe slot interface card, and then up through the SSD riser card and into the SSD.

The NVMe slots support U.2 NVMe flash SSD drives. Four 2.5-inch 7 mm form factor SSDs or two 2.5-inch 15 mm NVMe SSDs can be used.

Slot locations P1-C2 and P1-C3 remain empty if #EJBD backplane is selected to allow for spacing of 15 mm option.

Choose the type of NVMe SSD drives carefully because some drives are intended to be used only as a boot drive and for non-data-intensive workloads.

Table 2-13 lists the internal NVMe adapters that are available as of this writing.

Table 2-13   Available internal NVMe SSD features codes

Feature code	CCIN	Description	Max.	OS support
EC5J	59B4	Mainstream 800 GB SSD NVMe U.2 module	16	AIX, and Linux
EC5K	59B5	Mainstream 1.6 TB SSD NVMe U.2 module	16	AIX, and Linux
EC5L	59B6	Mainstream 3.2 TB SSD NVMe U.2 module	16	AIX, and Linux
EC7Q	59B4	800 GB Mainstream NVMe U.2 SSD 4k for AIX/Linux	16	AIX, and Linux

The (#EC5J/#EC5K/#EC5L) - 800 GB/1.6 TB/3.2 TB Mainstream SSDs are rated for a limited number of Drive Write per Day (DWPD) and as such are intended for use as a boot drive and for non-data-intensive workloads.

Internal SSD plug order

For redundancy purposes, it is a preferred practice to distribute the NVMe drives across the system nodes if they are present as follows:

▶ Populate slot C0 in each system node starting with node 1 and then, other C0 slots if other nodes are present.

▶ Populate slot C2 in each system node starting with node 1 and then, other C2 slots if other nodes are present.

▶ Populate slot C1 in each system node starting with node 1 and then, other C1 slots if other nodes are present.

▶ Populate slot C3 in each system node starting with node 1 and then, other C3 slots if other nodes are present.

Each NVMe U.2 drive has two LEDs at the top of the drive that indicate the following status:

▶ A power and activity LED (green)
▶ An error and identify function LED (amber).

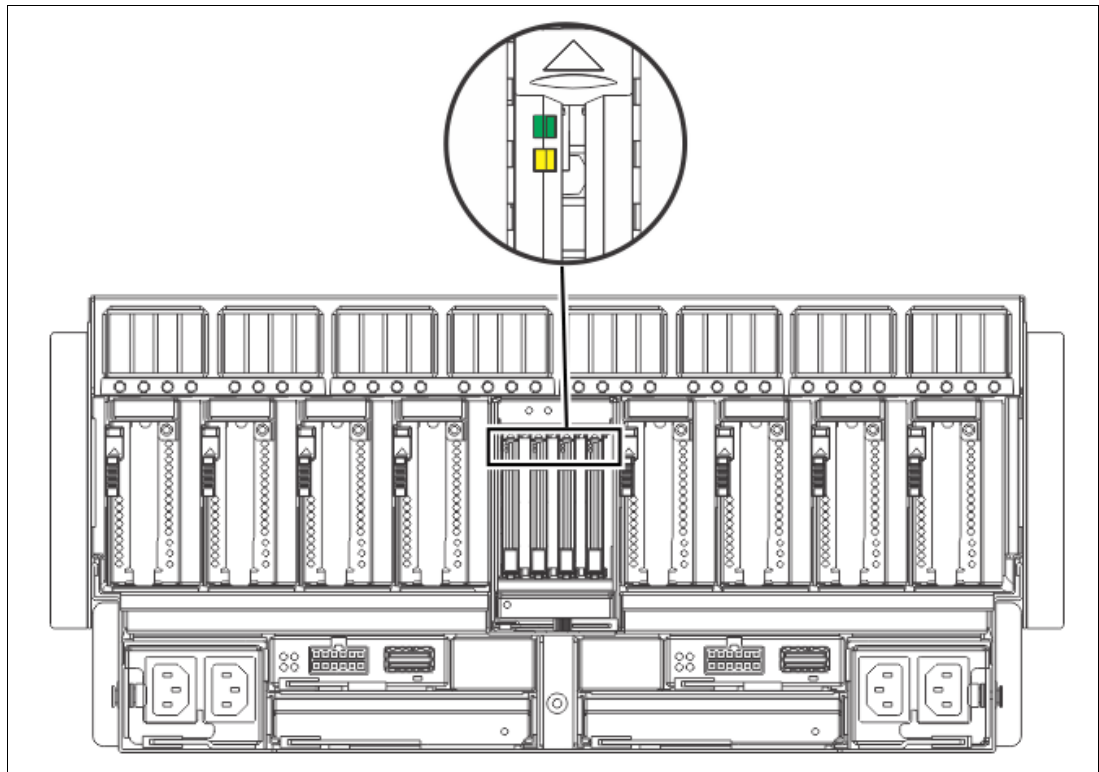The location of the NVMe U.2 drive LEDs in the system are shown in Figure 2-21.

Figure 2-21   Rear view of a E1080 server node with NVMe slot LED location

Concurrent maintenance of the NVMe drives are supported.

You can find the remaining life on a NVMe device from the logical partition that owns the device.

To determine the remaining life of an NVMe device, complete the following steps:

► For an IBM AIX operating system:

   a. From the AIX command line, enter `diag` and press **Enter**.

   b. From the Function Selection menu, select **Task Selection** → **NVMe general health information**.

   c. Select the NVMe device that you want to check the remaining life for and press **Enter**.

   d. View the Percentage of NVM subsystem life used field.

► For Linux operating system.

   a. From the Linux command line, enter the following command and press **Enter**:

      `nvme smart-log /dev/`*nvmeX* `-H`

      Where *nvmeX* is the resource name of the NVMe device.

   b. View the Percentage used field.

The NVMe device that is nearing its end of life must be replaced. The device soon reaches the limit for the number of write operations that are supported. Write operations to the NVMe device become slower over time, and at some point the NVMe device becomes a read-only device.

When the operating system writes data to a read-only device, the write operations are rejected, and the operating system considers the device as if a failure occurred. To support normal write operations, the NVMe device must be replaced.

2.5.3 USB subsystem

The Power E1080 has no integrated USB controlled or ports in the system nodes. The USB PCIe adapter must be used for enabling USB access. The PCIe2 2-PORT USB 3.0 LP adapter (#EC6J) is supported in the system node. This adapter typically is placed in the PCIe Slot 4/5 of system node1 or node2.

A USB Cable (#EC6N) is used for connecting the #EC6J adapter port to the rear side USB port in the system control unit (SCU), which is internally routed to a front-accessible USB port on the SCU (see Figure 2-22).

Figure 2-22   Rear view of the SCU showing the USB port for connection

The following USB devices are supported and available for the Power E1080:

► #EUA5 - Standalone USB DVD Drive with cable
► #EUA4 - RDX USB External Docking Station

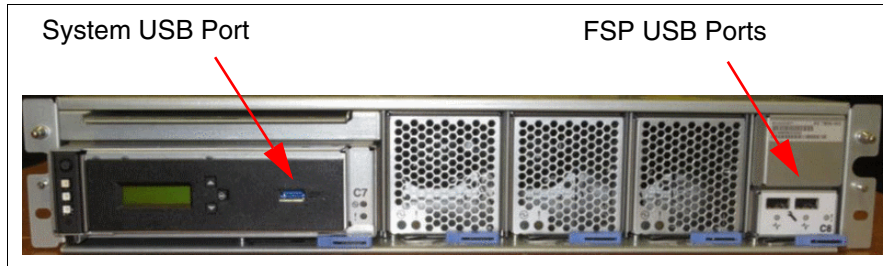The front view of SCU in which the System USB port and FSP USB ports are highlighted is shown in Figure 2-23.

Figure 2-23   Front view of SCU showing System USB port and FSP USB ports

2.5.4 PCIe slots features

Each system node provides 8 PCIe Gen5 hot-plug enabled slots; therefore, a 1-node system provides eight slots, a 2-node system provides 16 slots, a 3-node system provides 24 slots, and a 4-node system provides 32 slots.

You can connect up to four I/O expansion drawer features EMX0 per node with the slot capacity that is listed in Table 2-14.

Table 2-14   PCIe slots availability for different system nodes configurations

System node number	I/O expansion number	LP slots	Full height slots
1	4	8	48
2	8	16	96
3	12	24	144
4	16	32	192

Each I/O expansion drawer consists of two Fanout Module features EMXH, each providing six PCIe slots. Each Fanout Module connects to the system by using a pair of CXP cable features. The CXP cable features are listed in Table 2-15. The RPO Only cables in this list are not available for ordering new or MES upgrade but for migration from a source system. Select a longer length feature for inter rack connection between the system node and the expansion drawer.

Table 2-15   Optical CXP cable feature

Feature code	CCIN	Description	OS support	Order type
ECCR		2M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both
ECCY		10M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both

Feature code	CCIN	Description	OS support	Order type
ECCZ		20M Active Optical Cable Pair for PCIe Gen 3 Expansion Drawer		Both

The one pair of CXP optical cables connect to a system node by using one 2-ports PCIe optical cable adapter feature EJ24 placed in the CEC.

The CXP optical cable pair and the optical cable adapter features are concurrently maintanable. Therefore, careful balancing of I/O, assigning adapters through redundant EMX0 expansion drawers, and different system nodes can ensure high-availability for the I/O resources assigned to partitions.

Note: At the time of this writing, feature EMXH is the only supported or orderable fanout module. Previous fanout module features EMXF and EMXG are not supported on Power E1080 system.

For more information about internal buses and the architecture of internal and external I/O subsystems, see 2.5, "Internal I/O subsystem" on page 83.

2.6 Supported PCIe adapters

In this section, we provide more information about supported PCIe I/O adapters on Power E1080 server as on system general availability date September 17th, 2021. This list tends to update frequently as more PCIe I/O adapters are tested and certified. The latest list of supported adapters through announcement letters, is available at this web page.

The Order type table column in the following subsections is defined as:

Initial   Denotes the orderability of a feature for *only* the new purchase of the system

MES   Denotes the orderability of a feature for *only* the MES upgrade purchases on the system

Both   Denotes the orderability of a feature for *both* new and MES upgrade purchases

Supported   Denotes that feature is *not* orderable, but is *supported*; that is, the feature can be migrated only from existing systems.

2.6.1 LAN adapters

The supported LAN adapters are listed in Table 2-16.

Table 2-16   LAN Adapters

Feature code	CCIN	Description	Min	Max	OS support	Order type
5260	576F	PCIe2 LP 4-port 1 GbE Adapter	0	32	AIX, IBM i, and Linux	Supported
5899	576F	PCIe2 4-port 1 GbE Adapter	0	192	AIX, IBM i, and Linux	Supported
EC2T	58FB	PCIe Gen 3 LP 2-Port 25/10 Gb NIC&ROCE SR/Cu Adapter[a]	0	32	AIX, IBM i, and Linux	Both
EC2U	58FB	PCIe Gen 3 2-Port 25/10 Gb NIC&ROCE SR/Cu Adapter[a]	0	192	AIX, IBM i, and Linux	Both
EC67	2CF3	PCIe Gen 4 LP 2-port 100 Gb ROCE EN LP Adapter	0	24	AIX, IBM i, and Linux	Both
EC77	2CFA	PCIe4 LP 2-port 100 Gb Crypto Connectx-6 DX QFSP56	0	24	AIX, IBM i, and Linux	Both
EN0S	2CC3	PCIe2 4-Port (10 Gb+1 GbE) SR+RJ45 Adapter	0	192	AIX, IBM i, and Linux[b]	Supported
EN0T	2CC3	PCIe2 LP 4-Port (10 Gb+1 GbE) SR+RJ45 Adapter	0	32	AIX, IBM i, and Linux[a]	Supported
EN0W	2CC4	PCIe2 2-port 10/1 GbE BaseT RJ45 Adapter	0	192	AIX, IBM i, and Linux[b]	Both
EN0X	2CC4	PCIe2 LP 2-port 10/1 GbE BaseT RJ45 Adapter	0	32	AIX, IBM i, and Linux[a]	Both
EC2R	58FA	PCIe3 LP 2-Port 10 Gb NIC&ROCE SR/Cu Adapter	0	32	AIX, IBM i, and Linux	Supported
EC2S	58FA	PCIe3 2-Port 10 Gb NIC&ROCE SR/Cu Adapter	0	192	AIX, IBM i, and Linux	Supported

a. Requires SFP to provide 10 Gb, 2 Gb, or 1 Gb BaseT connectivity
b. IBM i supported through VIOS

2.6.2 Fibre Channel adapters

The supported Fibre Channel adapters are listed in Table 2-17.

Table 2-17   Fibre Channel adapters

Feature code	CCIN	Description	Min	Max	OS support	Order type
EN1A	578F	PCIe Gen 3 32 Gb 2-port Fibre Channel Adapter	0	192	AIX, IBM i, and Linux	Both

Feature code	CCIN	Description	Min	Max	OS support	Order type
EN1B	578F	PCIe Gen 3 LP 32 Gb 2-port Fibre Channel Adapter	0	32	AIX, IBM i, and Linux	Both
EN1C	578E	PCIe Gen 3 16 Gb 4-port Fibre Channel Adapter	0	192	AIX, IBM i, and Linux	Both
EN1D	578E	PCIe Gen 3 LP 16 Gb 4-port Fibre Channel Adapter	0	32	AIX, IBM i, and Linux	Both
EN1E	579A		0	192	AIX, IBM i, and Linux[a]	Both
EN1F	579A	PCIe Gen 3 LP 16 Gb 4-port Fibre Channel Adapter	0	32	AIX, IBM i, and Linux[a]	Both
EN1G	579B		0	192	AIX, Linux	Both
EN1H	579B	PCIe Gen 3 LP 2-Port 16 Gb Fibre Channel Adapter	0	32	AIX, Linux	Both
EN1J	579C		0	192	AIX, IBM i, and Linux	Both
EN1K	579C	PCIe Gen 4 LP 32 Gb 2-port Optical Fibre Channel Adapter	0	32	AIX, IBM i, and Linux	Both
EN2A	579D	PCIe Gen 3 16 Gb 2-port Fibre Channel Adapter	0	192	AIX, IBM i, and Linux	Both
EN2B	579D	PCIe Gen 3 LP 16 Gb 2-port Fibre Channel Adapter	0	32	AIX, IBM i, and Linux	Both

a. IBM i supported through VIOS

2.6.3 SAS adapters

The supported SAS adapters are listed in Table 2-18.

Table 2-18   SAS adapters

Feature code	CCIN	Description	Min	Max	OS support	Order type
EJ0J	57B4	PCIe Gen 3 RAID SAS Adapter Quad-port 6 Gb x8	0	128	AIX, IBM i, and Linux	Both
EJ0M	57B4	PCIe Gen 3 LP RAID SAS Adapter Quad-Port 6 Gb x8	0	32	AIX, IBM i, and Linux	Both

Feature code	CCIN	Description	Min	Max	OS support	Order type
EJ10	57B4	PCIe Gen 3 SAS Tape/DVD Adapter Quad-port 6 Gb x8	0	128	AIX, IBM i, and Linux	Both
EJ11	57B4	PCIe Gen 3 LP SAS Tape/DVD Adapter Quad-port 6 Gb x8	0	32	AIX, IBM i, and Linux	Both
EJ14	57B1	PCIe Gen 3 12 GB Cache RAID PLUS SAS Adapter Quad-port 6 Gb x8	0	128	AIX, IBM i, and Linux	Both

2.6.4 Crypto adapter

The supported crypto adapter is listed in Table 2-19.

Table 2-19   Crypto adapters

Feature code	CCIN	Description	Min	Max	OS support	Order type
EJ37	C0AF	PCIe Gen 3 Crypto Coprocessor BSC-Gen3 4769	0	192	AIX, IBM i, and Linux	Both

2.6.5 USB adapter

The supported USB adapter is listed in Table 2-20.

Table 2-20   USB adapter

Feature code	CCIN	Description	Min	Max	OS support	Order type
EC6J	590F	PCIe2 LP 2-Port USB 3.0 Adapter	0	32	AIX, IBM i, and Linux	Both

2.6.6 I/O expansion drawers

The supported I/O expansion drawers are listed in Table 2-21.

Table 2-21   I/O expansion drawer feature

Feature code	CCIN	Description	Min	Max	OS support	Order type
EMX0		PCIe Gen3 I/O Expansion Drawer	0	16	AIX, IBM i, and Linux	Both
EMXH	50CD	PCIe Gen 3 6-Slot Fanout Module for PCIe Gen 3 Expansion Drawer	0	32	AIX, IBM i, and Linux	Both

2.6.7 Disk drawer

The supported disk drawer is listed in the Table 2-22.

Table 2-22   Disk drawer feature

Feature code	CCIN	Description	Min	Max	OS support	Order type
ESLS	78D1	EXP24SX SAS Storage Enclosure	0	168	AIX, IBM i, and Linux	Both

2.6.8 SFP transceiver

The supported SFP transceivers are listed in Table 2-23. The SFP provides connectivity to the #EC2T and EC2U LAN adapters.

Table 2-23   SFP / QSFP features

Feature code	CCIN	Description	Min	Max	OS support	Order type
EB46		10 GbE Optical Transceiver SFP+ SR	0	9999		Both
EB47		25 GbE Optical Transceiver SFP28	0	9999		Both
EB48		1 GbE Base-T Transceiver RJ45	0	9999	AIX, IBM i, and Linux	Both
EB49		QSFP28 to SFP28 Connector	0	32		Both
EB57		QSFP+ 40 GbE Base-SR4 Transceiver	0	32		Both
EB59		100 GbE Optical Transceiver QSFP28	0	9999		Both

2.7 External I/O subsystems

The I/O capacity of the Power E1080 can be expanded by using extra external PCIe Expansion Drawers and external SAS Disk Expansion Drawers.

This section describes the PCIe Gen3 I/O Expansion Drawer (#EMX0) that can be attached to the Power E1080 server.

At the initial availability, the Power E1080 supports four PCIe Gen3 I/O drawer per system node, which yields a maximum of eight I/O drawers per 2-node Power E1080 server configuration. One I/O drawer supports two Fanout Modules that offer six PCIe Gen 3 adapter slots each. This configuration delivers an extra 48 PCIe Gen 3 slot capacity per system node and a maximum of 96 PCIe Gen 3 slots per 2-node server.

All eight slots in a system node can be used to cable the four I/O drawers. It is always recommended to balance the I/O Drawer connectivity across the available system nodes.

With the availability of 3-node and 4-node Power System E1080 configurations in later GA period, the number of PCIe Gen 3 I/O drawers scale accordingly with a maximum of 16 per 4-node Power E1080. A maximum of 48 PCIe Gen 3 slots per system node and a maximum of 192 PCIe Gen 3 slots per 4-node Power E1080 server to be available at that date.

Each Fanout Module in the PCIe Gen 3 I/O Expansion Drawer (#EMX0) is attached by one optical cable adapter, which occupies one x16 PCIe Gen 4 slot of a system node.

The maximum number of supported I/O drawers and the total number of PCIe slots are listed in Table 2-24.

Table 2-24   Maximum number of supported I/O drawers and the total number of PCIe slots

System nodes	Maximum #EMX0 drawers	Total number of slots		
		PCIe Gen 3, x16	PCIe Gen 3, x8	Total PCIe Gen 3
One system node	4	16	32	48
Two system nodes	8	32	64	96
Three system nodes	12	48	96	144
Four system nodes	16	64	128	192

2.7.1 PCIe Gen 4 cable adapter

The PCIe Gen3 I/O Expansion Drawer (#EMX0) is connected to the system node by using a PCIe Gen 4 I/O expansion card (#EJ24) that is placed in any of the PCIe slots with in the system node. One expansion card adapter connects to one PCIe Gen 3 6-slot Fanout Module (#EMXH) of the PCIe Gen3 I/O Expansion Drawer(#EMX0) using a CXP 16X Active Optical Cable (AOC) Pair. Therefore, the adapter is also named as PCIe Gen 4 CXP cable adapter.

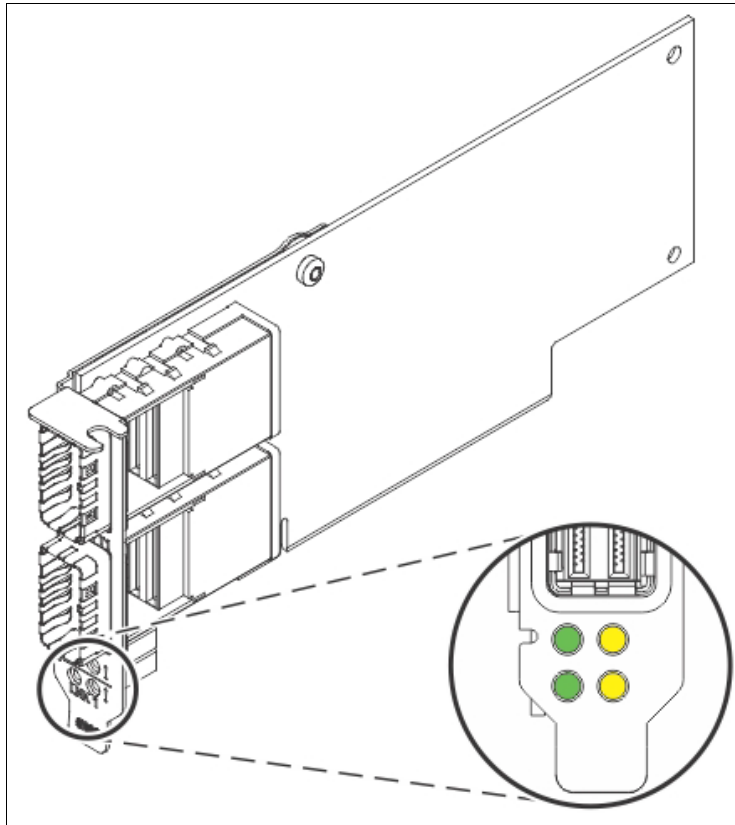The PCIe Gen 4 cable adapter is shown in Figure 2-24.

Figure 2-24   PCIe Gen 4 cable adapter (#EJ24)

The PCIe Gen 4 cable adapter (#EJ24) is a half-height, half-length, low-profile PCIe Gen 4 x16 adapter. The adapter provides two ports for the attachment of CXP 16x AOC pair. One adapter supports the attachment of one PCIe Gen 3 6-slot fanout module in a EMX0 PCIe Gen3 I/O expansion drawer.

Although these cables are not redundant, the loss of one cable reduces the I/O bandwidth (that is, the number of lanes that are available to the I/O module) by 50%.

A total of eight PCIe Gen 4 cable adapters (#EJ24) are allowed per system node to connect to maximum of eight PCIe Gen 3 6-slot Fanout Modules (#EMXH). Each Fanout Module must go with one CXP 16x AOC cable pair with both ports wired to the same PCIe Gen 4 cable adapter from the system node.

The cable adapter can be placed in any slot in the system node. The following adapter placement sequence in the system node is recommended:

```
System node slot 1, 7, 3, 6, 2, 8, 4, 5
```

The maximum number of adapters that is supported is 8 per system node and 32 per system, having four system nodes.

Whenever PCIe Gen3 I/O Expansion Drawer(#EMX0) is used, it is recommended to reserve the slots in the system nodes for the PCIe Gen 4 cable adapter (#EJ24) adapters. Place the high-profile adapters within the I/O Expansion Drawers.

Three different types of CXP 16x AOC cable pairs are available for the connection between PCIe Gen 4 cable adapter (#EJ24) and PCIe Gen 3 fanout module (#EMXH) of the I/O Expansion Drawer:

► 2.0M (#ECCR)

  This cable is used when I/O drawer is in the *same* rack as the E1080 server.

► 10.0M (#ECCY)

  This cable is used when the I/O drawer is in a *different* rack from the E1080 server.

► 20.0M (#ECCZ)

► Like the 10.0M (#ECCY) cable, these cables are used when I/O drawer is in a different rack from the E1080 server.

> **Note:** One #ECCR, one #ECCY, or one #ECCZ includes two AOC cables.

2.7.2 PCIe Gen3 I/O Expansion Drawer

The PCIe Gen3 I/O Expansion Drawer (#EMX0) is a 4U high, PCIe Gen3-based and rack-mountable I/O drawer. It offers two PCIe Fanout Modules (#EMXH), each providing six PCIe Gen3 full-high, full-length slots (two x16 and four x8).

The older Fanout Modules (#EMXF & #EMXG) that are used by Power E870, Power E870C, Power E880, and Power E880C systems cannot be ordered and is not supported with the Power system E1080.

For more information about the dimensions of the drawer, see 1.6, "I/O drawers" on page 26.

PCIe Gen 4 cable adapter (#EJ24) CXP 16X AOCs connects the system node to a PCIe Fanout Module in the I/O expansion drawer.

Concurrent repair and addition or removal of PCIe adapters is done by using HMC-guided menus or operating system support utilities.

A blind-swap cassette (BSC) is used to house the full-high adapters that are installed into these slots. The BSC is the same BSC that was used with the previous generation server's #5802, #5803, #5877, and #5873 12X attached I/O drawers.

Power consumption is up to 50W per adapter slot.

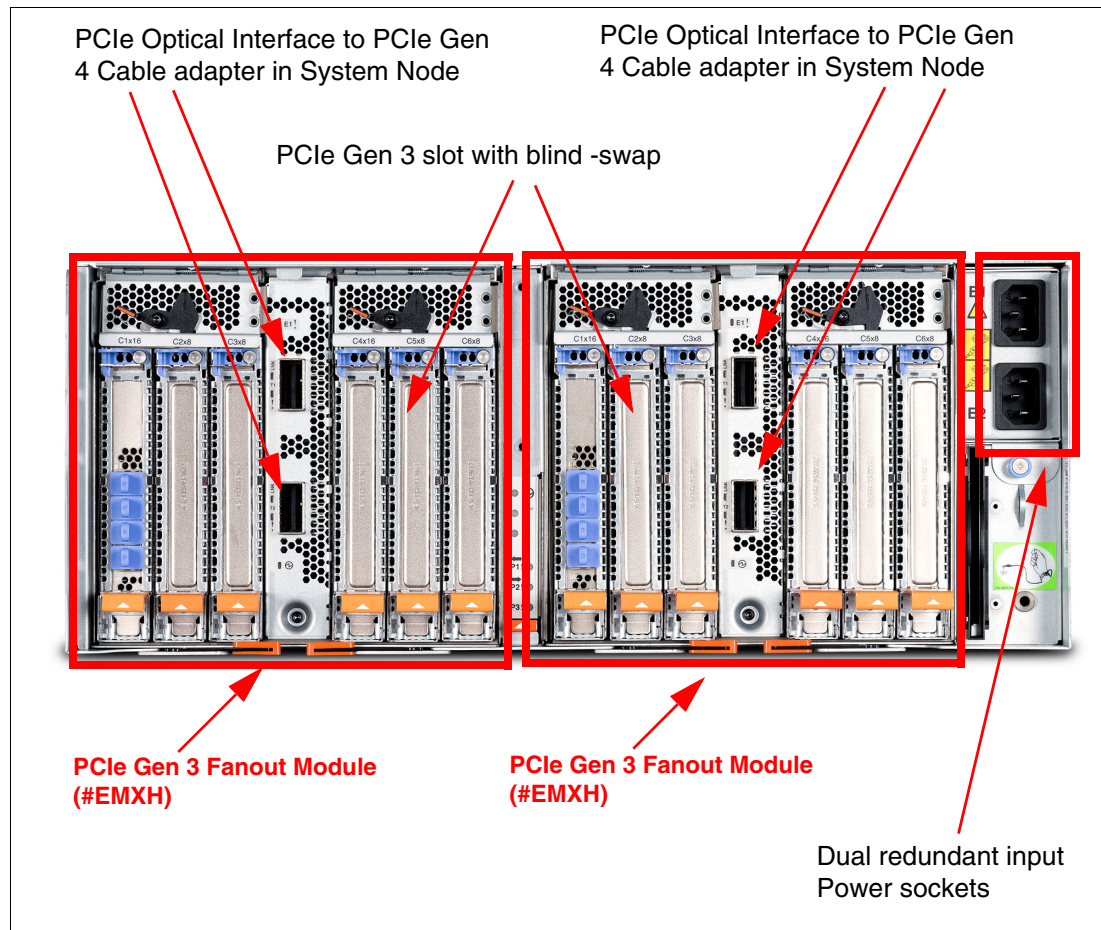Figure 2-25 shows the rear view of the PCIe Gen3 I/O Expansion Drawer.

Figure 2-25 Rear view of the PCIe Gen3 I/O Expansion Drawer

2.7.3 PCIe Gen3 I/O Expansion Drawer optical cabling

I/O drawers are connected to the adapters in the system node by any of the following data transfer cables:

► 2.0 m CXP 16x Active Optical Cable (AOC) Pair (#ECCR)
► 10.0 m CXP 16x Active Optical Cable (AOC) Pair (#ECCY)
► 20.0 m CXP 16x Active Optical Cable (AOC) Pair (#ECCZ)

Cable lengths: Use the 2.0 m cables for intra-rack installations. Use the 10.0 m or 20.0 m cables for inter-rack installations.

Older cables #ECC6, #ECC8, and #ECC9 that are used in Power System E980 are not supported or oderable on Power System E1080. They are supported only as model conversions and announced as RPO.

A minimum of one PCIe Gen 4 Optical Cable Adapter for PCIe Gen 3 Expansion Drawer (#EJ24) is required to connect to the PCIe Gen 3 6-slot Fanout Module in the I/O expansion drawer. The top port of the Fanout Module must be cabled to the top port of the #EJ24 port. Also, the bottom two ports must be cabled together.

To perform the cabling correctly, complete the following steps:

1. Connect an AOC to connector T0 on the PCIe Gen 4 cable adapter (#EJ24) in your server.

2. Connect the other end of the optical cable to connector T1 on one of the PCIe Gen 3 6-slot Fanout Modules in your I/O expansion drawer.

3. Connect another AOC cable to connector T1 on the PCIe Gen 4 cable adapter (#EJ24).

4. Connect the other end of the cable to connector T2 on the same PCIe Gen 3 6-slot Fanout Module in I/O expansion drawer.

5. Repeat the steps 1 to 4 for the other PCIe Gen 3 6-slot Fanout Module in the I/O expansion drawer, if required.

Drawer connections: Each Fanout Module in a PCIe Gen 3 Expansion Drawer (#EMX0) can be connected to only a single PCIe Gen 4 Cable Adapter (#EJ24). However, the two Fanout Modules in a single I/O expansion drawer can be connected to different system nodes in the same server.

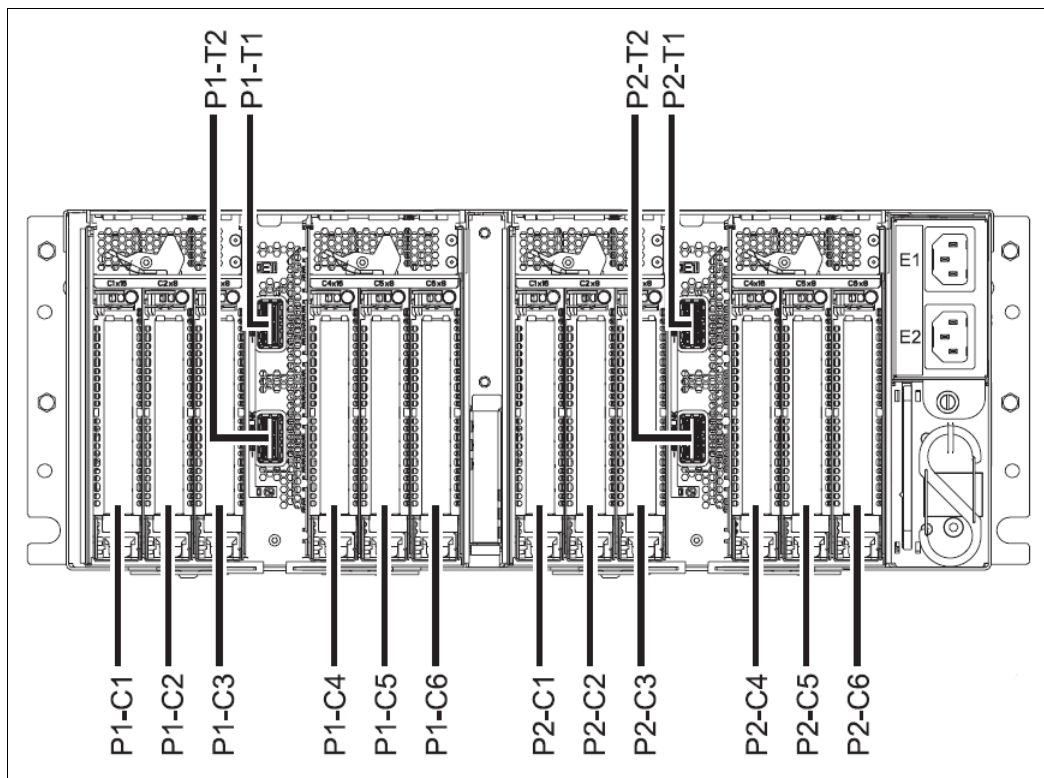Figure 2-26 shows the connector locations of the PCIe Gen3 I/O Expansion Drawer.
.

Figure 2-26   Connector locations for the PCIe Gen3 I/O Expansion Drawer

Table 2-25 lists the PCIe slots in the PCIe Gen3 I/O Expansion Drawer.

Table 2-25   PCIe slot locations and descriptions for the PCIe Gen3 I/O Expansion Drawer

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

Consider the following points:

► All slots support full-length, regular-height adapters or short (low-profile) adapters with a regular-height tailstock in single-wide, Gen3 BSCs.

► Slots C1 and C4 in each PCIe Gen 3 6-slot Fanout Module are x16 PCIe Gen 3 buses, and slots C2, C3, C5, and C6 are x8 PCIe buses.

► All slots support enhanced error handling (EEH).

► All PCIe slots are hot-swappable and support concurrent maintenance.

Typical Slot priorities are listed in Table 2-26.

Table 2-26   PCIe adapter slot priorities for the PCIe Gen3 I/O Expansion Drawer

Feature code	Description	Slot priorities
EMX0	Typical Slot priorities	1, 7, 4, 10, 2, 8, 3, 9, 5, 11, 6, 12

For more information about slot priorities for individual adapters, see IBM Documentation.

If the EMX0 PCIe Gen 3 Expansion Drawer is configured with two PCIe Gen 3 6-slot Fanout Modules, distribute the PCIe adapters across both I/O modules whenever possible

Slots 1, 7, 4, and 10 support SR-IOV capabilities. However, if the total amount of physical memory is less than 128 GB, Slot 4 (P1-C4) and Slot 10 (P2-C4) are not SR-IOV capable.

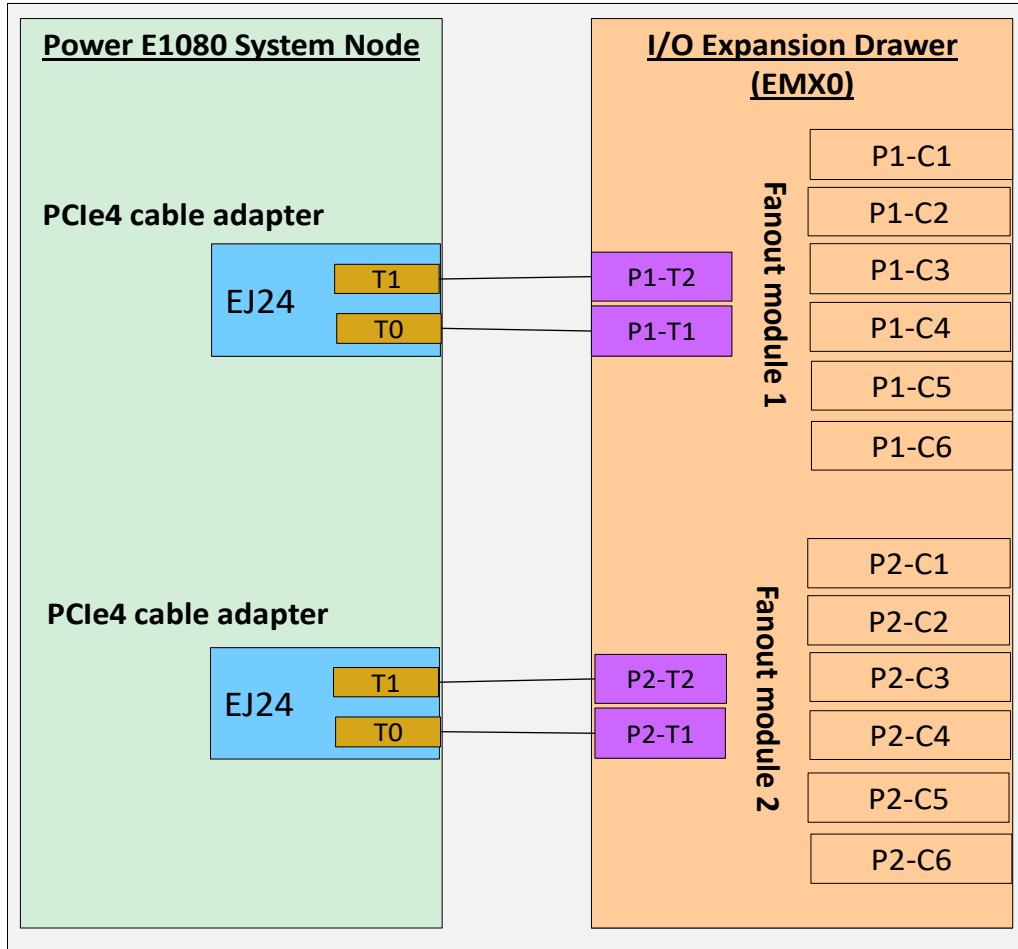Figure 2-27 shows the connection schematics between system node and PCIe Gen3 I/O Expansion Drawer.

Figure 2-27   Connection schematics between system node and Expansion Drawer

General rules for the PCIe Gen3 I/O Expansion Drawer configuration

The PCIe Gen 4 Cable Adapter (#EJ24) can be in any of the PCIe adapter slots in a Power E1080 server node. However, we advise that you first populate the PCIe adapter slots as per the slot priorities that are discussed in this section. If an I/O expansion drawer is present, the PCIe Gen 4 cable adapter (#EJ24) must be given the highest priority.

Each processor module drives two PCIe Gen5 slots, and all slots are equal regarding their bandwidth characteristics. If you first use the slots following the slot priorities, you ensure that one PCIe Gen5 slot per processor module is populated before you use the second PCIe Gen5 slot of the processor modules.

Table 2-27 lists the PCIe adapter slot priorities in the Power E1080 server. If the sequence is chosen as listed in the slot priority column, the adapters are assigned to the SCM in alignment with the internal enumeration order: SCM0, SCM1, SCM2, and SCM3.

Table 2-27   PCIe adapter slot priorities

Feature code	Description	Slot priorities
EJ24	PCIe Gen 4 Optical Cable Adapter for PCIe Gen 3 Expansion Drawer	1, 7, 3, 6, 2, 8, 4, 5

The following figures show several examples of supported configurations. (For simplification, we do not show every possible combination of the I/O expansion drawer to server attachments.)

Figure 2-28 shows an example of a Power 1080 that features a single system node and four PCIe Gen3 I/O Expansion Drawers. Each lane between the #EJ24 adapter and I/O drawer fanout module represent an CXP 16x AOC pair.
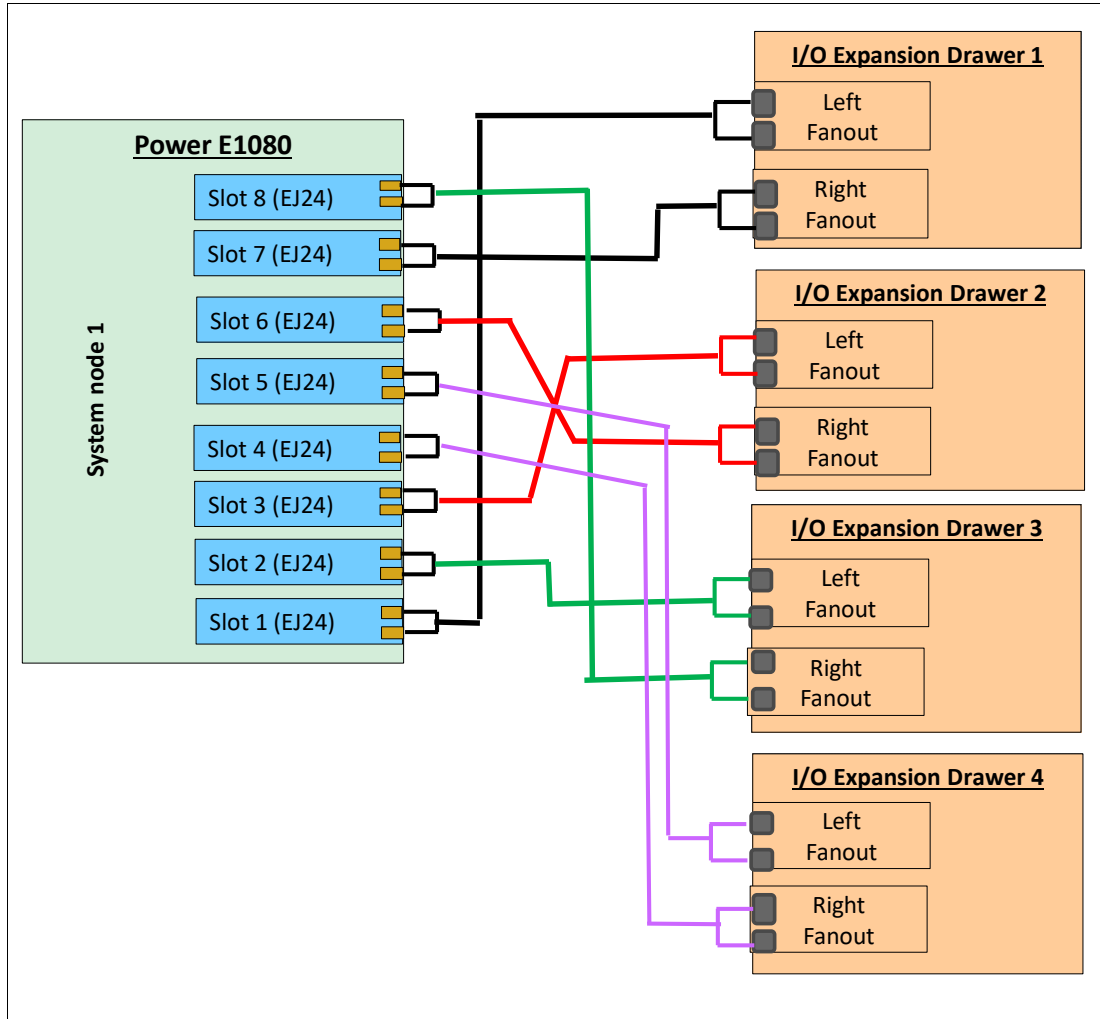

Figure 2-28   Single system node to four PCIe Gen3 I/O Expansion Drawer connections

Figure 2-29 shows an example of a Power E1080 having two system nodes and four PCIe Gen3 I/O Expansion Drawers. Each lane between the #EJ24 adapter and I/O drawer fanout module represent an CXP 16x AOC pair.
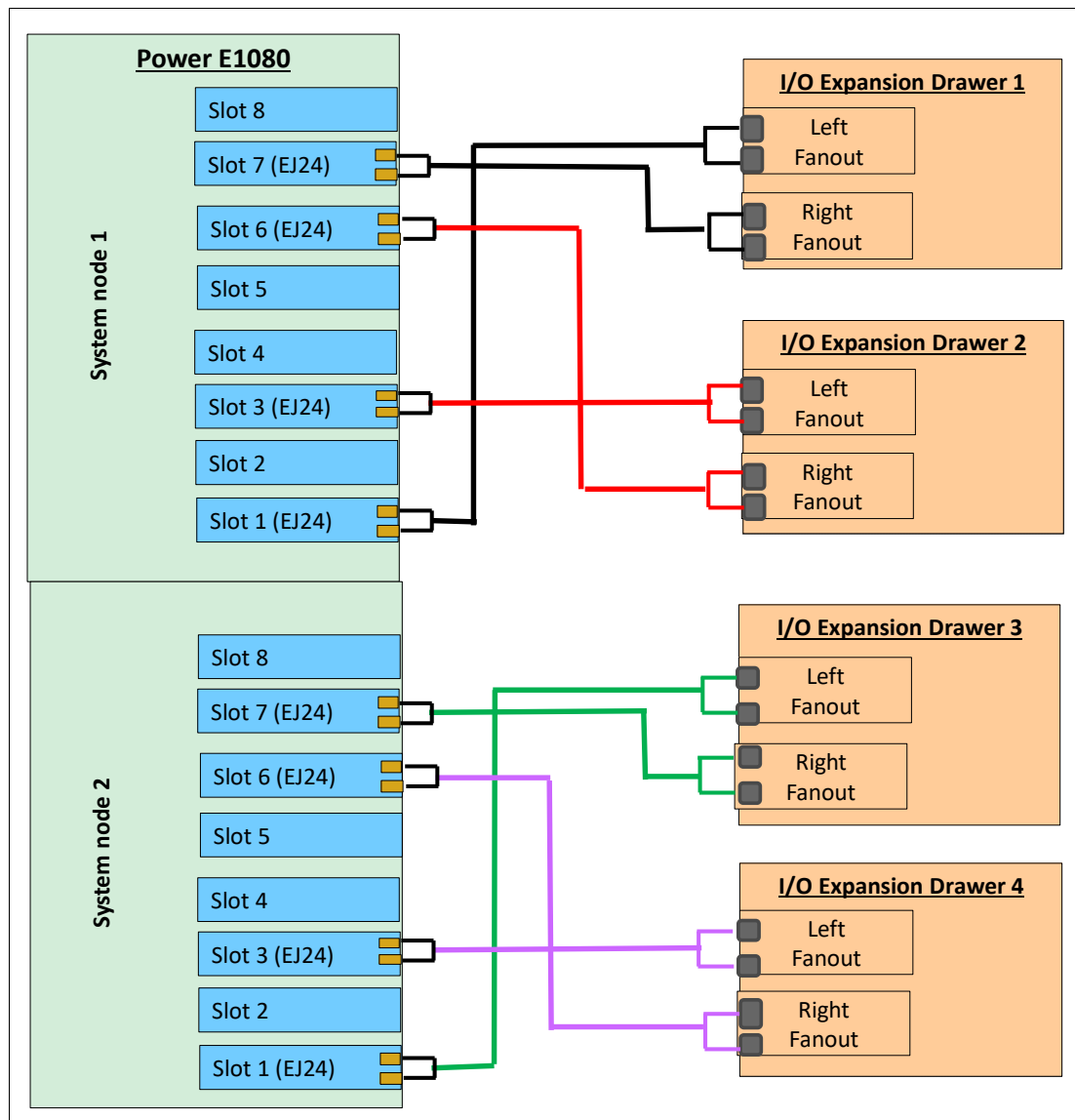

Figure 2-29   Dual system node to four PCIe Gen3 I/O Expansion Drawer connections

2.7.4 PCIe Gen3 I/O Expansion Drawer SPCN cabling

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

2.8 External disk subsystems

This section describes the following external disk subsystems that can be attached to the Power E1080 server:

► EXP24SX SAS Storage Enclosures
► IBM System Storage

2.8.1 IBM EXP24SX SAS Storage Enclosure

The EXP24SX SAS storage enclosure (#ESLS) is the only DASD drawer that is available for Power E1080 that provides extra disk drives. The ESLS supports SSDs and HDDs. The ESLS storage enclosures are connected to system units by using a serial-attached SCSI (SAS) port in the SAS adapters.

The EXP24SX is a storage expansion enclosure with 24 2.5-inch small form factor (SFF) SAS bays. It supports 10K and 15K SAS HDD having 512 and 4K format drives and enterprise and mainstream class SAS SSDs.

The enclosures can be split logically into one, two, or four independent groups:

► One set of 24 bays (mode 1) (D1-D24),
► Two sets of 12 bays (mode 2), (D1-D12) (D13-D24)or
► Four sets of 6 bays (mode 4).(D1-D6) (D7-D12) (D13-D18) (D19-D24)

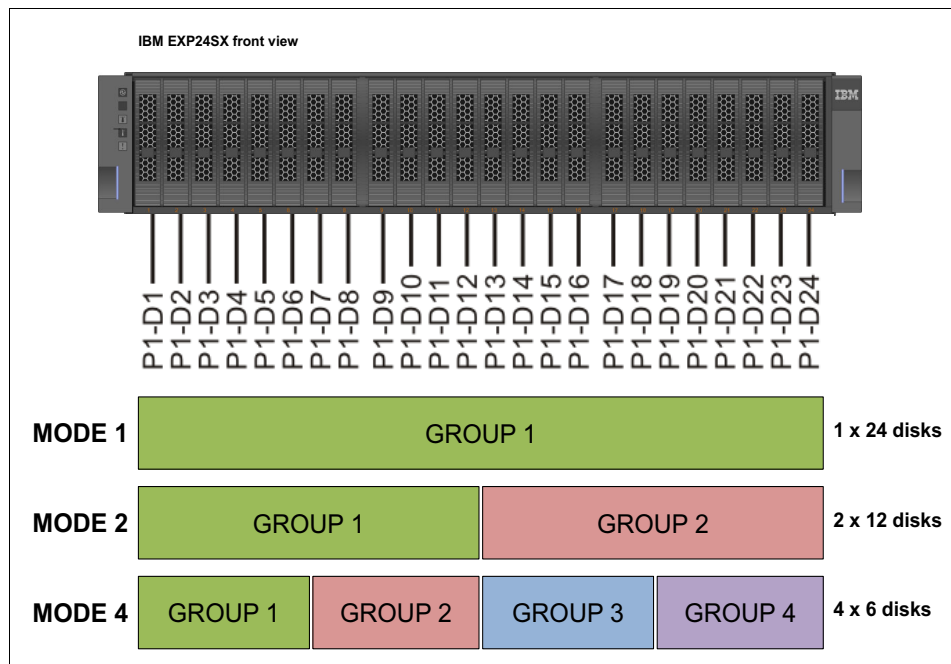The front view of the ESLS storage enclosure with drive locations is shown in Figure 2-30.

Figure 2-30   Front view of the ESLS storage enclosure with mode groups & drive locations

The SAS storage enclosures support the following operating systems:

► AIX
► IBM i
► Linux
► VIOS

The following PCIe Gen 3 SAS adapters support the EXP24SX drawers:

► PCIe Gen 3 RAID SAS Adapter Quad-port 6 Gb x8 (#EJ0J)
► PCIe Gen 3 LP RAID SAS Adapter Quad-Port 6 Gb x8 (#EJ0M)
► PCIe3 SAS Tape/DVD Adapter Quad-port 6 Gb x8 (#EJ10)
► PCIe3 LP SAS Tape/DVD Adapter Quad-port 6 Gb x8 (#EJ11)
► PCIe Gen 3 12 GB Cache RAID Plus SAS Adapter Quad-port 6 Gb x8 (#EJ14)

IBM i configurations require the drives to be protected (RAID or mirroring). Protecting the drives is highly advised, but not required for other operating systems. All Power servers operating system environments that use SAS adapters with write cache require the cache to be protected by using pairs of adapters.

The EXP24SX drawers feature the following high-reliability design points:

► SAS bays that support hot-swap.

► Redundant and hot-plug power and fan assemblies.

► Dual power cords.

► Redundant and hot-plug Enclosure Services Managers (ESMs).

► Redundant data paths to all drives.

► LED indicators on drives, bays, ESMs, and power supplies that support problem identification.

► Through the SAS adapters and controllers, drives that can be protected with RAID and mirroring and hot-spare capability.

Notes: Consider the following points about mixing SSDs and HDDs:

► SSDs and HDDs cannot be mixed when configured in mode 1.

► SSDs and HDDs can be mixed when configured in mode 2: One disk partition can be SSDs and the other disk partition can be HDDs, but you cannot mix within a disk partition.

► SSDs and HDDs can be mixed when configured in mode 4. Each disk partition can be SSDs *or* HDDs, but cannot mix within a disk partition.

For example, in a mode 2 drawer with two sets of 12 bays, one set can hold SSDs and one set can hold HDDs, but you cannot mix SSDs and HDDs in the same set of 12-bays.

The cables that are used to connect an #ESLS storage enclosure to a server are different from the cables that are used with the 5887 disk drive enclosure. Attachment between the SAS controller and the storage enclosure SAS ports is through the suitable SAS YO12 or X12 cables. The PCIe Gen3 SAS adapters support 6 Gb throughput. The EXP12SX drawer supports up to 12 Gb throughput if future SAS adapters support that capability.

The following cable options are available:

► 3.0M SAS X12 Cable (Two Adapter to Enclosure (#ECDJ)
► 4.5M SAS X12 AOC (Two Adapter to Enclosure (#ECDK)
► 10M SAS X12 AOC (Two Adapter to Enclosure (#ECDL)
► 1.5M SAS YO12 Cable (Adapter to Enclosure (#ECDT)
► 3.0M SAS YO12 Cable (Adapter to Enclosure (#ECDU)
► 4.5M SAS YO12 AOC (Adapter to Enclosure (#ECDV)
► 10M SAS YO12 AOC (Adapter to Enclosure (#ECDW)

Six SAS connectors are at the rear of the EXP24SX drawers to which SAS adapters or controllers are attached. They are labeled T1, T2, and T3; two T1s, two T2s, and two T3s connectors. Consider the following points:

► In mode 1, two or four of the six ports are used. Two T2 ports are used for a single SAS adapter, and two T2 and two T3 ports are used with a paired set of two adapters or a dual adapters configuration.

► In mode 2 or mode 4, four ports are used, two T2s and two T3 connectors, to access all the SAS bays.

► The T1 connectors are not used.

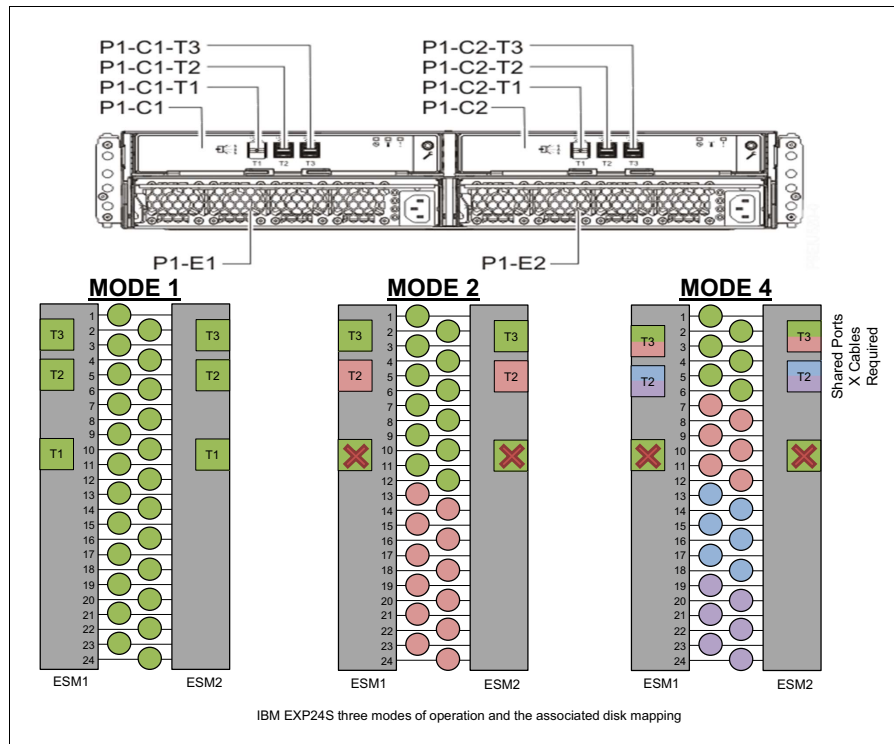Figure 2-31 shows the connector locations for the EXP24SX storage enclosure.

Figure 2-31   Rear view of the EXP24SX with location codes and different split modes

Mode setting is done by IBM Manufacturing. If you need to change the mode after installation, ask your IBM System Services Representative (IBM SSR) for support.

For more information about SAS cabling and cabling configurations, see "Connecting an #ESLS storage enclosure to your system" in IBM documentation.

2.9 System control and clock distribution

Power E1080 server requires UPIC, FSP & SMP cables for inter-drawer connections:

► UPIC cables allow the system nodes to power the System Control Unit (SCU)

► FSP cables are required to provide system control to the components

► SMP cables are necessary to extend the A bus interface that connect the Power10 processors together across the system nodes

Similar to the previous generation of Power System E980 server, two service processors are available for redundancy. They are hosted in the SCU and communicates with the system nodes by using the FSI/PSI bus connectors that are at the rear of the SCU and the system nodes.

All of the service processor communication between the control unit and the system nodes flows though the service processor cables. In comparison to previous generations, the Power E1080 associated SCU is no longer hosting the system clock. Each system node hosts its own redundant clocks.

The cables that are required for communications between the SCU and system nodes depend on the number of system nodes that are installed. When a system node is added, a new set of cables must be also added.

The cables that are necessary for each system node are grouped under a single feature code, which allows for an easier configuration. Each cable set includes a pair of FSP cables, and when applicable SMP cables and Universal Power Interconnect Cables (UPIC) cables.

Table 2-28 lists the available feature codes.

Table 2-28   Features for cable sets

Feature code	Description
EFCH	Cable set for system node drawer 1 (FSP + UPIC)
EFCE	Cable set for system node drawer 2 (FSP + SMP)
EFCF	Cable set for system node drawer 3 (FSP + SMP)
EFCG	Cable set for system node drawer 4 (FSP + SMP)

Initial orders of Power E1080 server includes one #EFCH, which is required to connect the system node with System Control Unit (SCU). This configuration does not require SMP cables, which are necessary only for configurations with two or more system nodes.

Cable sets feature codes are incremental and depend on the number of installed system nodes:

► One system node: #EFCH
► Two system nodes: #EFCH and #EFCE
► Three system nodes: #EFCH, #EFCE, and #EFCF
► Four system nodes: #EFCH, #EFCE, #EFCF, and #EFCG

The redundant Flexible Service Processor (FSP) provides proprietary interface communication, such as FRU Service Interface (FSI) and Processor Support Interface (PSI) to the system nodes.

PSI is used for FSP-host processor unit communication. PSI is a clock synchronous bidirectional interface for control communication. Each FSP in the SCU has four PSI interfaces and are connected such that whichever FSP becomes Primary can control the entire system.

FSI in the Power E1080 server is a serial point-to-point connection used for device communication in the overall System Control Structure design.

The FSI connection network is across FSP to FSP connections inside the SCU, FSP to system node through clocking and control logic. They connect FSPs in the SCU to system node elements and the Power10 to Power10 processor chips on the system node planar inside the system node. Similar to PSI network, whichever FSP becomes Primary can control the entire server.

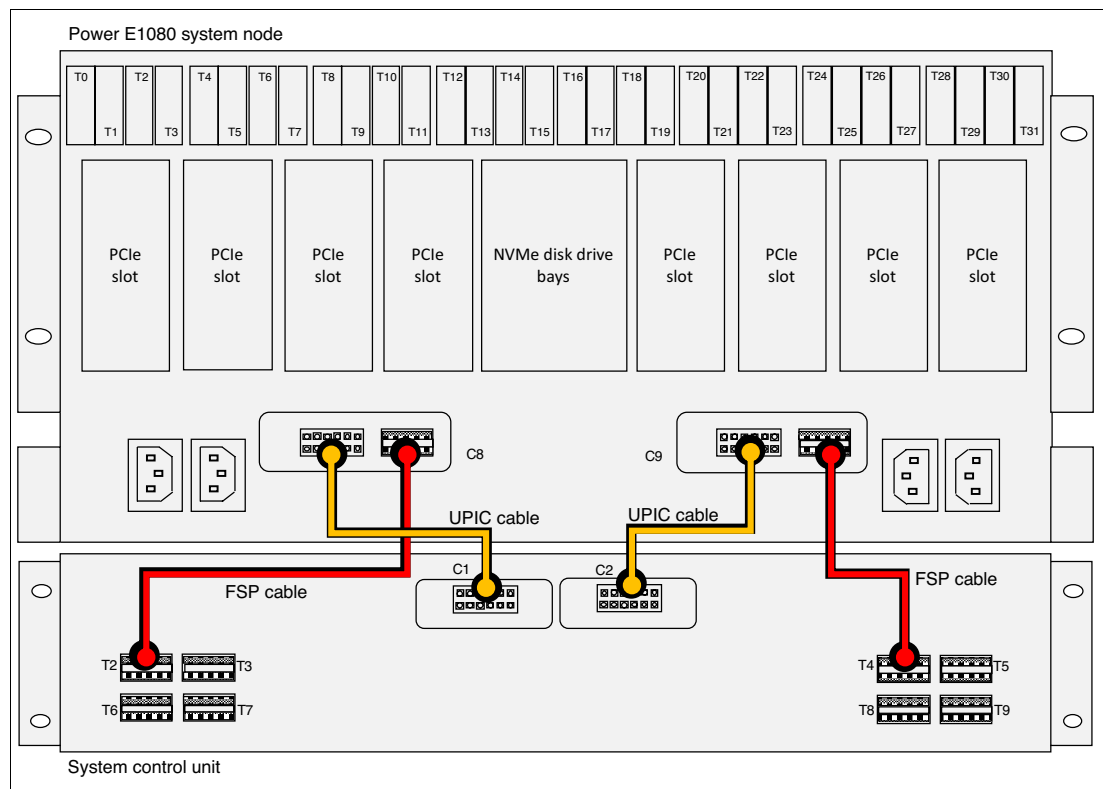Figure 2-32 shows UPIC and FSP cabling between a single system node and SCU.

Figure 2-32   UPIC & FSP connection between SCU and single System Node

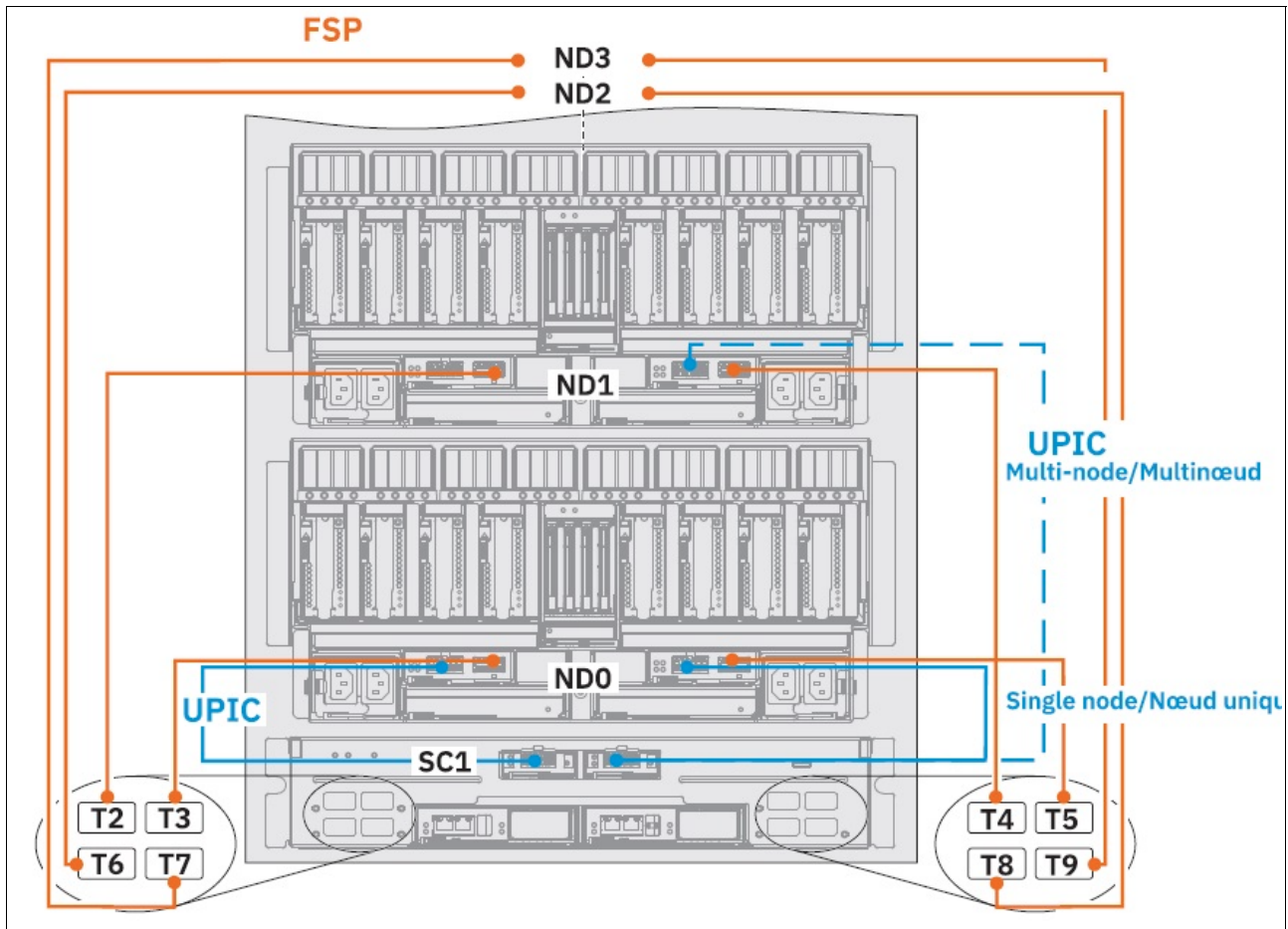Figure 2-33 shows UPIC & FSP cabling between two system nodes and SCU.

Figure 2-33   UPIC & FSP connection between SCU and two System Nodes

In a single system node configuration, two UPIC cables connect to the SCU and provide redundant power source from the system node. The SCU power source for two or more system nodes is supplied by the first and second system node. When Power E1080 supports more than two system nodes, power output ports on the third and forth system node are not used.

The system reference clock source is responsible for providing a synchronized clock signal to all functional units. Each system node of a Power E1080 server uses its own private set of two redundant system clock or control cards. If a failure occurs in any of the clock or control cards, the second card ensures continued operation of the system until a replacement is scheduled. Similar to Power E980, Power E1080 server does not require a global reference clock source in the system control.

Figure 2-34 shows UPIC and FSP cabling between three system nodes and SCU.
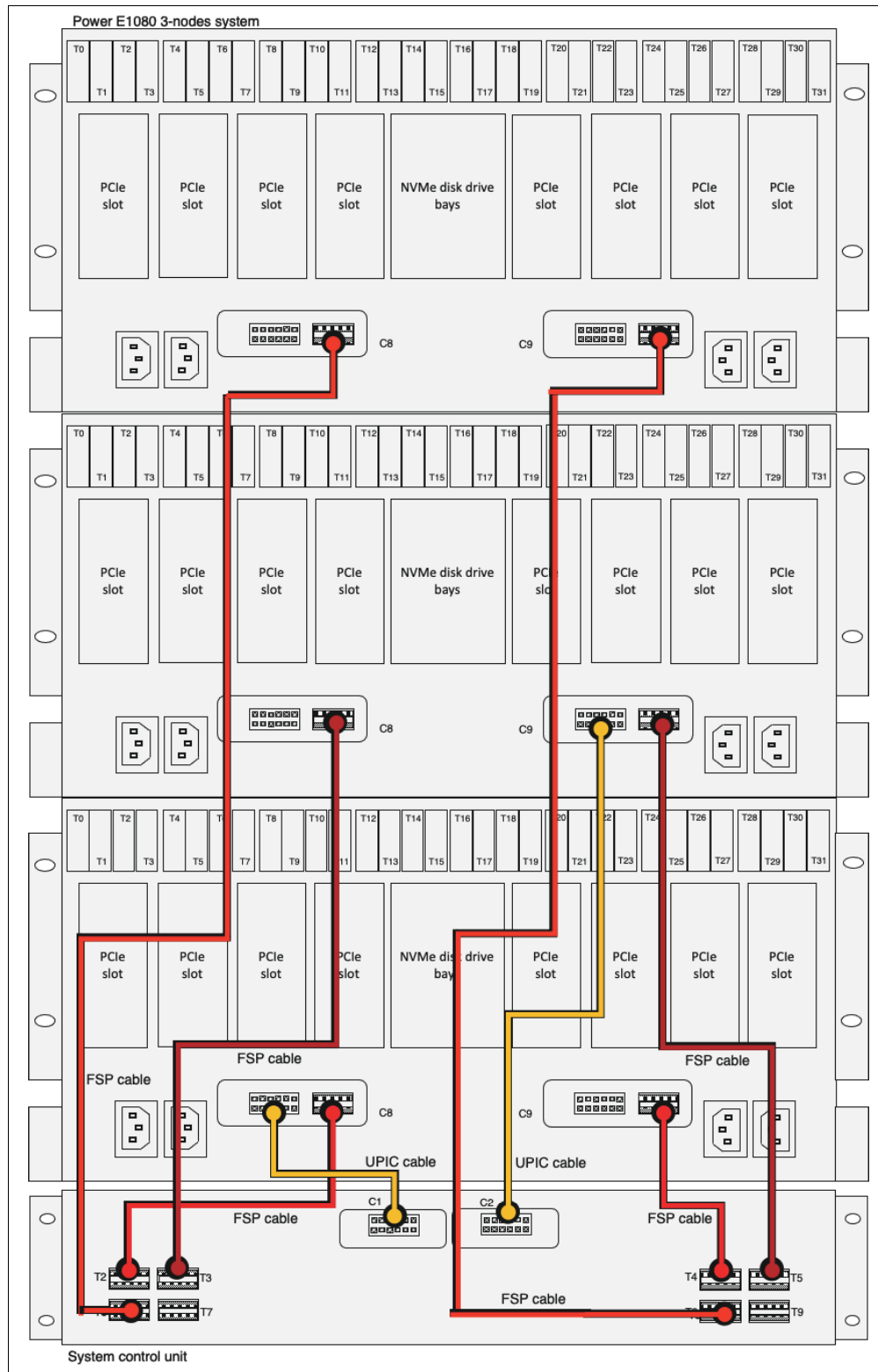

Figure 2-34   UPIC and FSP connection between SCU and three system nodes

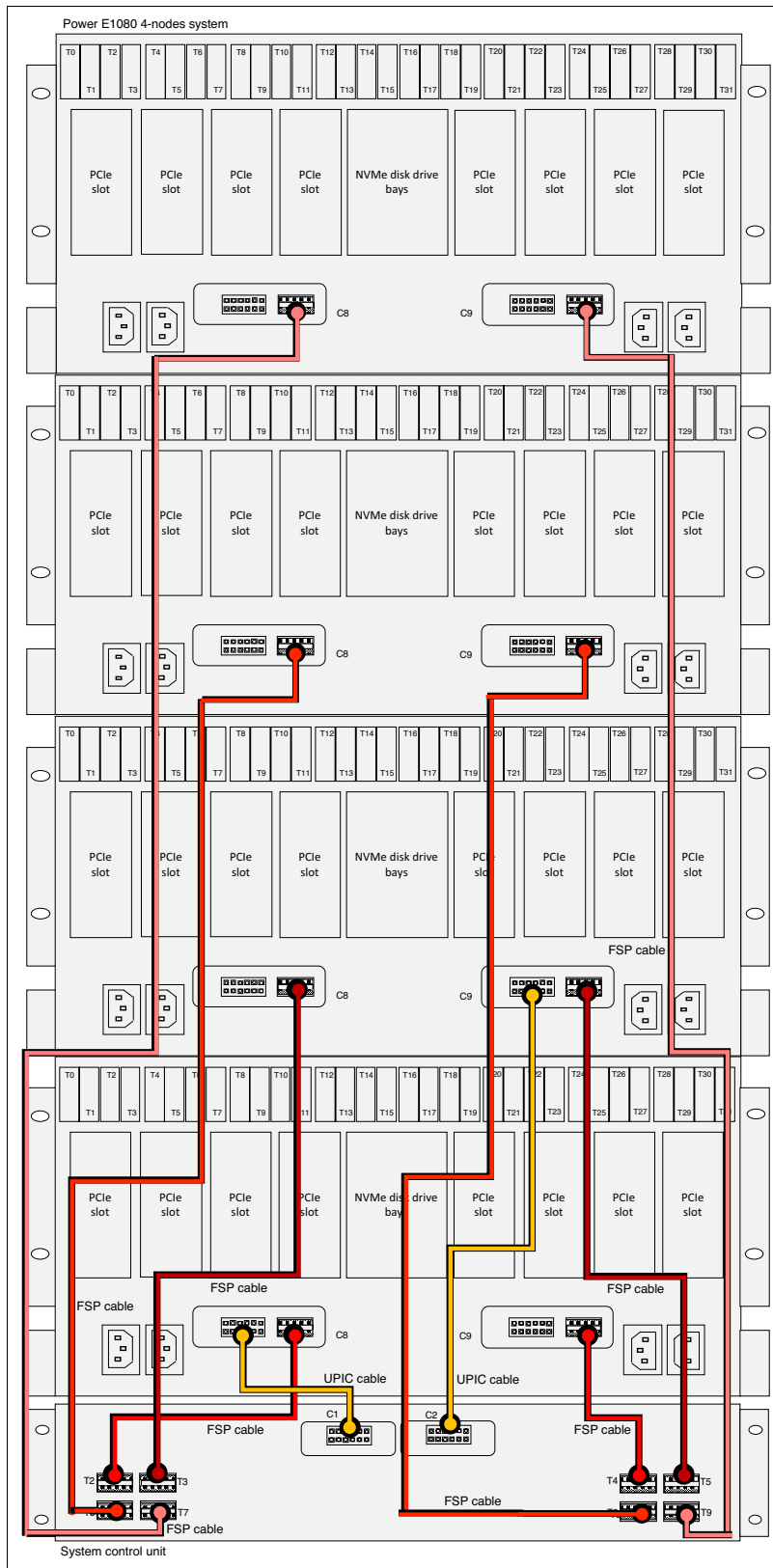Figure 2-35 shows UPIC and FSP cabling between four system nodes and SCU.

Figure 2-35   UPIC and FSP connection between SCU and four system nodes

2.10 Operating system support

The Power E1080 server supports the following operating systems:

► AIX
► IBM i
► Linux

In addition, the VIOS can be installed in special partitions that provide support to other partitions running AIX, IBM i, or Linux OSes for the use of various features, such as virtualized I/O devices, or PowerVM Live Partition Mobility (LPM).

For more information about the software that is available on Power Systems, see IBM Power Systems Software.

2.10.1 Power E1080 prerequisites

The minimum supported levels of IBM AIX, IBM i, and Linux at the time of this writing are described in the following sections. For more information about hardware features, see IBM Power Systems Prerequisites.

This tool helps also to plan a successful system upgrade by providing the prerequisite information for features currently in use or planned to be added to a system. It's possible to choose a machine type and model (9080-HEX for Power E1080) and find out all the prerequisites, operating system levels supported and other information.

2.10.2 AIX operating system

At announcement, the Power E1080 supports the following minimum level of AIX:

► AIX 7.1 TL 5 SP 5
► AIX 7.2 TL 4 SP 1

AIX 7.1 must run on an LPAR in POWER8 compatibility mode with VIOS-based virtual storage and networking. The I/O can be supplied only by the Virtual I/O Server and the system is LPM enabled, excluding Native VF (SR-IOV) directly in the customer LPAR.

AIX 7.2 can run with physical and virtual I/O and requires POWER9 compatibility mode.

NVMe over Fibre Channel fabrics (NVMe-OF) is now available on Power E1080 running AIX when the 2-port 32 Gb adapter (#EN1A or #EN1B) and the selected IBM FlashSystem high-end (FS9110) and mid-range (FS7200) model are used.

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX operating system. For more information about these packages, downloading, and obtaining the CD-ROM, see Fix Central.

The Service Update Management Assistant (SUMA), which can help you automate the task of checking and downloading operating system downloads, is part of the base operating system. For more information about the `suma` command, see IBM documentation.

AIX is available by using standalone (Standard Edition), Enterprise Edition, or through bundle options:

► AIX 7.1 Standard Edition
► AIX 7.1 Enterprise Edition V1.6
► AIX 7.2 Standard Edition
► AIX 7.2 Enterprise Edition
► AIX 7.2 Enterprise Cloud Edition

A major update for Enterprise Cloud Edition Bundles version includes PowerSC 2.0.

Customers that are purchasing new Power E1080 can get this stack with their system purchase. Customers with existing systems and active SWMAs by using either of the Bundles can get this Product by using ESS and start using it immediately.

Subscription licensing model

Per the Power E1080 announcement, a new subscription licensing model is available. This model provides customers with more options and flexibility in how they use the software.

A subscription license is a licensing model that provides access to an IBM program and IBM software maintenance for a specified subscription term (one or three years). The subscription term begins on the start date and ends on the expiration date, which is reflected in Entitled Systems Support (ESS).

Customers are licensed to run the product through the expiration date of the 1- or 3-year subscription term and then, can renew at the end of the subscription to continue using the product. This model provides flexible and predictable pricing over a specific term, with lower up-front entry cost.

Another benefit of this model is that the licenses are customer number entitled, which means they are not tied to a specific hardware serial number as with perpetual licenses. Therefore, the licenses can be moved between on-premises and cloud if needed, something that is becoming more of a requirement with hybrid workloads.

The new Product ID's for the subscription licenses are listed in Table 2-29.

Table 2-29   New subscription license PIDs (one or three years terms)

PID	Description
5765-2B1	AIX 7 standard edition
5765-2E1	AIX 7 enterprise edition 1.6
5765-2C1	Enterprise Cloud edition 1.6 with AIX 7
5765-6C1	Enterprise Cloud edition 1.6

The licenses are orderable through IBM configurator. The AIX perpetual and monthly term licenses for standard edition are still available.

AIX 7.3 Open Beta

In February 2021, the registration for the Open Beta of the new AIX 7.3 operating system was announced. The AIX 7.3 Open Beta allows you to download, install, and run an early version of the software on suitable IBM systems in a non-production, non-supported mode.

The Open Beta program provides an opportunity to gain early experience with AIX 7.3. Not all capabilities of AIX are available in the Open Beta.

Customers can request to join the Open Bets program by registering at this web page.

2.10.3 IBM i

IBM i is supported on the Power E1080 server by the following minimum required levels:

► IBM i 7.3 TR11 or later
► IBM i 7.4 TR5 or later

For compatibility information for hardware features and the corresponding IBM i Technology Levels, see IBM Prerequisites.

IBM i operating system transfer

IBM i customers can move to new Power E1080 servers, just like previous new system non-serial preserving upgrades/replacements.

IBM i terms and conditions require that IBM i operating system license entitlements remain with the machine for which they were originally purchased. Under qualifying conditions, IBM allows the transfer of IBM i processor and user entitlements from one machine to another. This capability helps facilitate machine replacement, server consolidation, and load rebalancing while protecting a customer's investment in IBM i software. When requirements are met, IBM i license transfer can be configured by using IBM configurator tools.

The following prerequisites must be met for transfers:

► The IBM i entitlements are owned by the user's enterprise.

► The donor machine and receiving machine are owned by the user's enterprise.

► The donor machine must have owned in the same user's enterprise as the receiving machine for a minimum of one year.

► Software Maintenance (SWMA) is on the donor machine. Each software entitlement to be transferred has SWMA coverage.

► An electronic Proof of Entitlement (ePoE) exists for the entitlements to be transferred.

► The donor machine entitlements is IBM i 5.4 or later.

► The receiving machine includes activated processors that are available to accommodate the transferred entitlements.

Each IBM i processor entitlement that is transferred to a target machine includes one year of new SWMA at no charge. Extra years of coverage or 24x7 support are available options for an extra charge.

2.10.4 Linux

The following types of Linux distributions are available to run on Power E1080:

► With native Power10 processor technology support
► Running in POWER9 compatibility mode

Distributions with native Power10 processor technology support also can run in POWER9 compatibility mode. This feature is important when doing LPM from a POWER9 processor-based server to Power E1080.

Red Hat

Red Hat Enterprise Linux (RHEL) version 8.4 and later can run in native Power10 processor mode. At the time of this writing, RHEL version 8.4 was available.

OpenShift and CoreOS also are supported.

SUSE

SUSE 15 SP3 is the first version with native Power10 processor technology support. Its regular support cycle is 18 months, plus long-term SP support.

Older distribution

The following selected older distribution also are supported on Power E1080:

► RHEL 8.2 is supported in POWER9 compatibility mode only
► SLES 12 SP5 is supported in POWER9 compatibility mode only

When a LPAR runs in POWER9 compatibility mode, it benefits from most of the features of the Power10 processor technology, including the full eight threads per core. However, program and kernel features that use new Power10 instructions or capabilities are not available.

The POWER9 compatibility mode is required when moving partitions back and forth between POWER9 processor based-systems and Power10 processor-based systems. After a partition is moved to a Power 10 processor-based systems, it can be upgraded to a distribution with native Power10 technology support and restarted in native Power10 mode.

> **Note:** No official support for POWER9 compatibility mode is available for older service pack of SLES 15, such as SLES 15 SP2.

Enhancements

One of the main features of the Power10 processor chip is the possibility to support up to 15 cores per SCM. Therefore, on a Power E1080 that is configured with up to 240 processor cores (each capable of running eight threads for up to 1920 possible threads), the Linux distribution with native Power10 technology support can use this capability.

LPARs running in POWER9 compatibility mode are restricted to 1536 threads per LPAR, which is the maximum for a POWER9 processor-based server. Table 2-30 lists the maximum logical CPUs, maximum memory, and maximum memory with LPM supported, according to the Linux distribution.

Table 2-30   Maximum threads versus processing mode

Linux distribution	Processor mode	Maximum logical CPUs	Maximum memory	Maximum memory with LPM
RHEL 8.2	POWER9	1536	64 TB	16 TB
RHEL 8.4	Power10	1920	64 TB	64 TB
SLES 12 SP5	POWER9	1536	64 TB	32 TB
SLES 15 SP3	Power10	1920	64 TB	64 TB

The Power10 specific toolchain is available in Advance Toolchain version 15.0, which allows customers and developers to use all new Power10 processor-based technology instructions when programming. Cross-module function call overhead was reduced because of a new PC-relative addressing mode.

One specific is a 10x-to-20x advantage over POWER9 processor-based technology on inferencing workloads because of new memory bandwidth and new instructions. One example is the new special purpose-built matrix math accelerator (MMA) that was tailored for the demands of machine learning and deep learning inference. It also includes many AI data types.

Network virtualization is an area with significant evolution and improvements, which benefit virtual and containerized environments. The following recent improvements were made for Linux networking features on Power E1080:

► SR-IOV allows virtualization of network cards at the controller level without the need to create virtual Shared Ethernet Adapters in the VIOS partition. It is enhanced with (vNIC) virtual Network Interface Controller, which allows data to be transferred directly from the partitions to or from the SR-IOV physical adapter without transiting through a VIOS partition.

► Hybrid Network Virtualization (HNV) allows Linux partitions to use the efficiency and performance benefits of SR-IOV logical ports and participate in mobility operations, such as active and inactive Live Partition Mobility (LPM) and Simplified Remote Restart (SRR). HNV is enabled by selecting a new Migratable option when an SR-IOV logical port is configured.

► NVMe over Fibre Channel fabrics (NVMe-OF) is now available on Power E1080 running Linux when the 2-port 32 Gb adapter (#EN1A or #EN1B) is and the selected IBM FlashSystem high-end (FS9110) and mid-range (FS7200) model are used.

Security

Security is a top priority for IBM and our distribution partners. Linux security on IBM Power is a vast topic that can be the subject of detailed separate material; however, improvements in the areas of hardening, integrity protection, performance, platform security, and certifications are introduced with this section.

Hardening and integrity protection deal with protecting the Linux kernel from unauthorized tampering while allowing upgrading and servicing to the kernel. These topics become even more important when running in a containerized environment with an immutable operating system, such as CoreOS in OpenShift.

Performance also is a security topic because specific hardening mitigation strategies (for example, against side-channel attacks), can have a significant performance effect. In addition, cryptography can use significant compute cycles.

The Power E1080 features transparent memory encryption at the level of the controller, which prevents an attacker from retrieving data from physical memory or storage-class devices that are attached to the processor bus.

2.10.5 Virtual I/O Server

The minimum required level of VIOS for the Power E1080 server is VIOS 3.1.3.10 or later.

IBM regularly updates the VIOS code. For more information, see IBM Fix Central.

2.10.6 Entitled System Support

The Entitled Systems Support website (ESS) is IBM's go-to place to view and manage Power and Storage software and hardware. In general, most products that are offered by IBM Systems that are purchased through our IBM Digital Sales representatives or Business Partners are accessed on this site when the IBM Configurator is used.

The site features the following three main sections:

► My entitled software: Activities related to Power and Storage software that offer to download licensed, free, and trial software media, place software update orders, and manage software keys.

► My entitled hardware: Activities related to Power and Storage hardware that offer to renew Update Access Keys, buy and use Elastic Capacity on Demand, assign or buy credits for new and existing pools in Enterprise Pools 2.0, download Storage Capacity on Demand codes, and manage Hybrid Capacity credits.

► My inventory: Activities related to Power and Storage inventory that offer to browse software license, software maintenance, and hardware inventory, manage inventory retrievals by way of Base Composer or generate several types of reports.

For more information, see this web page.

2.10.7 Update Access Keys

Since the introduction of the POWER8 processor-based server servers, IBM also introduced Update Access Key (UAK).

When system firmware updates are applied to the system, UAK and its expiration date are checked (see Figure 2-37 on page 122). System firmware updates include a release date.

When attempting to apply system firmware updates, if the release date for the firmware updates passed the expiration date for the update access key, the updates are not processed. As update access keys expire, they need to be replaced by using the Hardware Management Console (HMC) or the ASMI on the service processor.

By default, newly delivered systems include an UAK that often expires after three years. Thereafter, the UAK can be extended every six months, but only if a maintenance contract exists. The contract can be verified in the ESS website (for more information, see 2.10.6, "Entitled System Support" on page 119).

Determining when the current Update Access Key runs through the HMC, GUI or CLI as provided in the following examples. However, it also is possible to display the expiration date by using the suitable AIX or IBM i command.

UAK expiration date by using AIX 7.1
In the case of AIX 7.1, use the following command:

```
lscfg -vpl sysplanar0 | grep -p "System Firmware"
```

The output is similar to the output that is shown in Example 2-1 (the Microcode Entitlement Date represents the UAK expiration date).

Example 2-1   Output of the command to check UAK expiration date via AIX 7.1

```
$ lscfg -vpl sysplanar0 | grep -p "System Firmware"
      System Firmware:
...
        Microcode Image.............SV860_138 SV860_103 SV860_138
        Microcode Level.............FW860.42 FW860.30 FW860.42
        Microcode Build Date........20180101 20170628 20180101
        Microcode Entitlement Date..20190825
        Hardware Location Code......U8284.22A.XXXXXXX-Y1
      Physical Location: U8284.22A.XXXXXXX-Y1
```

UAK expiration date by using AIX 7.2

In the case of AIX 7.2, the output is slightly different from AIX 7.1. Use the following command:

```
lscfg -vpl sysplanar0 |grep -p "System Firmware"
```

The output is similar to the output that is shown in Example 2-2 (the Update Access Key Exp Date represents the UAK expiration date).

Example 2-2   Output of the command to check UAK expiration date via AIX 7.2

```
$ lscfg -vpl sysplanar0 |grep -p "System Firmware"
      System Firmware:
...
        Microcode Image.............SV860_138 SV860_103 SV860_138
        Microcode Level.............FW860.42 FW860.30 FW860.42
        Microcode Build Date........20180101 20170628 20180101
        Update Access Key Exp Date..20190825
        Hardware Location Code......U8284.22A.XXXXXXX-Y1
      Physical Location: U8284.22A.XXXXXXX-Y1
```

UAK expiration date by using IBM i

By using IBM i as operating system, it is possible to check the status of the UAK by using the Display Firmware Status window.

If the update access key expired, proceed to the ESS website to replace your update access key. Figure 2-36 on page 121 shows the output in the IBM i 7.1 and 7.2 releases. In the 7.3 release, the text changes to `Update Access Key Expiration Date`. The line that is highlighted in Figure 2-36 on page 121 is displayed whether the system is operating system managed or HMC managed.

```
                    Display Firmware Status

Service partition . . . . . . . . . . . . . . :    Yes
Firmware update policy  . . . . . . . . . . :    *OPSYS
Server IPL source . . . . . . . . . . . . . :    Permanent
Firmware product ID/release . . . . . . . . :    5733900   VOROMO
Firmware entitlement date . . . . . . . . . :    05/07/15
Server IPL required . . . . . . . . . . . . :    No


          --Server firmware--   -Operating system--
          Fix          PTF       Fix          PTF        Update
Copy      pack         ID        pack         ID         Status
*ACTIVE   ZL780_101    MH00000
          FW780.00
*TEMP     ZL780_111    LA00109   ZL780_111    LA00109    None
          FW780.00              FW780.00
*PERM     ZL780_101    MH00000   *NONE        *NONE      None
          FW780.00              *NONE


Press Enter to continue

F3=Exit    F5=Refresh    F12=Cancel    F14=Display firmware PTFs
```

Figure 2-36   Display Firmware Status window

2.11 Manageability

Manageability of the system is a key component for customers to manage, service, test, and monitor the system performance, security, and reliability.

2.11.1 Service user interface

To manage and service the system, the E1080 offers the following service interfaces:

► ASMI
► HMC GUI/CLI
► Operator panel
► Operating system

ASMI is a graphical interface that is part of the service processor firmware. The ASMI manages and communicates with the service processor. The ASMI is required to set up the service processor and to perform service tasks, such as reading service processor error logs, reading vital product data, and controlling the system power.

ASMI can be accessed by using the HMC GUI or CLI or directly attaching a laptop to the FSP port with an Ethernet cable that is crossed or through a switch (the same network subnet is required). Some of the functions in the ASMI are limited only to IBM service representatives, which requires celogin or dev user credentials during repair activities.

The control panel functions allow you to interface with the server. Control panel functions range in complexity from functions that display a status (such as IPL speed) to service functions that only service representatives can access.

For more information about a full list of all functions, see this web page.

The HMC GUI/CLI is the main manageable interface for customers. With HMC V10R1, the layout and the steps changed from previous versions.

2.11.2 System firmware maintenance

Starting with HMC V10R1 M1010, the main menu was re-arranged to classify the features that makes more sense in grouping.

Complete the following steps to update the firmware:

1. Click **Update Firmware** → **System FIrmware** → **Update**.
2. Accept the License Agreement (see Figure 2-37). You cannot proceed if you do not accept the agreement. Click **Next**.

Figure 2-37   Update System Firmware

3. In the next window, the Readiness Status column shows whether the system is ready. If the message is too long, hover around the box to see the entire message.

If two or more systems are selected and one is in a ready state and one is not in a ready state, the system that is *not* in a ready state is skipped and only the system with the ready state progresses (see Figure 2-38).

Click **Next**.
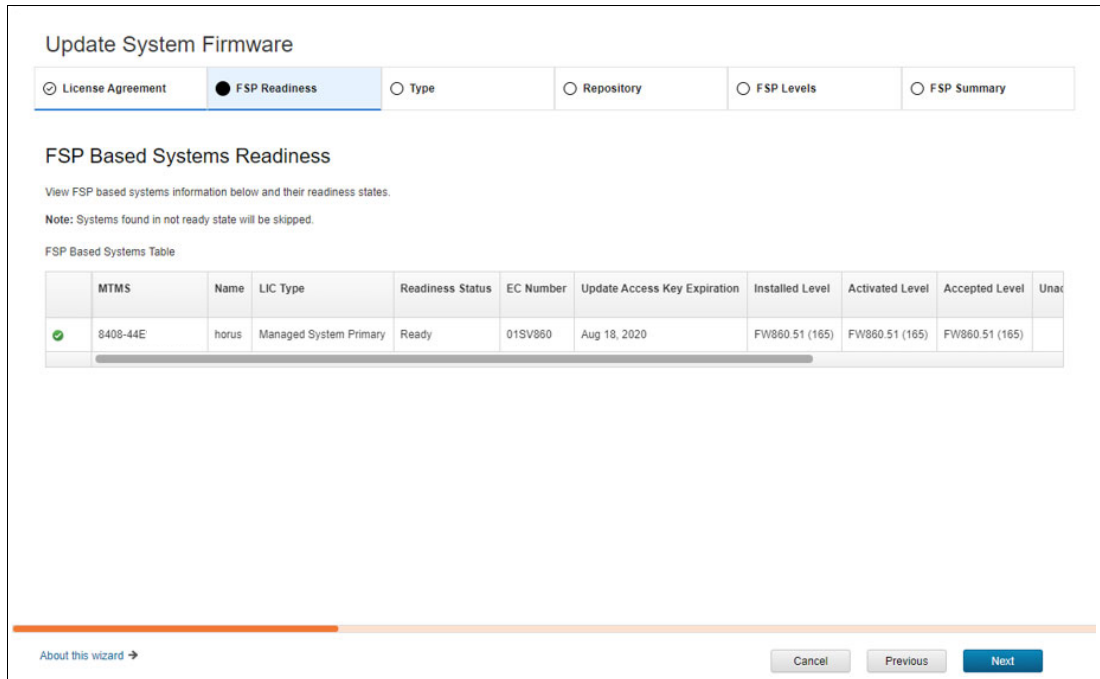
Figure 2-38   FSP Based Systems Readiness window

4. The System Firmware Type window opens. Two options are available: update or upgrade (see Figure 2-39). Click **Update** and then, click **Next**.
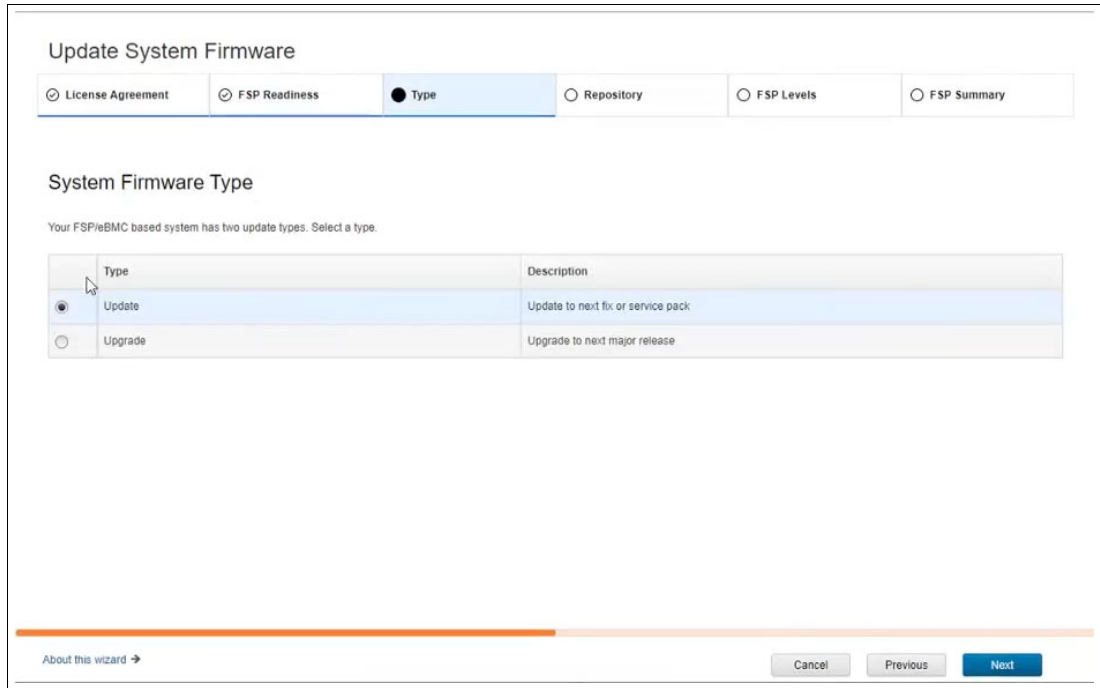
Figure 2-39   System Firmware Type window

5. The System Firmware Repository window opens (see Figure 2-40). Select from one of the following locations for the repository:
   – IBM Service Web Site
   – Removable Media
   – FTP Site
   – Secure FTP Site
   – Import Staging Location
   – Mount Point

   Click **Next**.
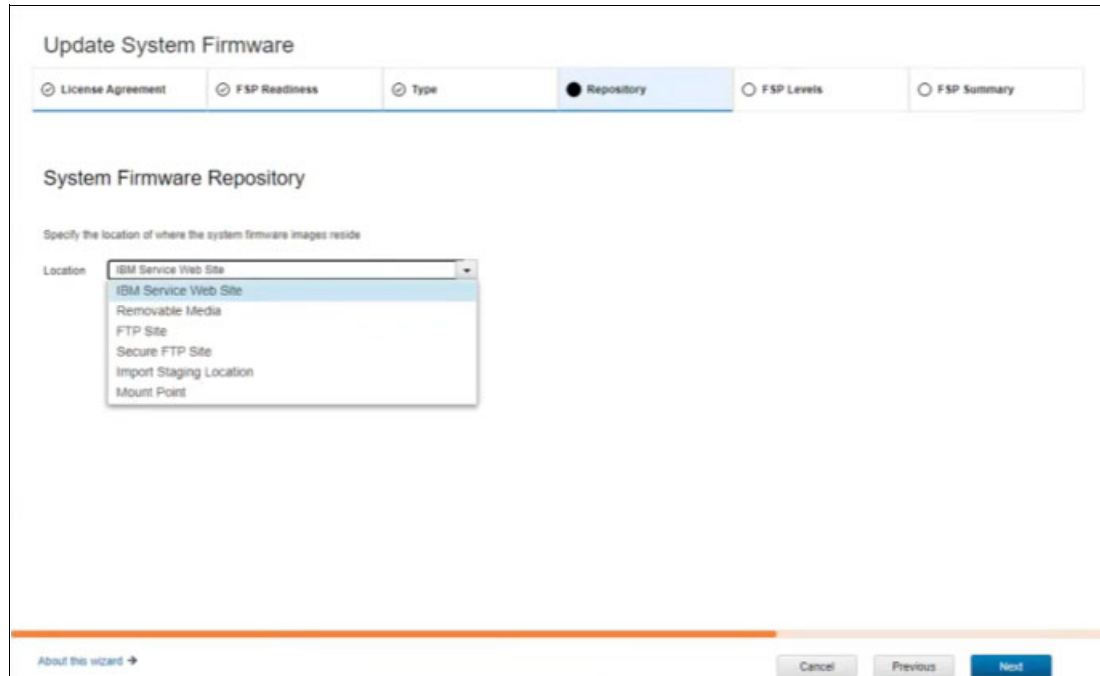
Figure 2-40   System Firmware Repository window

6. The Available Levels for Update window opens. In the target level column, the nature of the update (disruptive or non-disruptive) is shown. If you want to see what is included in the update or upgrade and which LPARs are affected, select the system.

From the drop-down menu, select **View Cover Data** or **View Impacted LPARs** (see Figure 2-41). then, click **Next**.
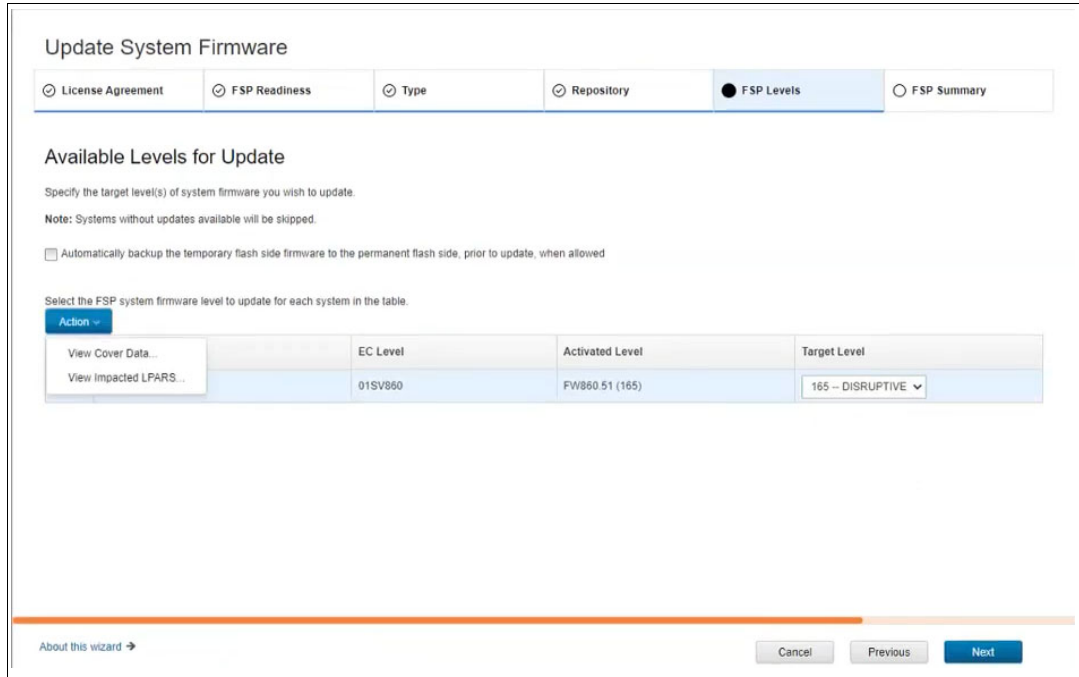
Figure 2-41 Available Levels of Update window

7. The Update Summary window opens. In this window, the progress of the update and whether it was successfully completed is shown. If a disruptive upgrade is to occurs, select the **I acknowledge that system disruption will occur** option. If this option is not selected when a disruptive upgrade occurs, the wizard cannot progress (see Figure 2-42).

Click **Next**.

Figure 2-42   Update Summary window

In the system Firmware Update Progress window, window (see Figure 2-43), the estimated time for and the progress of the installation are displayed.
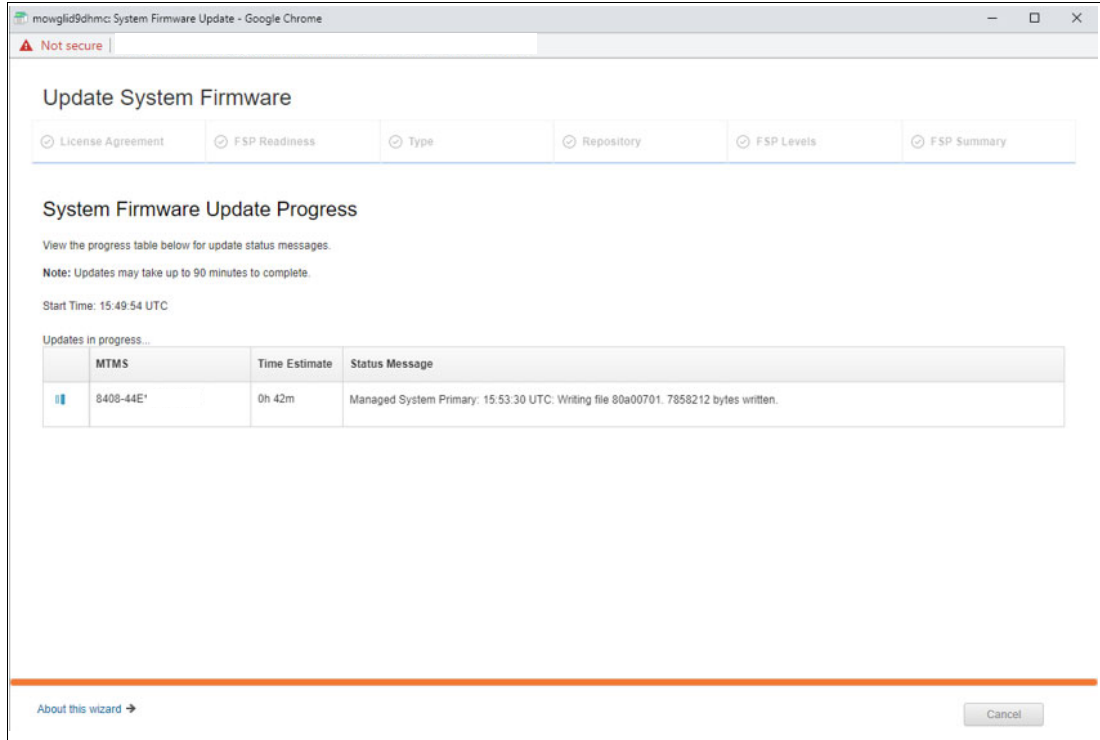
Figure 2-43   System Firmware Update Progress window

2.11.3 I/O firmware update

The I/O firmware update was integrated in the old GUI. However, to make the GUI more clear and simple, now it features a new layout that is similar to the system firmware update GUI.

For example, the first and second steps of I/O firmware update includes the License Agreement and Repository windows, which are similar to the system firmware update windows.

The next step of the wizard shows the system name and partition name that owns the I/O device, the current level, the available level of the related device microcode, and suggested actions and device description (see Figure 2-44).
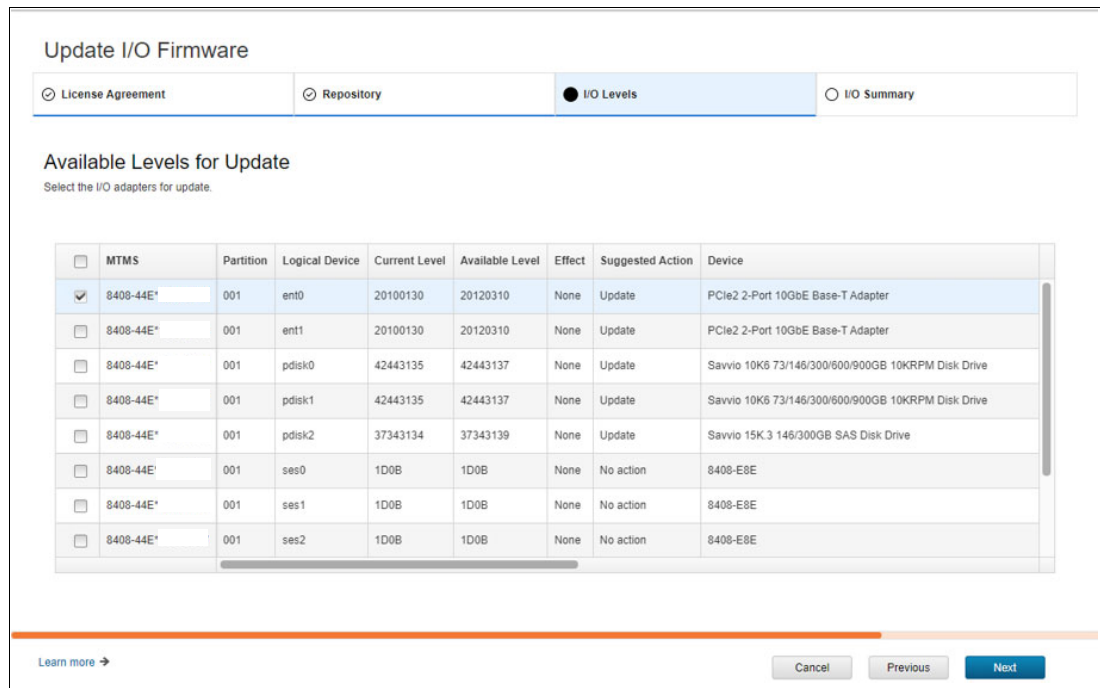
Figure 2-44   I/O firmware update interface

2.12 Serviceability

Instances exist in which maintenance procedures are performed on systems and subsystems, including replacing failed components, engineering change implementations, MES hardware upgrades, and migrations. Any of these procedures need the system operations to completely halt or run with a degradation in service.

The productive hours that are lost in performing the service negatively affects the business; therefore, the service strategy includes plans for not only how efficiently the system can be brought back to resume its operations normally, but also to reduce the hours that are needed to service the system.

The success of achieving these goals determines the availability of the system; that is, efficient serviceability means high availability. Power E1080 server carries forward the rich legacy of enterprise class Power systems to deliver the best enterprise RAS features compared to any OEM vendor.

Power E1080 server inherits the fundamental guiding principles and elements of serviceability from the Power E980 enterprise server. The following attributes of its service environment and the interfaces that help it achieve the objectives of serviceability are key:

► Errors detection
► Diagnose
► Reporting
► Notification
► Ease of location and service

These key attributes are discussed next.

2.12.1 Error detection

System ability to detect errors immediately can greatly enhance its capabilities to take suitable actions in time. For many soft errors, it can recover by self-healing properties. For other non-recoverable errors, it can notify through service interfaces for further actions. The overview of these service interfaces as applicable for Power E1080 server are discussed next.

Service interface

Service engineers are assisted by multiple system service interfaces that communicate with the service support applications in a server that is using the operator console, the GUI that is on the management console or service processor menu, or an operating system terminal. The service interface helps the support team to efficiently manage system resources and service information.

Applications that are available through the service interface are configured and placed to give service engineers access to important service functions. Depending on the system state, hypervisor, and the operating environment, one or more service interfaces can be useful in accessing logs and service information and communicating with the system. The following primary service interfaces are available:

- ► Light path diagnostics (LPD)
- ► Operator panel
- ► Service processor menu
- ► ASMI
- ► Operating system service menu
- ► Service focal point on the HMC or vHMC

The system can identify components for replacement by using Field Replaceable Unit (FRU) specific LEDs. The service engineer can use the identify function to set the Field replaceable unit (FRU) level LED to blink, which lights the blue enclosure locate and system locate LEDs. The enclosure LEDs turn on solid and guide the service engineer to follow the light path from the system to the enclosure and down to the specific FRU in error.

Similar to LPD notifies, other interfaces in the previous bulleted list provide tools to capture logs/dumps and other essential information to identify and detect errors.

For more information about service interfaces and the available service functions, see this web page.

Error check, first failure data capture, and fault isolation registers

Power processor-based systems feature specialized hardware detection circuits that are used to detect erroneous hardware operations. Error-checking hardware ranges from parity error detection that is coupled with Processor Instruction Retry and bus try again, to ECC correction on caches and system buses.

Within the processor or memory subsystem error checker, error-checker signals are captured and stored in hardware FIRs. The associated logic circuitry is used to limit the domain of an error to the first checker that encounters the error. In this way, runtime error diagnostic tests can be deterministic so that for every check station, the unique error domain for that checker is defined and mapped to field-replaceable units (FRUs) that can be repaired when necessary.

First-failure data capture (FFDC) is a technique that helps ensure that the root cause of the fault is captured without the need to re-create the problem or run any sort of extending tracing or diagnostics program when a fault is detected in a system. For most faults, a good FFDC design means that the root cause also can be detected automatically without service engineers' intervention.

FFDC information, error data analysis, and fault isolation are necessary to implement the advanced serviceability techniques that enable efficient service of the systems and to help determine the failing items.

In the rare absence of FFDC and Error Data Analysis, diagnostics are required to re-create the failure and determine the failing items.

2.12.2 Diagnostics

Diagnostics refers to identification of errors, symptoms, and determination of the potential causes of the errors identified.

Power E1080 server is supplemented with several advanced troubleshooting and diagnostic routines that are available through multiple service interfaces, as discussed in 2.12.1, "Error detection" on page 130. For more information about the available aids, see the following IBM Documentation web pages:

► Analyzing problems
► Isolation procedures

Reference codes

These codes represent the system IPL status progress codes, OS IPL progress codes, dump progress codes, service request numbers (SRN) and so on, which serve as diagnostic aid to help determine the source of various hardware errors. Diagnostic applications report problems by means of SRNs.

The support team and service engineers use this information with reference code-specific information to analyze and determine the source of errors or find more information about other isolation procedures.

Automatic diagnostics

The processor and memory FFDC is designed to perform without the need to recreate the problems or user intervention. Solid and intermittent errors are detected early and isolated at the time of failure. Runtime and boot-time diagnostics fall into this category.

Stand-alone diagnostics

These routines provide methods to test system resources by using diagnostics that are packaged on CDROM. They can be accessed by starting the system in service mode.

Service processor diagnostic

A service processor is self-sufficient to monitor unrecoverable errors in the system processor without the need of resources from the system processor. It also can monitor HMC connections and system thermal and operating environments, along with remote power control, reset, and maintenance functions.

2.12.3 Reporting

If a system hardware or environmentally induced failure occurs, the system runtime error diagnostics analyze the hardware error signature to determine the cause of failure.

The analysis is stored in the system NVRAM. The identified errors are reported to the operating system and recorded in the system logs of the operating system.

For an HMC-managed system in PowerVM environment, an ELA routine analyzes the error, forwards the event to the Service Focal Point (SFP) application that is running on the HMC, and notifies the system administrator about isolation of likely cause of the system problem. The service processor event log also records unrecoverable checkstop conditions and forwards them to the SFP application.

The system can call home from IBM i and AIX operating systems to report platform recoverable errors and errors that are associated with PCIe adapters and devices. In an HMC-managed system environment, a Call Home service request is started from the HMC and the failure report that carries parts information and part location is sent to IBM service organization.

Along with such information, customer contact information and system-specific information, machine type, model, and serial number, and error logs also are sent to IBM service organization electronically.

2.12.4 Notification

In this section, we describe the types of notifications that are available.

Call Home

HMC can notify service events that need IBM service organization attention, which uses the Call Home feature. Call Home helps to transmit error logs, server status, or other service-related information to the IBM service organization electronically.

This feature is optional, but customers can gain significant advantages implementing the feature, such as faster problem determination and resolution and, in most cases, without customer notice or direct involvement. This feature also allows the IBM Service organization to register auto service tickets and send customer replaceable units (CRU) directly to the customer or dispatching a service engineer to the customer location.

IBM Electronic Service Agent

The Electronic Service Agent (ESA) provides automatic problem reporting functions and can potentially predict and prevent hardware errors by early detection of potential problems. It is supported by IBM i and AIX and is not separately charged for because it is part of the base operating system.

ESA monitors and collects system inventory and service information, which is accessible from a secured web portal.

For more information about ESA planning, implementation, and configuration, see the following web pages:

► For AIX
► For IBM i

Customers who are skeptical about the information that is shared with IBM can be assured that security protocols are in place. Also, by registering with the IBM Electronics Services portal, the service information is accessible to customers.

2.12.5 Ease of location and service

For maintenance procedures that need service engineers to interact physically with the system, such as hardware MES execution, concurrent maintenance, or replacement of non-concurrent maintanable components, the time that is required for overall service can be greatly improved by using specific system aided tools that can help with replacement procedures.

Some the aids that are available with Power1080 system are discussed next.

Service labels

Placed at various locations on the system hardware, these labels provide the following ready-to-use, graphics-based information to perform physical maintenance on the system:

► Location diagrams

Typically, these diagrams consist of placement view of various hardware components (FRUs and CRUs) that are inside the server chassis. This information can help the service engineer to find the location of the FRUs that is being serviced or removed.

► Remove/ or replace procedures

A pictorial aid that shows the way to remove hardware components from the server chassis. For example, a picture can show thumbs that are pressing the latches, which can help release the component from its place.

► Arrows

Numbered arrows are used to indicate the order of actions and the serviceability direction of components. Some serviceable parts, such as latches, levers, and touch points, must be pulled or pushed in a specific direction and in a specific order for the mechanical mechanisms to engage or disengage.

► Physical address diagrams

These diagrams can help map the logical address of failed components to the physical address so that only intended components can be identified and serviced.

QR labels

QR labels are placed on the system to provide access to key service functions through a mobile device. When the QR label is scanned, the mobile device goes to server specific landing web pages that contain many of the service functions of interest while physically located beside the server rack in the data center. These functions include installation and repair instructions, service diagrams, and reference code look up.

Packaging for service

Power systems deliver great experiences with its ease of service operations design and unique physical packaging, which includes the following characteristics:

► Color coding (touch points): Blue-colored touch points delineate touch points on service components where the component can be safely handled for service actions, such as removal or installation.

► Tool-less design: Selected IBM systems support tool-less or simple tool designs. These designs require no tools or simple tools, such as flathead screw drivers, to service the hardware components.

► Positive retention: Positive retention mechanisms help to assure proper connections between hardware components, such as cables to connectors, and between two cards that attach to each other. Without positive retention, hardware components run the risk of becoming loose during shipping or installation, which prevents a good electrical connection.

   Positive retention mechanisms, such as latches, levers, thumb-screws, pop Nylatches (U-clips), and cables are included to help prevent loose connections and aid in installing (seating) parts correctly. These positive retention items do not require tools.

Concurrent maintenance

The Power E1080 includes many physical components that allow concurrent maintenance, which frees the service engineer to bring the system down for maintenance. The following concurrent maintanable components of E1080 are available:

► EXP24S SAS storage enclosure drawer

► Drives in the EXP24S storage enclosure drawer

► NVMe U.2 drives

► PCIe extender cards, optical PCIe link I/O expansion card

► PCIe I/O adapters

► PCIe I/O drawers

► PCIe to USB conversion card

► SMP cables

► System node AC power supplies: Two functional power supplies must remain installed at all times while the system is operating

► System node fans

► System control unit fans

► System control unit operations panel

► Time of Day clock battery

► UPIC interface card in SCU

► UPIC power cables from system node to system control unit

Enterprise solutions

In this chapter, we describe the major pillars that can help enterprises achieve their business goals and the reasons on why Power E1080 provides a significant contribution to that end.

This chapter includes the following topics:

3.1 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that delivers industry-leading virtualization for enterprises. It is the umbrella branding term for Power processor-based server virtualization; that is, IBM POWER Hypervisor, logical partitioning, IBM Micro-Partitioning®, VIOS, Live Partition Mobility (LPM), and more. PowerVM is a combination of hardware enablement and software.

> **Note**: PowerVM Enterprise Edition License Entitlement is included with each Power E1080 server. PowerVM Enterprise Edition is available as a hardware feature (#5228), supports up to 20 partitions per core, VIOS, multiple shared processor pools (MSPPs) and also offers LPM.

3.1.1 IBM POWER Hypervisor

Power processor-based servers are combined with PowerVM technology and offer the following key capabilities that can help to consolidate and simplify IT environments:

► Improve server usage and share I/O resources to reduce the total cost of ownership (TCO) and better use IT assets.

► Improve business responsiveness and operational speed by dynamically reallocating resources to applications as needed to better match changing business needs or handle unexpected changes in demand.

► Simplify IT infrastructure management by making workloads independent of hardware resources so that business-driven policies can be used to deliver resources that are based on time, cost, and service-level requirements.

Combined with features in the Power E1080, the IBM POWER Hypervisor delivers functions that enable other system technologies, including logical partitioning (LPAR) technology, virtualized processors, IEEE virtual local area network (VLAN)-compatible virtual switch, virtual SCSI adapters, virtual Fibre Channel adapters, and virtual consoles.

The POWER Hypervisor is a basic component of the system's firmware and offers the following functions:

► Provides an abstraction between the physical hardware resources and the LPARs that use them.

► Enforces partition integrity by providing a security layer between LPARs.

► Controls the dispatch of virtual processors to physical processors.

► Saves and restores all processor state information during a logical processor context switch.

► Controls hardware I/O interrupt management facilities for LPARs.

► Provides VLAN channels between LPARs that help reduce the need for physical Ethernet adapters for inter-partition communication.

► Monitors the Flexible Service Processor (FSP) and performs a reset or reload if it detects the loss of one of the FSP, notifying the operating system if the problem is not corrected.

The POWER Hypervisor is always active, regardless of the system configuration or whether it is connected to the managed console. It requires memory to support the resource assignment of the LPARs on the server. The amount of memory that is required by the POWER Hypervisor firmware varies according to several factors:

► Memory usage for hardware page tables (HPTs)
► Memory usage to support I/O devices
► Memory usage for virtualization

Memory usage for hardware page tables

Each partition on the system includes its own hardware page table (HPT) that contributes to hypervisor memory usage. The HPT is used by the operating system to translate from effective addresses to physical real addresses in the hardware. This translation from effective to real addresses allows multiple operating systems to run simultaneously in their own logical address space. Whenever a virtual processor for a partition is dispatched on a physical processor, the hypervisor indicates to the hardware the location of the partition HPT that can be used when translating addresses.

The amount of memory for the HPT is based on the maximum memory size of the partition and the HPT ratio. The default HPT ratio is 1/128th (for AIX, Virtual I/O Server [VIOS], and Linux partitions) of the maximum memory size of the partition. AIX, VIOS, and Linux use larger page sizes (16 and 64 KB) instead of using 4 KB pages. The use of larger page sizes reduces the overall number of pages that must be tracked; therefore, the overall size of the HPT can be reduced. For example, the HPT is 2 GB for an AIX partition with a maximum memory size of 256 GB.

When defining a partition, the maximum memory size that is specified is based on the amount of memory that can be dynamically added to the dynamic partition (DLPAR) without changing the configuration and restarting the partition.

In addition to setting the maximum memory size, the HPT ratio can be configured. The `hpt_ratio` parameter for the `chsyscfg` Hardware Management Console (HMC) command can be issued to define the HPT ratio that is used for a partition profile. The valid values are 1:32, 1:64, 1:128, 1:256, or 1:512.

Specifying a smaller absolute ratio (1/512 is the smallest value) decreases the overall memory that is assigned to the HPT. Testing is required when changing the HPT ratio because a smaller HPT might incur more CPU consumption because the operating system might need to reload the entries in the HPT more frequently. Most customers choose to use the IBM provided default values for the HPT ratios.

Memory usage for I/O devices

In support of I/O operations, the hypervisor maintains structures that are called the *translation control entities* (TCEs), which provide an information path between I/O devices and partitions. The TCEs provide the address of the I/O buffer, indications of read versus write requests, and other I/O-related attributes. Many TCEs are used per I/O device, so multiple requests can be active simultaneously to the same physical device. To provide better affinity, the TCEs are spread across multiple processor chips or drawers to improve performance while accessing the TCEs.

For physical I/O devices, the base amount of space for the TCEs is defined by the hypervisor that is based on the number of I/O devices that are supported. A system that supports high-speed adapters also can be configured to allocate more memory to improve I/O performance. Linux is the only operating system that uses these extra TCEs so that the memory can be freed for use by partitions if the system uses only AIX.

Memory usage for virtualization features

Virtualization requires more memory to be allocated by the POWER Hypervisor for hardware statesave areas and various virtualization technologies. For example, on Power10 processor-based systems, each processor core supports up to eight simultaneous multithreading (SMT) threads of execution, and each thread contains over 80 different registers.

The POWER Hypervisor must set aside save areas for the register contents for the maximum number of virtual processors that are configured. The greater the number of physical hardware devices, the greater the number of virtual devices, the greater the amount of virtualization, and the more hypervisor memory is required. For efficient memory consumption, wanted and maximum values for various attributes (processors, memory, and virtual adapters) must be based on business needs, and not set to values that are significantly higher than actual requirements.

Predicting memory that is used by the POWER Hypervisor

The IBM System Planning Tool (SPT) is a resource that can be used to estimate the amount of hypervisor memory that is required for a specific server configuration. After the SPT executable file is downloaded and installed, you can define a configuration by selecting the correct hardware platform and the installed processors and memory, and defining partitions and partition attributes. SPT can estimate the amount of memory that is assigned to the hypervisor, which assists you when you change a configuration or deploy new servers.

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

► Virtual SCSI

   The POWER Hypervisor provides a virtual SCSI mechanism for the virtualization of storage devices. The storage virtualization is accomplished by using two paired adapters: a virtual SCSI server adapter and a virtual SCSI customer adapter.

► Virtual Ethernet

   The POWER Hypervisor provides a virtual Ethernet switch function that allows partitions fast and secure communication on the same server without any need for physical interconnection or connectivity outside of the server if a Layer 2 bridge to a physical Ethernet adapter is set in one VIOS partition, also known as *Shared Ethernet Adapter* (SEA).

► Virtual Fibre Channel

   A virtual Fibre Channel adapter is a virtual adapter that provides customer LPARs with a Fibre Channel connection to a storage area network through the VIOS partition. The VIOS partition provides the connection between the virtual Fibre Channel adapters on the VIOS partition and the physical Fibre Channel adapters on the managed system.

► Virtual (TTY) console

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

Logical partitions

LPARs and virtualization increase the usage of system resources and add a level of configuration possibilities.

Logical partitioning is the ability to make a server run as though it were two or more independent servers. When you logically partition a server, you divide the resources on the server into subsets, called *LPARs*. You can install software on an LPAR, and the LPAR runs as an independent logical server with the resources that you allocated to the LPAR.

LPAR also is referred to in some documentation as a *virtual machine* (VM), which making it look similar to what other hypervisors offer. However, LPARs provide a higher level of security and isolation and other features that are described in this chapter.

Processors, memory, and I/O devices can be assigned to LPARs. AIX, IBM i, Linux, and VIOS can run on LPARs. VIOS provides virtual I/O resources to other LPARs with general-purpose operating systems.

LPARs share a few system attributes, such as the system serial number, system model, and processor FCs. All other system attributes can vary from one LPAR to another.

Micro-Partitioning

When you use the Micro-Partitioning technology, you can allocate fractions of processors to an LPAR. An LPAR that uses fractions of processors is also known as a *shared processor partition* or *micropartition*. Micropartitions run over a set of processors that is called a *shared processor pool* (SPP), and virtual processors are used to enable the operating system manage the fractions of processing power that are assigned to the LPAR.

From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system is enhanced to determine the difference. Physical processors are abstracted into virtual processors that are available to partitions.

On the Power10 processor-based server, a partition can be defined with a processor capacity as small as 0.05 processing units. This number represents 0.05 of a physical core. Each physical core can be shared by up to 20 shared processor partitions, and the partition's entitlement can be incremented fractionally by as little as 0.05 of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under the control of the Power Hypervisor. The shared processor partitions are created and managed by the HMC.

The Power E1080 supports up to 240 cores in a single system and 1000 micropartitions (1000 is the maximum that PowerVM supports).

> **Note:** Although the Power E1080 supports up to 1000 micropartitions, the real limit depends on application workload demands in use on the server.

Processing mode

When you create an LPAR, you can assign entire processors for dedicated use, or you can assign partial processing units from an SPP. This setting defines the processing mode of the LPAR.

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to partitions. The SMT feature in the Power10 processor core allows the core to run instructions from two, four, or eight independent software threads simultaneously.

Shared dedicated mode

On Power10 processor-based servers, you can configure dedicated partitions to become processor donors for idle processors that they own, which allows for the donation of spare CPU cycles from dedicated processor partitions to an SPP. The dedicated partition maintains absolute priority for dedicated CPU cycles. Enabling this feature can help increase system usage without compromising the computing power for critical workloads in a dedicated processor mode LPAR.

Shared mode

In shared mode, LPARs use virtual processors to access fractions of physical processors. Shared partitions can define any number of virtual processors (the maximum number is 20 times the number of processing units that are assigned to the partition). The Power Hypervisor dispatches virtual processors to physical processors according to the partition's processing units entitlement. One processing unit represents one physical processor's processing capacity. All partitions receive a total CPU time equal to their processing unit's entitlement. The logical processors are defined on top of virtual processors. Therefore, even with a virtual processor, the concept of a logical processor exists, and the number of logical processors depends on whether SMT is turned on or off.

3.1.2 Multiple shared processor pools

MSPPs are supported on Power10 processor-based servers. This capability allows a system administrator to create a set of micropartitions with the purpose of controlling the processor capacity that can be used from the physical SPP.

Micropartitions are created and then identified as members of the default processor pool or a user-defined SPP. The virtual processors that exist within the set of micropartitions are monitored by the Power Hypervisor. Processor capacity is managed according to user-defined attributes.

If the Power server is under heavy load, each micropartition within an SPP is assured of its processor entitlement, plus any capacity that might be allocated from the reserved pool capacity if the micropartition is uncapped.

If specific micropartitions in an SPP do not use their processing capacity entitlement, the unused capacity is ceded and other uncapped micropartitions within the same SPP can use the extra capacity according to their uncapped weighting. In this way, the entitled pool capacity of an SPP is distributed to the set of micropartitions within that SPP.

All Power servers that support the MSPPs capability have a minimum of one (the default) SPP and up to a maximum of 64 SPPs.

This capability helps customers reduce total cost of ownership (TCO) significantly when the cost of software or database licenses are dependant on the number of assigned processor-cores.

3.1.3 Virtual I/O Server

The Virtual I/O Server (VIOS) is part of PowerVM. It is the specific appliance that allows the sharing of physical resources among LPARs to allow more efficient usage (for example, consolidation). In this case, the VIOS owns the physical I/O resources (SCSI, Fibre Channel, network adapters, or optical devices) and allows customer partitions to share access to them, which minimizes and optimizes the number of physical adapters in the system.

The VIOS eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The VIOS supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses.
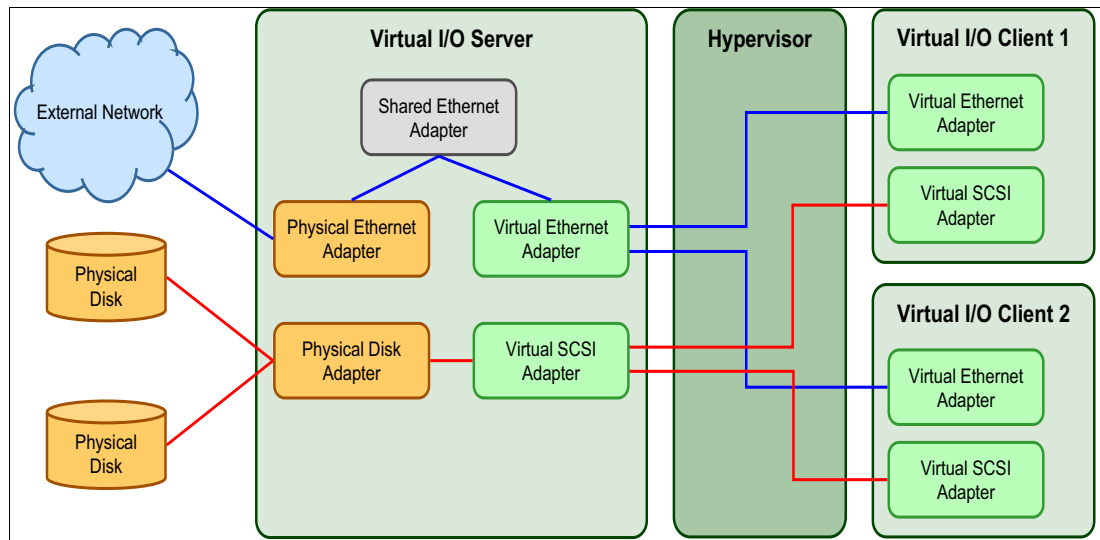
Figure 3-1 shows an overview of a VIOS configuration.

Figure 3-1   Architectural view of the VIOS

It is a preferred practice to run dual VIO servers per physical server.

Shared Ethernet Adapter

A Shared Ethernet Adapter (SEA) can be used to connect a physical Ethernet network to a virtual Ethernet network. The SEA provides this access by connecting the Power Hypervisor VLANs to the VLANs on the external switches. Because the SEA processes packets at Layer 2, the original MAC address and VLAN tags of the packet are visible to other systems on the physical network. IEEE 802.1 VLAN tagging is supported.

By using the SEA, several customer partitions can share one physical adapter. You also can connect internal and external VLANs by using a physical adapter. The SEA service can be hosted only in the VIOS (not in a general-purpose AIX or Linux partition) and acts as a Layer 2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (Etherchannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the Power Hypervisor on the VIOS.

Virtual SCSI

Virtual SCSI is used to view a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a customer/server relationship. The VIOS LPAR owns the physical I/O resources and acts as a server or, in SCSI terms, a target device. The client LPARs access the virtual SCSI backing storage devices that are provided by the VIOS as clients.

The virtual I/O adapters (a virtual SCSI server adapter and a virtual SCSI client adapter) are configured by using an HMC. The virtual SCSI server (target) adapter is responsible for running any SCSI commands that it receives, and is owned by the VIOS partition. The virtual SCSI client adapter allows a client partition to access physical SCSI and SAN-attached devices and LUNs that are mapped to be used by the client partitions. The provisioning of virtual disk resources is provided by the VIOS.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a technology that allows multiple LPARs to access one or more external physical storage devices through the same physical Fibre Channel adapter. This adapter is attached to a VIOS partition that acts only as a pass-through that manages the data transfer through the Power Hypervisor.

Each partition features one or more virtual Fibre Channel adapters, each with their own pair of unique worldwide port names. This configuration enables you to connect each partition to independent physical storage on a SAN. Unlike virtual SCSI, only the client partitions see the disk.

For more information and requirements for NPIV, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

3.1.4 Live Partition Mobility

LPM allows you to move a running LPAR from one system to another without disruption. Inactive partition mobility allows you to move a powered-off LPAR from one system to another one.

LPM provides systems management flexibility and improves system availability by avoiding the following situations:

► Planned outages for hardware upgrade or firmware maintenance.
► Unplanned downtime. With preventive failure management, if a server indicates a potential failure, you can move its LPARs to another server before the failure occurs.

For more information and requirements for LPM, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

HMCV10R1 and VIOS 3.1.3 or later provide the following enhancements to the Live Partition Mobility Feature:

► Automatically choose fastest network for LPM memory transfer
► Allow LPM when a virtual optical device is assigned to a partition

3.1.5 Active Memory Expansion

Active Memory Expansion (AME) is an optional feature code (#EM8F) that belongs to the technologies under the PowerVM umbrella and enables memory expansion on the system.

AME is an innovative technology that supports the AIX operating system. It helps enable the effective maximum memory capacity to be larger than the true physical memory maximum. Compression and decompression of memory content can enable memory expansion up to 100% or more. This expansion can enable a partition to complete significantly more work or support more users with the same physical amount of memory. Similarly, it can enable a server to run more partitions and do more work for the same physical amount of memory.

AME uses CPU resources to compress and decompress the memory contents. The trade-off of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies about how compressible the memory content is. It also depends on having adequate spare CPU capacity available for this compression and decompression.

The Power E1080 includes a hardware accelerator that is designed to boost AME efficiency and uses less processor core resources. Each AIX partition can turn on or turn off AME. Control parameters set the amount of expansion that is wanted in each partition to help control the amount of CPU used by the AME function.

An IPL is required for the specific partition that is turning memory expansion. When enabled, monitoring capabilities are available in standard AIX performance tools, such as `lparstat`, `vmstat`, `topas`, and `svmon`.

A planning tool is included with AIX, which enables you to sample workloads and estimate how expandable the partition's memory is and much CPU resource is needed. The feature can be ordered with the initial order of the Power E1080 or as an MES order. A software key is provided when the enablement feature is ordered, which is applied to the system node. An IPL is not required to enable the system node. The key is specific to an individual system and is permanent. It cannot be moved to a different server.

> **Note:** IBM i does not support AME.

3.1.6 Remote Restart

Remote Restart is a high availability option for partitions. If an error occurs that causes a server outage, a partition that is configured for Remote Restart can be restarted on a different physical server. At times, it might take longer to start the server, in which case the Remote Restart function can be used for faster reprovisioning of the partition. Typically, this can be done faster than restarting the server that stopped and then restarting the partitions. The Remote Restart function relies on technology that is similar to LPM where a partition is configured with storage on a SAN that is shared (accessible) by the server that hosts the partition.

HMC V10R1 provides an enhancement to the Remote Restart Feature that enables remote restart when a virtual optical device is assigned to a partition.

3.1.7 POWER processor modes

Although they are not virtualization features, the POWER processor modes are described here because they affect various virtualization features.

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

► POWER8

This native mode for POWER8 processors implements Version 2.07 of the IBM Power ISA. For more information, see IBM documentation.

► POWER9

This native mode for POWER9 processors implements Version 3.0 of the IBM Power ISA. For more information, see IBM documentation.

► POWER10

This native mode for Power10 processors implements Version 3.1 of the IBM Power ISA. For more information, see IBM documentation.

Figure 3-2 shows the available processor modes on a Power E1080.

Logical Partition Profile Properties: default_profile @ vm112

| General | **Processors** | Memory | I/O | Virtual Adapters | Power Controlling | Settings |

Detailed below are the current processing settings for this partition profile.

Processing mode
○ Dedicated
◉ Shared

Processing units
Total managed system processing units : 48.00
Minimum shared processing units : `0.1`
Desired shared processing units : `1.0`
Maximum shared processing units : `8.0`
Shared processor pool: `DefaultPool (0)` ▼

Virtual processors
Minimum processing units required for each virtual processor : 0.10
Newer operating system levels support : 0.05
Minimum virtual processors : `1.0`
Desired virtual processors : `2.0`
Maximum virtual processors : `8.0`

Sharing mode
☑ Uncapped   Weight : `128` ▲▼

Processor compatibility mode: `default` ▼

| OK | Cancel | Help |

default
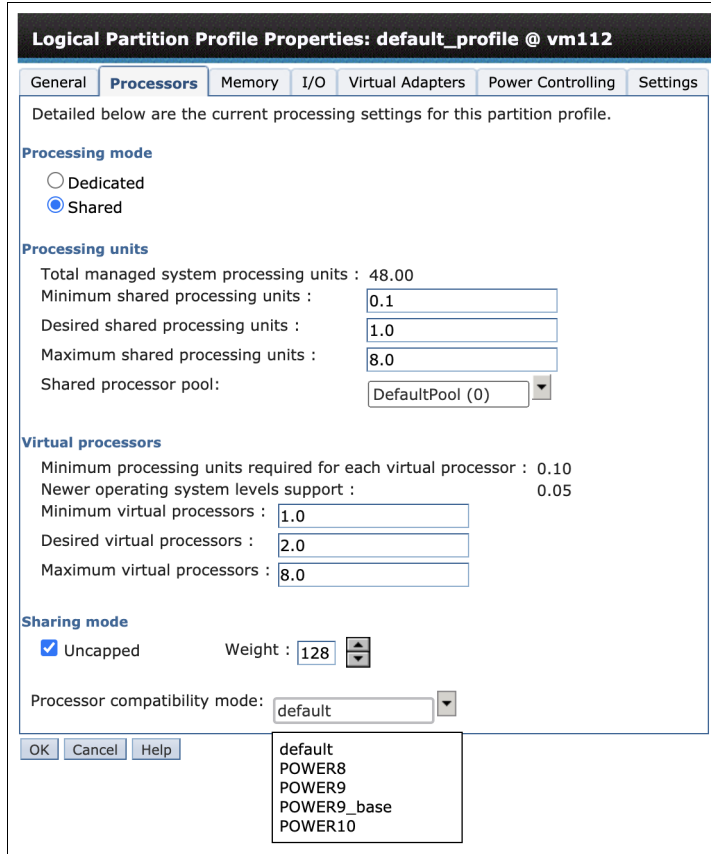POWER8
POWER9
POWER9_base
POWER10

Figure 3-2   Processor modes

Processor compatibility mode is important when LPM migration is planned between different generation of servers. An LPAR that might be migrated to a machine that is managed by a processor from another generation must be activated in a specific compatibility mode.

> **Note:** Migrating a LPAR from a POWER7 processor-based server to Power E1080 by using LPM is not supported; however, the following steps can be completed to accomplish this task:
>
> 1. Migrate LPAR from POWER7 processor-based server to POWER9 processor-based server using LPM.
>
> 2. Migrate then the LPAR from POWER9 processor-based server to Power E1080.

3.1.8 Single Root I/O Virtualization

Single Root I/O Virtualization (SR-IOV) is an extension to the Peripheral Component Interconnect Express (PCIe) specification that allows multiple operating systems to simultaneously share a PCIe adapter with little or no runtime involvement from a hypervisor or other virtualization intermediary.

SR-IOV is PCI standard architecture that enables PCIe adapters to become self-virtualizing. It enables adapter consolidation through sharing, much like logical partitioning enables server consolidation. With an adapter capable of SR-IOV, you can assign virtual *slices* of a single physical adapter to multiple partitions through logical ports; all of this is done without a VIOS.

3.1.9 More information about virtualization features

The following IBM Redbooks publications provide more information about the virtualization features:

► *IBM PowerVM Best Practices*, SG24-8062
► *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
► *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
► *IBM Power Systems SR-IOV: Technical Overview and Introduction*, REDP-5065

3.2 IBM PowerVC overview

IBM Power Virtualization Center (PowerVC) is an advanced virtualization and cloud management offering (which is built on OpenStack) that provides simplified virtualization management and cloud deployments for IBM AIX, IBM i, and Linux VMs running on IBM Power Systems. PowerVC is designed to improve administrator productivity and simplify the cloud management of VMs on Power Systems servers.

By using PowerVC, you can perform the following tasks:

► Create VMs and resize the VMs CPU and memory.
► Attach disk volumes or other networks to those VMs.
► Import VMs and volumes so that they can be managed by IBM PowerVC.
► Monitor the use of resources in your environment.
► Take snapshots of or cone a VM.
► Migrate VMs while they are running (live migration between physical servers).
► Remote restart VMs in case of a server failure.
► Use advanced storage technologies, such as VDisk mirroring or IBM Global mirror.
► Improve resource usage to reduce capital expense and power consumption.
► Increase agility and execution to respond quickly to changing business requirements.
► Increase IT productivity and responsiveness.
► Simplify Power Systems virtualization management.
► Accelerate repeatable, error-free virtualization deployments.

IBM PowerVC can manage AIX, IBM i, and Linux-based VMs that are running under PowerVM virtualization that are connected to an HMC or by using NovaLink. This release supports the scale-out and the enterprise Power Systems servers that are built on IBM POWER8, IBM POWER9, and subsequent technologies.

3.2.1 IBM PowerVC functions and advantages

Why IBM PowerVC? Why do you need another virtualization management offering? When more than 70% of IT budgets are spent on operations and maintenance, IT customers legitimately expect vendors to focus their new development efforts to reduce IT costs and foster innovation within IT departments.

IBM PowerVC gives IBM Power Systems customers the following advantages:

► It is deeply integrated with Power Systems.

► It provides virtualization management tools.

► It eases the integration of servers that are managed by PowerVM in automated IT environments, such as clouds.

► It is a building block of IBM Infrastructure as a Service (IaaS), based on Power Systems.

► PowerVC integrated with other cloud management tool, such as Ansible, Terraform, or OpenShift, and can be integrated into orchestration tools, such as the IBM Cloud® Automation Manager, VMware vRealize, or the SAP Landscape Management (LaMa).

► PowerVC also provides an easy exchange of VM images between private and public clouds.

IBM PowerVC is an addition to the PowerVM set of enterprise virtualization technologies that provide virtualization management. It is based on open standards and integrates server management with storage and network management.

Because IBM PowerVC is based on the OpenStack initiative, Power Systems can be managed by tools that are compatible with OpenStack standards. When a system is controlled by IBM PowerVC, it can be managed in one of three ways:

► By a system administrator by using the IBM PowerVC graphical user interface (GUI)

► By a system administrator that uses scripts that contain the IBM PowerVC Representational State Transfer (REST) application programming interfaces (APIs)

► By higher-level tools that call IBM PowerVC by using standard OpenStack API

The following PowerVC offerings are positioned within the available solutions for Power Systems cloud:

► IBM PowerVC: Advanced Virtualization Management
► IBM PowerVC for Private Cloud: Basic Cloud
► IBM Cloud Automation Manager: Advanced Cloud
► VMware vRealize: Advanced Cloud

PowerVC provides a systems management product that enterprise customers require to manage effectively the advanced features that are offered by IBM premium hardware. It reduces resource use and manages workloads for performance and availability.

3.3 System automation with Ansible

Enterprises can spend many precious administrative time performing repetitive tasks and using manual processes. Tasks, such as updating, patching, compliance checks, provisioning new virtual machines or LPARs, and ensuring that the correct security updates are in place, are taking time away from more valuable business activities.

The ability to automate by using Ansible returns valuable time to the system administrators.

Red Hat Ansible Automation Platform for Power is fully enabled, so enterprises can automate several tasks within AIX, IBM i, and Linux that include deploying applications. Ansible also can be combined with HMC, PowerVC, and Power Virtual Server to provision infrastructure anywhere you need, including cloud solutions from other IBM Business Partners or third-party providers based on Power processor-based servers.

A first task after the initial installation or set up of a new LPAR is to ensure that the correct patches are installed. Also, extra software (whether it is open source software, ISV software, or perhaps their own enterprise software) must be installed. Ansible features a set of capabilities to roll out new software, which makes it popular in Continuous Delivery/Continuous Integration (CD/CI) environments. Orchestration and integration of automation with security products represent other ways in which Ansible can be applied within the data center.

Despite the wide adoption of AIX and IBM i in many different business sectors by different types of customers, Ansible can help introduce the Power processor-based technology to customers who believe that AIX and IBM i skills are a rare commodity that are difficult to locate in the marketplace but want to take advantage of all the features of the hardware platform. The Ansible experience is identical across Power or x86 processor-based technology and the same steps can be repeated in IBM Cloud.

AIX and IBM i skilled customers also can benefit from the extreme automation solutions that are provided by Ansible.

The Power processor-based architecture features unique advantages over commodity server platforms, such as x86, because the engineering teams that are working on the processor, motherboards, virtualization. and management appliances collaborate closely to ensure an integrated stack that works seamlessly. This approach is in stark contrast to the multivendor x86 processor-based technology approach, in which the processor, server, management, and virtualization must be purchased from different (and sometimes competing) vendors.

The Power stack engineering teams partnered closely to deliver the enterprise server platform, which results in an IT architecture with industry-leading performance, scalability, and security (see Figure 3-3).
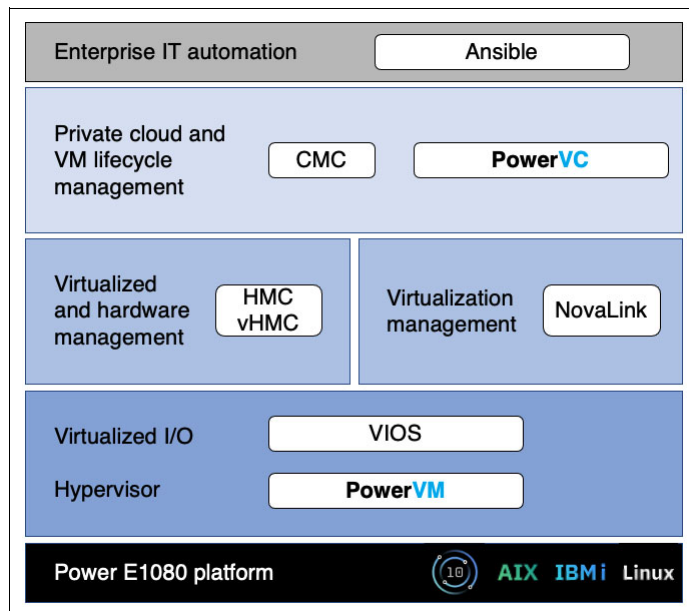

Figure 3-3   Power stack

Every layer in the Power stack is optimized to make the Power10 processor-based technology the platform of choice for mission-critical enterprise workloads. This stack includes the Ansible Automation Platform, which is described next.

3.3.1 Ansible Automation Platform

Ansible Automation Platform integrates with IBM Power processor-based technology, which is included in the Certified Integrations section of the Red Hat Ansible web site.

The various Ansible collections for IBM Power processor-based technology, which (at the time of this writing) were downloaded more than 25,000 times by customers, are now included in the Red Hat Ansible Automation Platform. As a result, these modules are covered by Red Hat's 24x7 enterprise support team, which collaborates with the respective Power processor-based technology development teams.

3.3.2 Power Systems in the Ansible ecosystem

A series of Ansible collections is available for the Power processor-based platform. A *collection* is a group of modules, Playbooks, and roles. Embracing Ansible in the Power community, our AIX and IBM i community have the right set of modules available. Some examples are development of tools to manage AIX logical volumes, which put module interfaces over the installation key command, or managing the AIX init tab entries.

From an IBM i perspective, an example is the ability to run SQL queries against the integrated IBM Db2® database that comes built into the IBM i platform, manage object authorities, and so on. All of these things mean something to an AIX administrator or IBM i administrator.

Our operating system teams develop modules that are sent to open source community (named Ansible Galaxy). Every developer can post any object that can be a candidate for a collection in the open Ansible Galaxy community and possibly certified to be supported by IBM with a subscription to Red Hat Ansible Automation Platform (see Figure 3-4).
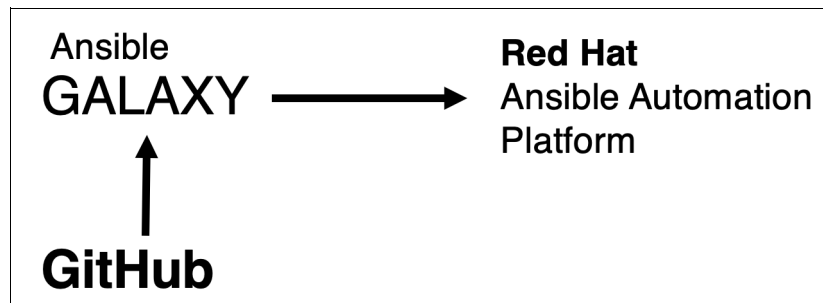

Figure 3-4   Power content in the Ansible ecosystem

3.3.3 Ansible modules for AIX

IBM Power Systems AIX collection, as part of the broader offering of Ansible Content for IBM Power Systems, is available from Ansible Galaxy and has community support.

The collection includes modules and sample playbooks that help to automate tasks and is available starting at hthis web page.

3.3.4 Ansible modules for IBM i

Ansible Content for IBM Power Systems - IBM i provides modules, action plug-ins, roles, and sample playbooks to automate tasks on IBM i, including the following example:

► Command execution
► System and application configuration
► Work management
► Fix management
► Application deployment

For more information about the collection, see this web page.

3.3.5 Ansible modules for HMC

IBM Power Systems HMC collection provides modules that can be used to manage configurations and deployments of HMC systems. The collection content helps to include workloads on Power processor-based platforms as part of an enterprise automation strategy through the Ansible ecosystem.

For more information about this collection, see this web page.

3.3.6 Ansible modules for VIOS

IBM Power Systems VIOS collection provides modules that can be used to manage configurations and deployments of Power VIOS systems. The collection content helps to include workloads on Power processor-based platforms as part of an enterprise automation strategy through the Ansible system.

For more information, see this web page.

3.4 Protect trust from core to cloud

Although enforcing a data encryption policy is an effective way to minimize the risk of a data breach that, in turn, minimizes costs, at the time of this writing, only 17% of enterprises surveyed indicated that they protected more than 50% of their sensitive data in cloud with encryption[1]. Only a few enterprises at the world-wide level have an encryption strategy applied consistently across the entire organization, largely because it adds complexity, costs, and negatively affects performance, which means missed SLAs to the business.

The rapidly evolving cyber threat landscape requires focus on cyber-resilience. Persistent and end-to-end security is the only way to reduce exposure to threats.

Power processor-based platforms always offered the most secure and reliable servers in its class. The Power E1080 further extends the industry-leading security and reliability of the Power processor-based platform, with focus on protecting applications and data across all of the hybrid cloud environments. It also introduces significant innovations along the following major dimensions:

► Advanced Data Protection that offers simple to use and efficient capabilities to protect sensitive data through mechanisms, such as encryption and multi-factor authentication.

---

[1] source: Thales Data Threat Report - Global Edition:
https://cpl.thalesgroup.com/resources/encryption/2021/data-threat-report

- Platform Security ensures that the server is hardened against tampering, continuously protects its integrity, and ensures strong isolation among multi-tenant workloads. Without strong platform security, all other system security measures are at risk.
- Security Innovation for Modern Threats provides the ability to stay ahead of new types of cybersecurity threats by using emerging technologies.
- Integrated Security Management addresses the key challenge of ensuring correct configuration of the many security features across the stack, monitoring them, and reacting if unexpected changes are detected.

The Power E1080 is enhanced to simplify and integrate security management across the stack, which reduces the likelihood of administrator errors.

In the Power E1080, all data is protected by a greatly simplified end-to-end encryption that extends across the hybrid cloud without detectable performance impact and prepares for future cyber threats.

Power10 processor-core technology features built-in security integration:

- Stay ahead of current and future data threats with better cryptographic performance and support for quantum-safe cryptography and fully homomorphic encryption (FHE).
- Enhance the security of applications with more hardened defense against return-oriented programming (ROP) attacks.
- Simplified single-interface hybrid cloud security management without any required setup.
- Protect your applications and data with the most secure VM isolation in the industry with orders of magnitude lower Common Vulnerability Exposures (CVEs) than hypervisors related to x86 processor-based servers.

Also, workloads on the Power E1080 benefit from cryptographic algorithm acceleration, which allows algorithms, such as Advanced Encryption Standard (AES), SHA2, and SHA3 to run significantly faster than POWER9 processor-based servers on a per core basis. This performance acceleration allows features, such as AIX Logical Volume Encryption, to be enabled with low performance overhead.

3.4.1 Crypto engines and transparent memory encryption

Power10 processor technology is engineered to achieve significantly faster encryption performance with quadruple the number of AES encryption engines. In comparison to IBM POWER9 processor-based servers, Power E1080 is updated for today's most demanding standards and anticipated future cryptographic standards, such as post-quantum and fully homomorphic encryption (FHE), and brings new enhancements to container security.

Transparent memory encryption is designed to simplify encryption and support end-to-end security without affecting performance by using hardware features for a seamless user experience. The protection that is introduced in all layers of an infrastructure is shown in Figure 3-5 on page 151.
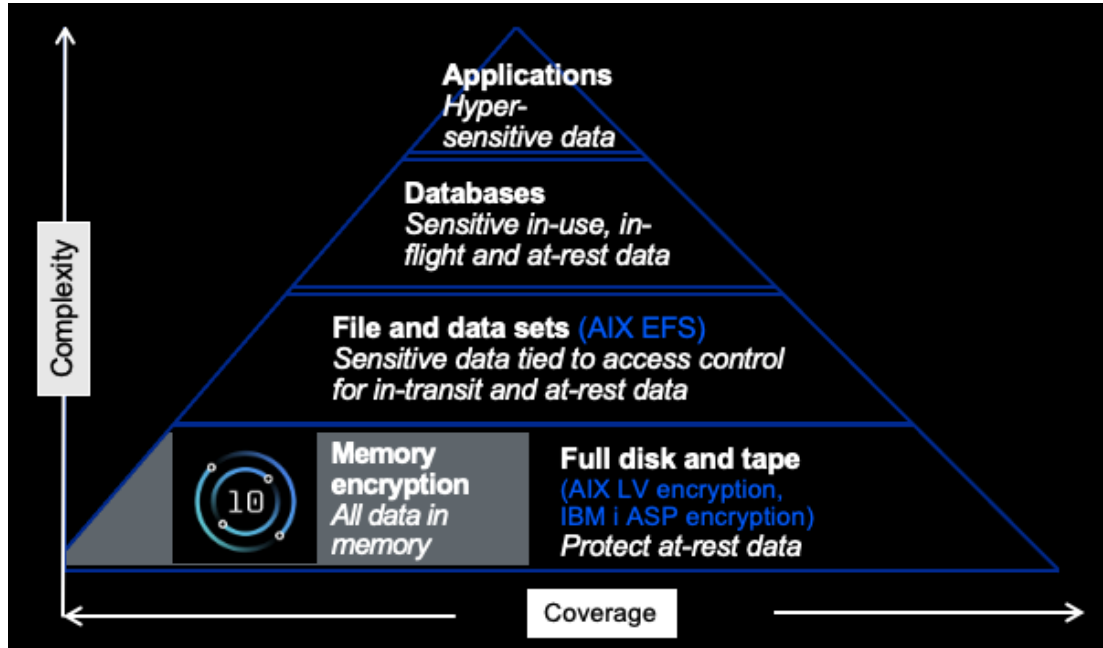
Figure 3-5   Protect data in memory with transparent memory encryption

3.4.2 Quantum-safe cryptography support

To be prepared for the Quantum era, the Power E1080 is built to efficiently support future cryptography, such as Quantum-safe cryptography and FHE. The software libraries for these solutions are optimized for the Power10 processor-chip instruction set architecture (ISA) and are or will be available in the respective open source communities.

Quantum-safe cryptography refers to the efforts to identify algorithms that are resistant to attacks by classical and quantum computers in preparation for the time when large-scale quantum computers are built.

Homomorphic encryption refers to encryption techniques that permit systems to perform computations on encrypted data without decrypting the data first. The software libraries for these solutions are optimized for the Power processor-chip instruction set architecture (ISA).

3.4.3 IBM PowerSC support

The Power E1080 benefits from the integrated security management capabilities that are offered by:

►   IBM PowerSC

►   The Power software portfolio for managing security and compliance on every Power processor-based platform that is running AIX

►   IBM i or the supported distributions

►   Versions of Linux

PowerSC is introducing more features to help customers manage security end-to-end across the stack to stay ahead of various threats. Specifically, PowerSC 2.0 adds support for Endpoint Detection and Response (EDR), host-based intrusion detection, block listing, and Linux.

3.5 Running artificial intelligence where operational data is stored

Separate platforms for AI and business applications make deploying AI in production difficult,. The result is reduced end-to-end availability for applications and data access, risk of violating service level agreements because of overhead of sending operational data and receiving predictions from external AI platforms, and increased complexity and cost of managing different environments and external accelerators.

As the world is set to deploy AI everywhere, attention is turning from how fast to build AI to how fast to inference with the AI.

To support this shift, the Power E1080 delivers faster business insights by running AI *in place* with four Matrix Math Assist (MMA) units to accelerate AI in each Power10 technology-based processor-core. The robust execution capability of the processor cores with MMA AI acceleration, enhanced SIMD, and enhanced data bandwidth provide an alternative to external accelerators, such as GPUs, and related device management for execution of statistical machine learning and inferencing workloads. These features, combined with the possibility to consolidate multiple environments for AI model execution on a Power E1080 platform together with other different types of environments, reduces costs and leads to a greatly simplified solution stack for AI.

Operationalizing AI inferencing directly on a Power E1080 brings AI closer to data. This ability allows AI to inherit and benefit from the Enterprise Qualities of Service (QoS): reliability, availability, and security of the Power10 processor-based platform and support a performance boost. Enterprise business workflows can now readily and consistently use insights that are built with the support of AI.

The use of data gravity on Power10 processor-cores enables AI to run during a database operation or concurrently with an application, for example. This feature is key for time-sensitive use cases. It delivers fresh input data to AI faster and enhances the quality and speed of insight.

As no-code application development, pay-for-use model repositories, auto-machine learning, and AI-enabled application vendors continue to evolve and grow, the corresponding software products are brought over to the Power10 platform. Python and code from major frameworks and tools, such as TensorFlow, PyTorch, and XGBoost, run on the Power10 processor-based platform without any changes.

Open Neural Network Exchange (ONNX) models can be brought over from x86 processor-based servers or other platforms and small-sized virtual machines or Power Virtual Server (PowerVS), for deployment on Power E1080, which gives customers the ability to build on commodity hardware but deploy on enterprise servers.

Power10 processor-core architecture includes an embedded matrix math accelerator (MMA). This MMA is extrapolated for an E1080 system to provide up to 5x faster AI inference for FP32 to infuse AI into business applications and drive greater insights, or up to 10x faster AI inference by using reduced precision data types, such as Bfloat16 and INT8, when compared with a prior generation Power9-based server.

IBM optimized math libraries so that AI tools benefit from acceleration that is provided by the MMA units of the Power10 chip. The benefits of MMA action can be realized for statistical machine learning and inferencing, which provides a cost-effective alternative to external accelerators or GPUs.

3.5.1 Train anywhere, deploy on Power E1080

IBM Power10 processor-based technology addresses challenges through hardware and software innovation. Models that are trained on any cloud that is based on Power or x86 processor-based servers (even with different endian-ness) can be deployed on the Power E1080 and run without any changes.

Power10 cores are equipped with four Math Engines for matrix and tensor math. Applications can run models with colocated data without the need for external accelerators, GPUs, or extra AI platforms. Power10 technology uses the "train anywhere, deploy here" principle to operationalize AI.

A model can be trained on a public or private cloud (see Figure 3-6) by using the following procedure:

1. The model is registered with its version in the so-called *model vault*, which is a VM or LPAR with tools, such as Watson Openscale, BentoML, or Tensorflow Serving.

2. The model is pushed out to the destination (in this case, it is a VM or an LPAR running a database with an application) and the model might be used by the database or the application.

3. Transactions that are received by the database and application trigger model execution and generate predictions or classifications. These predictions also can be stored locally. For example, these predictions can be the risk or fraud that is associated with the transaction or product classifications to be used by downstream applications.

4. A copy of the model output (prediction or classification) is sent to the model operations (ModelOps) engine for calculation of drift by comparison with Ground Truth.

5. If the drift exceeds a threshold, the model retrain triggers are generated.

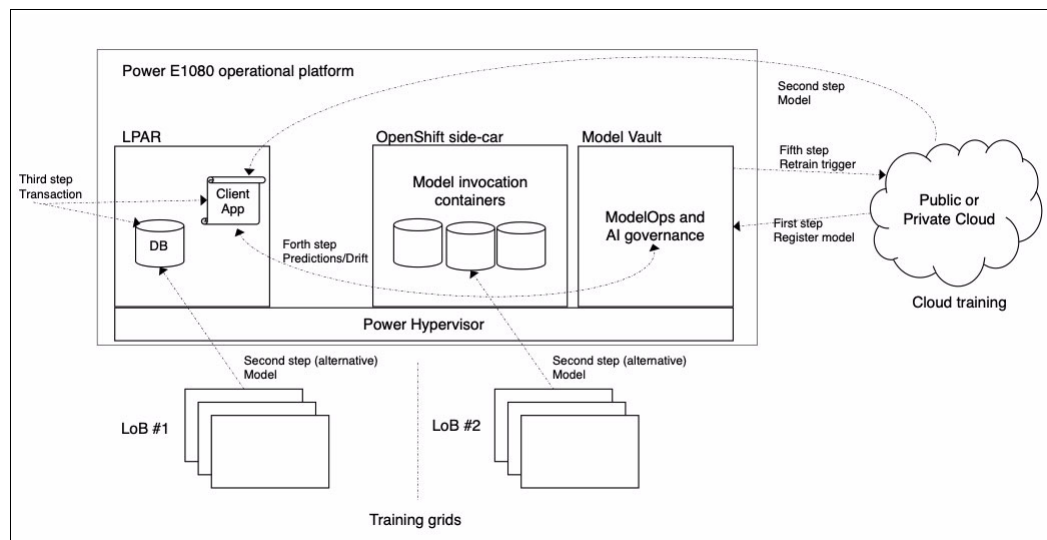6. Retrained models are then taken through steps 1 - 5.

Figure 3-6   Operationalizing AI

By using a combination of software and hardware innovation, Power E1080 can meet the model performance, response time, and throughput KPIs of databases and applications that are infused with AI.

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only:

► *IBM Power Systems Private Cloud with Shared Utility Capacity: Featuring Power Enterprise Pools 2.0,* SG24-8476

► *SAP HANA Data Management and Performance on IBM Power Systems*, REDP-5570

► *IBM PowerAI: Deep Learning Unleashed on IBM Power Systems Servers*, SG24-8409

► *IBM Power System AC922 Technical Overview and Introduction*, REDP-5494

► *IBM Power System E950: Technical Overview and Introduction*, REDP-5509

► *IBM Power System E980: Technical Overview and Introduction*, REDP-5510

► *IBM Power System L922 Technical Overview and Introduction*, REDP-5496

► *IBM Power System S822LC for High Performance Computing Introduction and Technical Overview*, REDP-5405

► *IBM Power Systems H922 and H924 Technical Overview and Introduction*, REDP-5498

► *IBM Power Systems LC921 and LC922: Technical Overview and Introduction*, REDP-5495

► *IBM Power Systems S922, S914, and S924 Technical Overview and Introduction Featuring PCIe Gen 4 Technology, REDP-5595*

► *IBM PowerVM Best Practices*, SG24-8062

► *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940

► *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

► *IBM PowerVC Version 2.0 Introduction and Configuration*, SG24-8477

You can search for, view, download, or order these documents and other Redbooks publications, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

The following websites are also relevant as further information sources:

- ► IBM Fix Central:

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

- ► IBM documentation:

  https://www.ibm.com/docs/en

- ► IBM Power10 documentation:

  https://www.ibm.com/docs/en/power10

- ► IBM Portal for OpenPOWER - POWER9 Monza Module:

  https://www.ibm.com/systems/power/openpower/tgcmDocumentRepository.xhtml?aliasId=POWER9_Monza

- ► IBM Power:

  https://www.ibm.com/it-infrastructure/power

- ► IBM Storage:

  https://www.ibm.com/it-infrastructure/storage

- ► IBM Systems Energy Estimator:

  https://ibm.biz/energy-estimator

- ► IBM System Planning Tool:

  https://www.ibm.com/support/pages/ibm-system-planning-tool-power-processor-based-systems-0

- ► OpenCAPI:

  http://opencapi.org/technical/use-cases/

- ► OpenPOWER Foundation:

  https://openpowerfoundation.org/

- ► Power Systems Capacity on Demand:

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

- ► Support for IBM Systems:

  https://www.ibm.com/mysupport/s/?language=en_US

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM ®

Printed in U.S.A.

Redbooks ®

ibm.com/redbooks