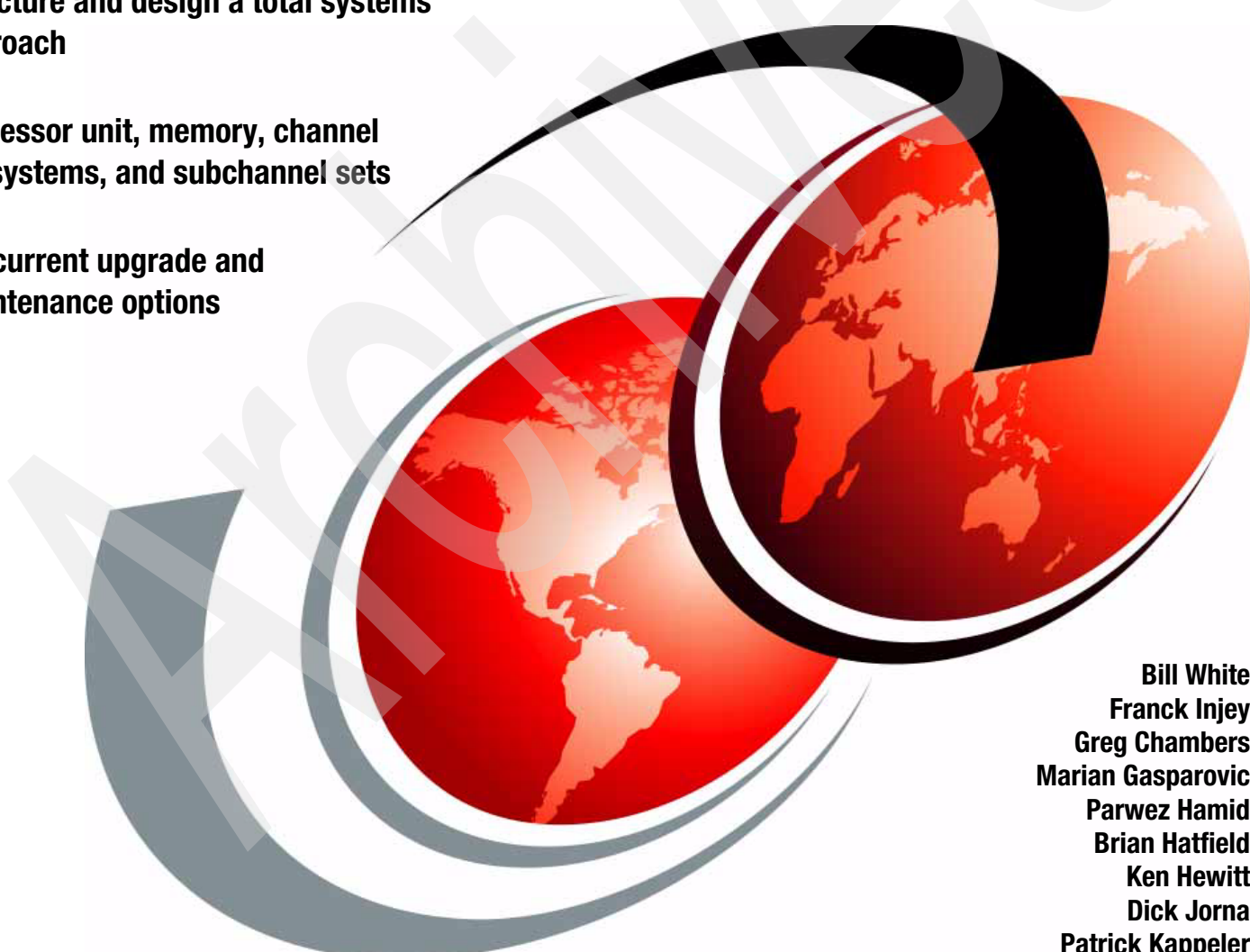


# IBM System z9 Enterprise Class Technical Guide

Structure and design a total systems approach

Processor unit, memory, channel subsystems, and subchannel sets

Concurrent upgrade and maintenance options



Bill White  
Franck Inje  
Greg Chambers  
Marian Gasparovic  
Parwez Hamid  
Brian Hatfield  
Ken Hewitt  
Dick Jorna  
Patrick Kappeler

**Redbooks**





International Technical Support Organization

**IBM System z9 Enterprise Class Technical Guide**

June 2007

Archived

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

Archived

**Third Edition (June 2007)**

This edition (SG24-7124-02) applies to the IBM® System z9 Enterprise Class server.

© Copyright International Business Machines Corporation 2005, 2006, 2007. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

# Contents

<b>Notices</b> .....	vii
Trademarks .....	viii
<b>Preface</b> .....	ix
The team that wrote this book .....	ix
Become a published author .....	x
Comments welcome .....	xi
<b>Chapter 1. Overview</b> .....	1
1.1 Introduction .....	2
1.2 z9 EC models .....	4
1.3 System functions and features .....	6
1.3.1 The CEC cage .....	7
1.3.2 I/O connectivity .....	10
1.3.3 Cryptography .....	12
1.3.4 Performance .....	14
1.3.5 Parallel Sysplex support .....	16
1.3.6 Server Time Protocol .....	17
1.3.7 Intelligent Resource Director (IRD) .....	18
1.3.8 Capacity On Demand .....	19
1.3.9 Reliability, Availability, and Serviceability (RAS) .....	20
1.3.10 Software .....	21
1.4 Service-Oriented Architecture (SOA) .....	21
1.5 Summary .....	24
<b>Chapter 2. System structure and design</b> .....	27
2.1 System structure .....	28
2.1.1 Book concept .....	28
2.1.2 Models .....	31
2.1.3 Memory .....	32
2.1.4 Ring topology .....	36
2.1.5 Connectivity .....	37
2.1.6 Frames and cages .....	39
2.1.7 The MCM .....	42
2.1.8 The PU, SC, SD, and MSC chips .....	42
2.1.9 Summary .....	44
2.2 System design .....	45
2.2.1 Design highlights .....	45
2.2.2 Book design .....	46
2.2.3 Processor Unit design .....	48
2.2.4 Processor Unit functions .....	54
2.2.5 Memory design .....	68
2.3 Model configurations .....	71
2.4 Logical partitioning .....	79
2.5 Storage operations .....	84
2.5.1 Reserved storage .....	86
2.5.2 Logical partition storage granularity .....	87
2.5.3 LPAR Dynamic Storage Reconfiguration (DSR) .....	87

<b>Chapter 3. I/O system structure</b> .....	89
3.1 Overview .....	90
3.2 I/O cages .....	91
3.2.1 Self-Timed Interconnect (STI) .....	93
3.2.2 STIs and I/O cage connections .....	94
3.2.3 Balancing I/O connections .....	96
3.3 I/O and cryptographic feature cards .....	97
3.3.1 I/O feature cards .....	97
3.3.2 Physical Channel IDs (PCHIDs) .....	98
3.4 Connectivity .....	101
3.4.1 I/O and cryptographic features support and configuration rules .....	101
3.4.2 ESCON channels .....	105
3.4.3 FICON channels .....	108
3.4.4 OSA-Express2 and OSA-Express adapters .....	115
3.4.5 Coupling links .....	124
3.4.6 External Time Reference feature (FC 6155) .....	126
3.4.7 Cryptographic features .....	127
<b>Chapter 4. Channel Subsystem</b> .....	129
4.1 Channel Subsystem (CSS) .....	130
4.1.1 Multiple CSSs concept .....	131
4.1.2 Multiple CSSs structure .....	131
4.1.3 Multiple Subchannel Sets (MSS) .....	133
4.1.4 CSS-related numbers .....	135
4.1.5 Physical Channel ID (PCHID) .....	137
4.1.6 Channel spanning .....	138
4.1.7 IOCP example .....	139
4.1.8 IODF Version 5 .....	142
4.1.9 Configuration management .....	143
4.1.10 System-initiated CHPID reconfiguration .....	144
4.1.11 Multipath Initial Program Load (IPL) .....	144
4.2 The MIDAW facility .....	144
<b>Chapter 5. Cryptography</b> .....	149
5.1 Cryptographic functions .....	150
5.1.1 Cryptographic synchronous functions .....	150
5.1.2 Cryptographic asynchronous functions .....	150
5.1.3 Cryptographic feature codes .....	153
5.2 CP Assist for Cryptographic Function (CPACF) .....	154
5.3 Crypto Express2 .....	154
5.3.1 Crypto Express2 coprocessor .....	155
5.3.2 Crypto Express2 accelerator .....	157
5.3.3 Configuration rules .....	157
5.4 TKE workstation feature .....	159
5.4.1 Optional TKE Feature .....	159
5.5 Cryptographic functions comparison .....	160
5.6 Software support .....	161
5.6.1 CPACF .....	161
5.6.2 Crypto Express2 .....	162
5.6.3 Web deliverables .....	162
5.6.4 z/OS ISCF FMIDs .....	162
<b>Chapter 6. Software support</b> .....	165
6.1 Operating systems summary .....	166

6.2 Support by operating system . . . . .	167
6.2.1 z/OS . . . . .	167
6.2.2 z/VM . . . . .	168
6.2.3 VSE/ESA and z/VSE. . . . .	170
6.2.4 Linux on System z. . . . .	171
6.2.5 TPF and z/TPF . . . . .	172
6.3 Support by function . . . . .	172
6.3.1 ICKDSF. . . . .	178
6.4 Software licensing considerations. . . . .	178
6.4.1 Workload License Charges. . . . .	178
6.4.2 Select Application License Charges (SALC). . . . .	179
6.5 Concurrent upgrade considerations . . . . .	180
6.6 References . . . . .	182
<b>Chapter 7. Sysplex functions . . . . .</b>	<b>183</b>
7.1 Parallel Sysplex. . . . .	184
7.2 Coupling Facility considerations . . . . .	185
7.2.1 Sysplex configurations and Time Synchronization . . . . .	185
7.2.2 Coupling Facility and CFCC considerations for z9 EC . . . . .	188
7.2.3 CFCC enhanced patch apply . . . . .	189
7.2.4 Coupling link connectivity . . . . .	190
7.2.5 ICF processor assignments . . . . .	192
7.2.6 Dynamic CF dispatching and Dynamic ICF expansion. . . . .	193
7.3 System-managed CF structure duplexing. . . . .	195
7.4 Intelligent Resource Director . . . . .	197
7.4.1 LPAR CPU management . . . . .	198
7.4.2 Dynamic Channel Path Management . . . . .	199
7.4.3 Channel Subsystem Priority Queuing . . . . .	201
7.4.4 WLM and Channel Subsystem priority . . . . .	202
7.4.5 Special considerations and restrictions. . . . .	203
<b>Chapter 8. Concurrent upgrades and availability . . . . .</b>	<b>205</b>
8.1 Availability enhancements. . . . .	206
8.2 Concurrent upgrades . . . . .	207
8.2.1 Capacity Upgrade on Demand (CUoD). . . . .	210
8.2.2 Customer Initiated Upgrade (CIU). . . . .	216
8.2.3 On/Off Capacity on Demand (On/Off CoD). . . . .	222
8.2.4 Capacity BackUp (CBU) . . . . .	228
8.3 Enhanced Book Availability (EBA) . . . . .	232
8.3.1 Planning consideration . . . . .	232
8.3.2 Enhanced Book Availability - Process. . . . .	234
8.4 Enhanced Driver Maintenance (EDM) . . . . .	242
8.5 Nondisruptive upgrades . . . . .	244
8.5.1 Planning for nondisruptive upgrades . . . . .	245
<b>Chapter 9. Environmental requirements . . . . .</b>	<b>249</b>
9.1 Power and cooling. . . . .	250
9.1.1 Power consumption . . . . .	250
9.1.2 Internal Battery Feature . . . . .	251
9.1.3 Emergency power-off . . . . .	251
9.1.4 Cooling requirements . . . . .	251
9.2 Weights . . . . .	252
9.3 Dimensions . . . . .	252
9.4 Frame tie down for raised floor and non-raised floor . . . . .	253

9.5 Restriction of Hazardous Substances . . . . .	253
<b>Appendix A. Hardware Management Console</b> . . . . .	255
HMC support for Server Time Protocol . . . . .	256
Remote operations . . . . .	257
HMC Console support. . . . .	259
HMC application . . . . .	260
<b>Appendix B. CHPID mapping tool.</b> . . . . .	263
CMT requirements . . . . .	264
CMT purpose and description . . . . .	264
z9 EC CHPID mapping . . . . .	265
<b>Appendix C. Fiber cabling services</b> . . . . .	273
Fiber cabling services options . . . . .	274
IBM Networking Services fiber cabling services . . . . .	274
Summary . . . . .	278
References. . . . .	279
<b>Related publications</b> . . . . .	295
IBM Redbooks . . . . .	295
Other publications . . . . .	295
Online resources . . . . .	296
How to get IBM Redbooks . . . . .	296
Help from IBM . . . . .	296
<b>Index</b> . . . . .	297



# Notices

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

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

*IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.*

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

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

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

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

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


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

## COPYRIGHT LICENSE:

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

## Trademarks

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

Chipkill™	Sysplex™	Sysplex Timer®
CICS®	GDPS®	System z™
CUA®	HiperSockets™	System z9™
DB2 Connect™	IBM®	System/360™
DB2®	IMS™	System/370™
developerWorks®	Parallel Sysplex®	Tivoli®
DFSMS™	Processor Resource/Systems Manager™	TotalStorage®
DRDA®	PR/SM™	VM/ESA®
DS8000™	Redbooks®	VSE/ESA™
e-business on demand®	Redbooks (logo)  ®	WebSphere®
Enterprise Systems Architecture/390®	Resource Link™	z/Architecture®
ECKD™	RACF®	z/OS®
ES/9000®	RETAIN®	z/VM®
ESCON®	RMF™	z/VSE™
FlashCopy®	S/360™	zSeries®
FICON®	S/370™	z9™
Geographically Dispersed Parallel	S/390®	

The following terms are trademarks of other companies:

SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

Java, JRE, JVM, J2EE, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

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

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

# Preface

This IBM® Redbooks® publication discusses the IBM System z9™ Enterprise Class (z9 EC), which offers a continuation of the IBM scalable mainframe servers. Based on z/Architecture®, the System z9 Enterprise Class server provides major extensions by:

- ▶ Increasing the maximum number of Processor Units and logical partitions
- ▶ Supporting larger configurations with the concept of Channel Subsystems and Multiple Subchannels Sets
- ▶ Providing a base for major server consolidation by further removing memory, processor, and channel constraints

In addition to increased performance and expansion options, improved facilities for nondisruptive maintenance and growth provide better operational support and availability.

This book provides an overview of the z9 EC and its functions, features, and associated software support. More details are offered in selected areas relevant to technical planning.

This book is intended for systems engineers, consultants, planners, and anyone wanting to understand the new IBM System z9 Enterprise Class functions and plan for their usage. It is not intended as an introduction to mainframes. Readers are expected to be generally familiar with existing System z™ technology and terminology.

This publication is part of a series. For a more complete understanding of System z9 capabilities, also refer to our companion Redbooks publications:

- ▶ *IBM System z9 Business Class Technical Introduction*, SG24-7241
- ▶ *IBM System z Connectivity Handbook*, SG24-5444

## The team that wrote this book

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

**Bill White** is a Project Leader and Senior Networking Specialist at the International Technical Support Organization, Poughkeepsie Center.

**Franck Inje** is a Project Leader and I/T Architect at the International Technical Support Organization, Poughkeepsie Center.

**Greg Chambers** is a Certified Field Technical Sales Specialist in the Western Region. During his 29 years at IBM, he has worked in various roles in service, design, and support of IBM large systems. Currently, he provides pre-sales and post-sales technical support for System z products

**Marian Gasparovic** is an IT Specialist from IBM Slovakia. He worked as an administrator for z/OS® as an IBM Business Partner for six years. He joined IBM in 2004 and now works as a Field Technical Sales Support for System z in the CEMA region as a member of a new workload team.

**Parwez Hamid** is a Consulting IT Specialist working for IBM Server and Technology Group in the UK. During the past 32 years, he worked in various roles within IBM and with a large number of IBM mainframe customers. He also worked on projects introducing new technology. Currently, he provides pre-sales and post-sales technical support for the IBM System z product portfolio. In addition, since 1995, Parwez worked in IBM Poughkeepsie, writing Redbooks and preparing technical material for the world-wide announcement and education for System z servers. He is a regular speaker at various System z conferences and events.

**Brian Hatfield** is a Certified Consulting Learning Specialist working for the IBM Systems and Technology Group in Atlanta, Georgia. He has over 28 years of experience in the IBM mainframe environment, starting his career as a Large System Customer Engineer in Southern California. He has been in education for the past 16 years and currently develops and delivers technical training for the System z environment.

**Ken Hewitt** is an IT Specialist in Australia. He has 18 years of experience in IBM Large Systems. He has worked in various roles within IBM and currently provides sales technical support for customers across Australia.

**Dick Jorna** is a Certified Consulting IT Specialist working for IBM Server and Technology Group in the Netherlands. During the past 38 years, he has worked in various roles within IBM and with a large number of mainframe customers. He currently provides pre-sales System z technical consultancy in support of large and small customers. In addition, he acts as a System z Product Manager in the Netherlands and is responsible for all System z related activities.

**Patrick Kappeler** is a System z Security Specialist working in France. He joined IBM in 1970 as a Diagnostic Program Designer and has held several mainframe technical support specialist and management positions, as well as international assignments, since. He joined the EMEA System z New Technology Center in Montpellier in 1996, where he now provides consulting and pre-sales technical support in the area of e-business security.

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

Jason Boxer  
IBM System z Technical Support Lead, Poughkeepsie

Catherine Cronin  
IBM Systems and Technology Group, Performance Analysis, Poughkeepsie

Jeffrey Berger  
IBM Software Group, Information Management DB2® Performance, San Jose

## Become a published author

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

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

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

[ibm.com/redbooks/residencies.html](http://ibm.com/redbooks/residencies.html)

## Comments welcome

Your comments are important to us!

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

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

[ibm.com/redbooks](http://ibm.com/redbooks)

- ▶ Send your comments in an email to:

[redbook@us.ibm.com](mailto:redbook@us.ibm.com)

- ▶ Mail your comments to:

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

Archived

## Overview

For over four decades, the IBM mainframe has been a leader in data and transaction serving. The announcement of the System z9 in July of 2005 provided a strong combination of past mainframe characteristics plus new functions designed around scalability, availability, and security. IBM further enhances the capabilities of the System z9 by optimized capacity settings with subcapacity central processors (CPs), the availability of the new System z9 Integrated Information Processor (zIIP), and improvements for FICON® performance and throughput. With the availability of new capacity settings, the System z9 has a comprehensive server range to meet the needs of businesses spanning mid-range companies to large enterprises.

The IBM System z9 Enterprise Class (z9 EC, formerly z9-109) continues the evolution of the mainframe, building upon the z/Architecture definitions. The z9 EC extends and integrates key platform characteristics: Dynamic and flexible partitioning, resource management in mixed and unpredictable workload environments, availability, scalability, clustering, and security and systems management with emerging e-business on demand® application technologies, for example, WebSphere®, Java™, and Linux®. All these technologies and improvements come into play when the System z9 becomes the heart of the enterprise SOA solutions.

Most topics mentioned in this chapter are discussed in greater detail later in this book. Here we introduce components of the total systems design and then focus in subsequent chapters on specific features and functions that are relevant to technical planning.

# 1.1 Introduction

The System z9 provides a significant increase in system scalability over the previous System z servers. With the increased performance and the total system capacity possible, an opportunity exists to continue to consolidate diverse applications on a single platform. New innovations help to ensure it is a security-rich platform that can help to maximize the resources and their utilization, and can help provide the ability to integrate applications and data across the infrastructure.

The z9 EC is built on more than 40 years of industry leadership and continues to take that leadership to new levels. The server is built using a modular multi-book design that supports one to four books per server. The book contains a multi-chip module (MCM), which hosts the processor units, memory, and high speed connectors for I/O. This approach enables many of the high-availability, nondisruptive capabilities that differentiate it from other servers.

The z9 EC has five model offerings, from one to 54 configurable processor units (PUs). The first four models (S08, S18, S28, and S38) have 12 PUs per book, and the high capacity model (the S54) has 16 PUs in each of its four books. Using a modular book design, the z9 EC Model S54 is designed to provide up to 95 percent more total system capacity than the z990 Model D32 and has up to double the available memory. The comparison of the z9 EC Model S54 and the z990 Model D32 is based on the LSPR mixed workload average.

The chart in Figure 1-1 shows continued growth improvements along all axes. While some of the previous generation of servers have grown more along one axis for a given family, later families focus on the other axes. The balanced design of z9 EC achieves improvement equally along all four axes.

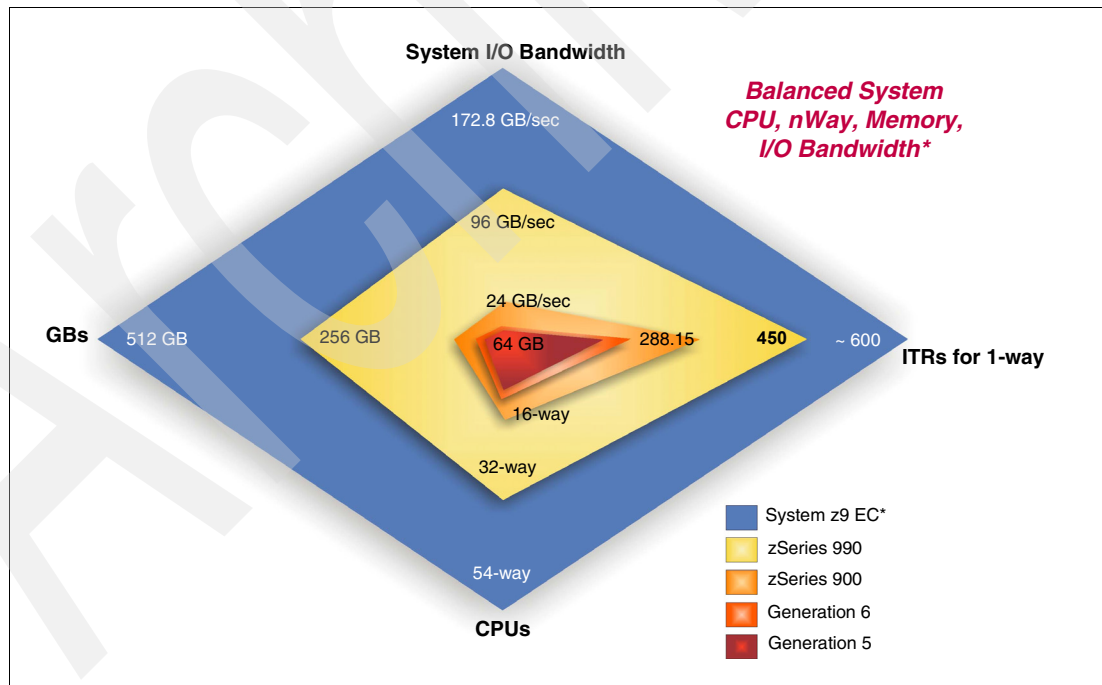


Figure 1-1 z9 EC balanced system design

With the System z9, the Parallel Sysplex® cluster takes the commercial strengths of the z/OS platform to improved levels of system management, competitive price/performance, scalable growth, and continuous availability.



The z9 EC provides a significant increase in system scalability and opportunity for server consolidation by providing a *multi-book* system structure. All books are interconnected with a very high-speed internal communications links through the L2 cache, which allows the system to be operated and controlled by the PR/SM™ facility as a symmetrical, memory-coherent multiprocessor.

The PU configuration is made up of two System Assist Processors (SAPs) per book and two spare PUs per server. The remaining PUs can be characterized as Central Processors (CPs), Integrated Facility for Linux (IFL) processors, System z Application Assist Processors (zAAPs), System z9 Integrated Information Processors (zIIPs), Internal Coupling Facility (ICFs) processors, or additional SAPs.

Each book supports up to 128 GB of memory, for a server maximum of 512 GB, and 16 high-performance Self-Timed Interconnects for data communications between memory and I/Os. The I/O infrastructure has been redesigned to increase the number of cards that can be installed in an I/O cage. The multiple Channel Subsystems architecture on the z9 EC allows up to four CSSs, each with 256 channels. I/O constraint relief using Multiple Subchannel Sets (MSS) allows access to a greater number of logical volumes.

PR/SM manages all the installed and enabled resources (processors and memory) of the installed books as a single large SMP. It provides the ability to configure and operate as many as 60 logical partitions, which have processors, memory, and I/O resources assigned from any of the installed books.

Figure 1-2 shows an external view of the z9 EC.



Figure 1-2 z9 EC - External view

The z9 EC is focused on providing higher availability and reducing planned and unplanned outages, which, when properly configured, may be accomplished with improved nondisruptive replace, repair, and upgrade functions for memory, books, and I/O, as well as extending nondisruptive capability to download Licensed Internal Code updates.

Integrated clear key encryption security-rich features include support for Advanced Encryption Standard, Secure Hash Algorithm-256, and integrated Pseudo Random Number Generation. Performing these security functions in hardware contributes to improved performance.

## 1.2 z9 EC models

The z9 EC has a machine type of 2094. There are five models. The last two digits of each model indicates the maximum number of PUs available for purchase: Models S08, S18, S28, S38, and S54.

A PU is the generic term for the z/Architecture processor on the Multichip Module (MCM) that can be characterized as:

- ▶ A Central Processor (CP).
- ▶ An Internal Coupling Facility (ICF) to be used by the Control Facility Control code (CFCC).
- ▶ An Integrated Facility for Linux (IFL).
- ▶ An additional System Assist Processor (SAP®) to be used by the Channel Subsystem.
- ▶ A System z Application Assist Processor (zAAP). One CP must be installed with or prior to any zAAPs being installed.
- ▶ A System z9 Integrated Information Processor (zIIP). One CP must be installed with or prior to any zIIPs being installed.

In the z9 EC five-model structure, only one CP, ICF, or IFL must be purchased and activated for any model. PUs can be purchased in single PU increments and are orderable by feature code. The total number of PUs purchased may not exceed the total number available for that model.

The development of a multi-book system provides an opportunity to concurrently increase the capacity of the system in three areas:

- ▶ Add capacity by concurrently activating more CPs, IFLs, ICFs, zAAPs, or zIIPs on an existing book.
- ▶ Add a new book concurrently and activate more CPs, IFLs, ICFs, zAAPs, or zIIPs.
- ▶ Add a new book to provide additional memory or STIs to support increasing storage or I/O requirements. I/O features or channel types supported on the z9 EC are:
  - FICON Express (upgrades from z990), FICON Express2, and FICON Express4
  - Coupling Links - Peer mode only
  - OSA-Express (upgrades from z990) and OSA Express2
  - ESCON®
  - Crypto Express2

## Model upgrade paths

With the exception of the z900 Model 100, any z900, z990, and z9 Business Class model S07 may be upgraded to a z9 EC, as shown in Figure 1-3.

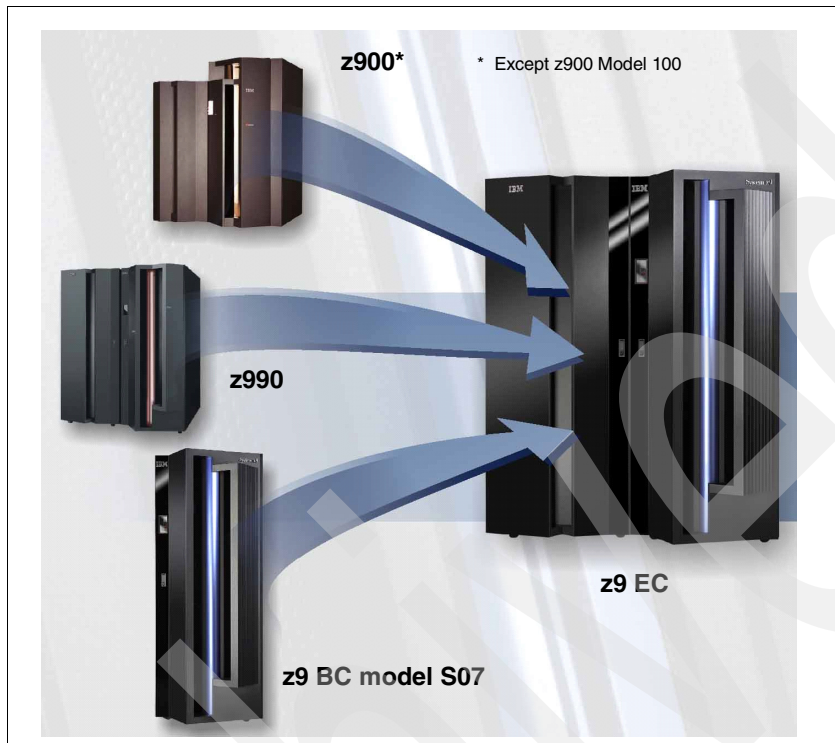


Figure 1-3 Upgrades

## Concurrent Processor Unit conversions

The z9 EC supports concurrent conversion between different PU types, providing flexibility to meet changing business environments. CPs, IFLs, zAAPs, zIIPs, or ICFs may be converted to CPs, IFLs, zAAPs, zIIPs, or ICFs. Unassigned IFLs can be converted to IFLs only.

## 1.3 System functions and features

The z9 EC is a 2-frame server; the frames contain the key components. Figure 1-4 shows an internal view of the server, which is made up of the following:

- ▶ The CEC cage with up to four books
- ▶ From one up to three I/O cages
- ▶ Power supplies
- ▶ An optional internal battery feature
- ▶ Modular cooling units
- ▶ Support elements

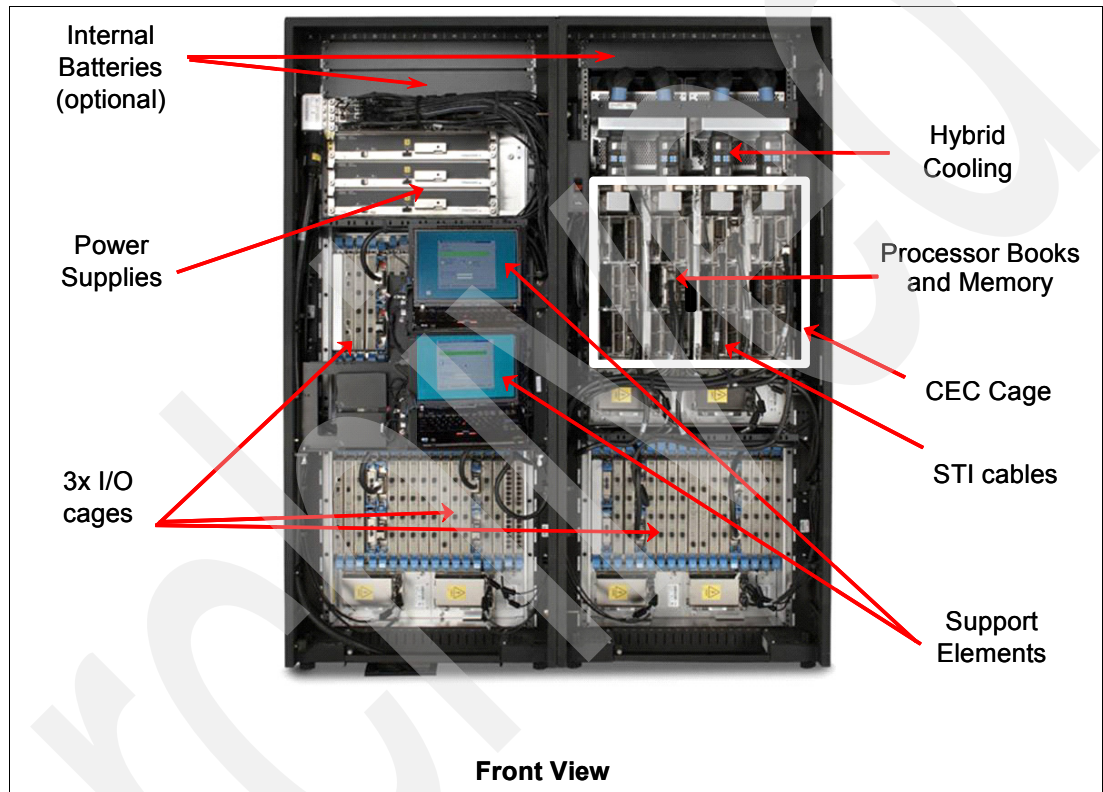


Figure 1-4 z9 EC internal view

The z9 EC is designed to provide:

- ▶ Uniprocessor performance improvement, which is expected to be up to 35 percent more than the z990, based on LSPR mixed workload average.
- ▶ Nearly double the total system capacity of z990, with up to 54 Processor Units compared to a maximum of 32 PUs on z990.
- ▶ Double the available memory per model, with up to 128 GB per book.
- ▶ Increased host bus bandwidth between memory and I/O.
- ▶ Up to 16 Self-Timed Interconnects (STIs) per book, with 2.7 GBps each for I/O, a 35 percent increase in STI speed compared to z990.

- ▶ Reduction in the impact of planned and unplanned server outages:
  - Enhanced Book Availability
  - Redundant I/O Interconnect
  - Enhanced Driver Maintenance
  - Dynamic oscillator switch-over
  - Concurrent MBA fanout card hot-plug
- ▶ Multiple Subchannel Sets (MSS), which are designed to allow improved device connectivity for Parallel Access Volumes (PAVs).
- ▶ Improved performance over native FICON channels for programs that process data sets exploiting striping and compression (such as DB2, VSAM, PDSE, HFS, and zFS) by reducing channel, director, and control unit overhead when using the Modified Indirect Data Address Word (MIDAW) facility.
- ▶ Increased connectivity for Cryptography, FICON, and OSA in one I/O cage.
- ▶ Increased number of open exchanges that may be active simultaneously for FICON Express2 and Express4, from 32 to 64 concurrent I/O operations per channel.
- ▶ Up to 336 FICON Express2 or FICON Express4 channels, a 40 percent increase compared to z990.
- ▶ Availability enhancements for FICON.
- ▶ Additional LAN connectivity with the OSA-Express2 1000BASE-T Ethernet.
- ▶ OSA-Express2 OSN (OSA for NCP) is designed to provide Channel Data Link Control (CDLC) protocol support on the z9 EC for the IBM Communication Controller for Linux on System z (CCL), which allows system administrators to configure, manage, and operate their CCL Network Control Programs (NCPs) in the same manner as their ESCON-attached 374x NCPs.

### 1.3.1 The CEC cage

All z9 EC parts come in a CEC cage that consists of two frames.

#### MCM technology

The System z9 is built on the superscalar microprocessor architecture. The z9 EC 12-PU and 16-PU MCM has 16 chips. The total number of transistors is approximately 4.5 billion, compared with approximately 3 billion for the z990. With this technology integration comes improvements in chip-to-substrate and substrate-to-board connections.

The z9 EC MCM has 102 layers in the glass ceramic substrate and is thinner, shortening the paths that signals have to travel to reach their destination (another chip or exiting the MCM). Inside the low dielectric glass ceramic substrate is 0.545 km of internal wiring that interconnects the 16 chips that are mounted on the top layer of the MCM. The internal wiring provides power and signal paths into and out of the MCM.

The number of CPs and SAPs assigned for each particular model depends on the configuration. Unlike the z990, the z9 EC system has only two spare PUs.

With the z9 EC, this design and the exploitation of the CMOS 10SK-SOI technology improves the uniprocessor performance by up to 35 percent compared to z990.

However, the true capacity increase of the system is driven by the increased number of Processor Units per system: From a maximum of 48 in the z990 to a maximum of 64 in the z9 EC.

## Memory

The z9 EC continues to employ storage size selection by Licensed Internal Code. Memory cards installed may have more usable memory than required to fulfill the machine order. Licensed Internal Code Configuration Control (LIC-CC) will determine how much memory is used from each card.

## MBA fanout card hot-plug

A Memory Bus Adapter (MBA) fanout card is designed to provide the path for data between memory and the I/O cards using Self-Timed Interconnect (STI) cables. The MBA fanout card is hot-pluggable in the z9 EC. Up to eight MBA fanout cards are available per book.

In the event of an outage, an MBA fanout card may be concurrently repaired without loss of access to its associated I/O cards using Redundant I/O Interconnect.

## Self-Timed Interconnect granularity

The MBA fanout card is designed to support Self-Timed Interconnect (STI) granularity: two STIs per MBA fanout card (up to eight MBA fanout cards per book) and up to 64 STIs on the z9 EC when four books are installed. An STI is an interface to the Memory Bus Adaptor (MBA), used to gather and send data.

Each STI has a bidirectional bandwidth of 2.7 GBps for I/O and 2.0 GBps for ICB-4s and STI-3 extender cards. When populated with 16 STIs, each book has a maximum bandwidth of 43.2 GBps.

## PR/SM

PR/SM enables logical partitioning of the z9 EC server.

### *Up to 60 logical partitions*

With the z9 EC, it is possible to define up to 60 logical partitions, which may provide even more flexibility to allocate hardware resources. With PR/SM and Multiple Image Facility, the installation can share:

- ▶ ESCON ports across logical partitions within a single Channel Subsystem
- ▶ FICON channels, ISC-3s, and OSA ports across logical partitions and across multiple Channel Subsystems.

Support of up to 60 logical partitions is exclusive to the z9 EC and is supported by z/OS, z/VM®, z/VSE™, z/TPF, and Linux on System z.

### *Separate PU management*

Separate PU management is a flexible way for managing Processor Units (PUs). PUs defined as Internal Coupling Facility (ICF) processors, Integrated Facility for Linux (IFL) processors, System z Application Assist Processors (zAAPs), or System z9 Integrated Information Processors (zIIPs) are managed separately.

The separate management of PU types simplifies capacity planning and management of the configured logical partitions and their associated processor resources.

The ability to specify a separate logical partition weight for shared zAAPs or zIIPs helps to simplify capacity planning and management of the configured logical partitions and their associated processor resources.

## Channel Subsystem (CSS)

The z9 EC CSS supports multiple Channel Subsystems on a single server. It provides:

- ▶ Four Channel Subsystems, each CSS having from one to 256 channels.
- ▶ Multiple Subchannel Sets.
- ▶ Each CSS can be configured with one to 15 logical partitions.

The I/O subsystem continues to be viewed as a single entity, through an Input/Output Configuration Data Set (IOCDS) across the entire system. Only one Hardware System Area (HSA) is used for the multiple CSSs.

**Note:** There is no change to the operating system, which continues to support a maximum of 256 channels.

## Multiple Subchannel Sets (MSS)

System z9 Multiple Subchannel Sets are designed to provide an increased number of subchannels. Two subchannel sets are available per CSS, enabling a total of 63.75 K subchannels in set 0 and the addition of 64 K-1 subchannels in set 1.

With the multiple subchannel set facility, one or two sets of subchannels may be configured to each CSS. z/OS V1R7 will allow only Parallel Access Volume Alias (PAV-alias) devices in the subchannel set 1. MSS is designed to provide greater I/O device configuration capabilities for large installations.

Multiple Subchannel Sets are exclusive to the z9 EC. MSS is supported by ESCON (CHPID type CNC) and FICON features (CHPID type FC) on z/OS V1R7 on System z9.

## Physical Channel IDs (PCHIDs)

A z9 EC can have up to 1024 physical channels, or PCHIDs. In order for an operating system to make use of that PCHID, it must be mapped to a CHPID within the IOCDS. Each CHPID is uniquely defined with a CSS and mapped to an installed PCHID. A PCHID is eligible for mapping to any CHPID in any CSS.

The z9 EC CHPID Mapping Tool (CMT) provides a method of customizing the CHPID assignments for a z9 EC system to avoid attaching critical channel paths to single points of failure. It should be used after the machine order is placed and before the system is delivered for installation. The PCHID assignments are fixed and cannot be changed.

There are no default CHPID assignments; CHPIDs are assigned when the IOCDS file is built.

## Spanned channels

The z9 EC Channel Subsystem is extended to provide sharing of some channel types in a manner that extends the shared channel function. Internal Channel types, such as HiperSocket (IQD) and Internal Coupling Channels (ICP), can be configured as *spanned* channels. External channels such as FICON channels, OSA features, and external coupling links can be defined as spanned channels. Spanned channels allow a channel to be configured to multiple CSSs, thus enabling them to be shared by any or all of the configured logical partitions.

## Performance assists for z/VM Linux guests

There is an important virtualization technology in System z9 designed to improve the performance of z/VM guest operating systems (such as Linux on System z) when Queued Direct Input/Output (QDIO) is used for HiperSockets™, FCP, and OSA.

This virtualization technology is designed to allow QDIO interruptions to be passed directly to guests for HiperSockets, Fibre Channel Protocol (FCP), and OSA channels.

The hardware assists allow a cooperating guest operating system to initiate QDIO operations directly to the applicable channel, without interception by z/VM, thereby helping to provide additional performance improvements. The performance assists are provided on the System z9 for:

- ▶ HiperSockets, CHPID type IQD
- ▶ All FICON features with CHPID type FCP
- ▶ All OSA features with CHPID type OSD

### 1.3.2 I/O connectivity

z9 EC provides many connectivity options.

#### I/O cage

The z9 EC has a minimum of one CEC cage and one I/O cage in the A frame. The Z frame can accommodate additional two I/O cages, making a total of three for the whole system. Figure 1-4 on page 6 shows the layout of the frames and I/O cages.

One I/O cage can accommodate the following card types:

- ▶ Up to eight Crypto Express2
- ▶ Up to 28 FICON Express4, FICON Express2, or FICON Express
- ▶ Up to 24 OSA-Express2 or OSA-Express
- ▶ Up to 28 ESCON

It is possible to populate the 28 I/O slots in one I/O cage with any mix of the of the above-mentioned cards.

#### FICON Express4, FICON Express2, and FICON Express channels

Up to 336 FICON Express4 or FICON Express2 channels and up to 120 FICON Express channels are supported on a z9 EC. The FICON Express4 features support a link data rate of 1, 2, or 4 Gbps auto negotiated. The FICON Express2 features support a link data rate of 1 or 2 Gbps auto negotiated.

The FICON Express4 and FICON Express2 features support:

- ▶ Native FICON and FICON Channel-to-Channel (CTC) traffic supporting connectivity to FICON devices in the z/OS, z/VM, z/VSE, z/TPF, and Linux on System z environments
- ▶ Fibre Channel Protocol traffic, CHPID type FCP, supporting connectivity to disks and tapes through Fibre Channel switches and directors in the z/VM, z/VSE (ESS disks only), and Linux on System z environments

The same FICON Express4, FICON Express2, and FICON Express channel cards used for FICON channels can also be used for FCP channels. FCP channels are enabled on these cards as a microcode load with an FCP mode of operation and CHPID type definition. FCP mode is available in long wavelength (LX) and short wavelength (SX) operation.

The z9 EC supports FCP channels, switches, and FCP/SCSI devices with full fabric connectivity under Linux on System z.



### ***FICON CTC function***

FICON CTC connectivity increases the bandwidth between z9 EC, z9 BC, z990, and z890. As the FICON CTC function is included as part of the native FICON (FC) mode of operation on System z, a FICON channel used for FICON CTC is not limited to intersystem connectivity; it also supports multiple device definitions. Native FC mode can support both device and CTC mode definition concurrently, allowing for greater connectivity flexibility.

### ***FICON Cascaded Directors***

With the FICON Cascaded Director function, a native FICON (FC) channel or a FICON CTC can connect a server to a device or other server through two (same vendor) FICON Directors in between. Cascaded support is important for disaster recovery and business continuity solutions. It can provide high availability and extended distance connectivity, and has the potential for fiber infrastructure cost savings by reducing the number of channels for interconnecting the two sites.

### ***FCP point-to-point attachments***

When a FICON feature is configured as CHPID type FCP, the direct attachment of devices (point-to-point connection) is supported without the need for an intermediate Fibre Channel switch or director. N\_Port ID virtualization (NPIV) is not supported with FCP point-to-point attachments.

Point-to-point connections may be used to access data stored on these devices, and also to IPL an operating system or other stand-alone program from such a device, using the SCSI IPL feature. The no-charge SCSI IPL feature (FC 9904) is required to use the SCSI IPL function.

FCP point-to-point attachments are supported on the z9 EC, z9 BC, z990, and z890, by the FICON Express 4, FICON Express2, and FICON Express features with CHPID type FCP. It is used by z/VM and Linux on System z. z/VM provides FCP point-to-point to its guests; z/VM V5R3 and later can use it also for system usage.

### **ESCON channels**

The high density ESCON feature (FC 2323) has 16 ports, of which 15 can be activated for customer use. One port is always reserved as a spare, in the event of a failure of one of the other ports.

ESCON channels are available on a port basis in increments of four. Each port utilizes a light emitting diode (LED) as the optical transceiver, and supports use of a 62.5/125-micrometer multimode fiber optic cable terminated with a small form factor, industry standard MT-RJ connector.

### **Modified Indirect Data Address Word facility**

The System z9 I/O subsystem supports a facility for indirect addressing, the Modified Indirect Data Address Word (MIDAW) facility, for both ESCON and FICON channels. The use of the MIDAW facility, by applications that currently use data chaining, may result in improved channel throughput in FICON environments.

The MIDAW facility is exclusive to the z9 EC, and is supported by ESCON using CHPID type CNC and by FICON using CHPID types FCV and FC. The MIDAW facility is exploited by z/OS.

### **Open Systems Adapter (OSA)**

The z9 EC can have up to 24 features of the Open Systems Adapter family of local area network (LAN) adapters, giving a maximum of 48 ports of LAN connectivity.

It is possible to choose any combination of the supported OSA Express2 or OSA Express Ethernet features on the z9 EC. The OSA-Express Token Ring is not supported.

### **OSA-Express2 OSN - Open Systems Adapter for NCP**

The OSA-Express2 Gigabit Ethernet and 1000BASE-T Ethernet features (FC 3364, 3365, and 3366) have the capability to provide channel connectivity from System z operating systems to IBM Communication Controller for Linux on System z (CCL) using the Open Systems Adapter for the Network Control Program (OSA for NCP) supporting the Channel Data Link Control (CDLC) protocol.

If SNA solutions that require NCP functions are required, CCL can be considered as a migration strategy to replace IBM Communications Controllers (374x). The CDLC connectivity option enables TPF environments to exploit CCL.

OSA-Express2 OSN support is exclusive for System z9, on the OSA-Express2 Gigabit Ethernet SX, Gigabit Ethernet LX, and 1000BASE-T Ethernet features, and requires the port to be configured as CHPID type OSN, which can be configured on a port-by-port basis.

### **HiperSockets - IPv6**

Internet Protocol Version 6 (IPv6) support is being offered for HiperSockets (CHPID type IQD). IPv6 is the protocol designed by the Internet Engineering Task Force (IETF) to replace Internet Protocol Version 4 (IPv4) to help satisfy the demand for additional IP addresses.

IPv6 support is currently available on the OSA-Express2 and OSA-Express features in the z/OS, z/VM, and Linux on System z environments. The support of IPv6 on HiperSockets is exclusive to System z9, and is supported by z/OS and z/VM.

## **1.3.3 Cryptography**

The z9 EC continues to build on the existing cryptographic capabilities.

### **CP Assist for Cryptographic Function (CPACF)**

The z9 EC continues to use the Cryptographic Assist Architecture first implemented on z990. CPACF performance is designed to scale with PU performance improvements.

CPACF offers the following on every Processor Unit (PU) characterized as a Central Processor (CP) or Integrated Facility for Linux (IFL):

- ▶ Data Encryption Standard (DES).
- ▶ Triple Data Encryption Standard (TDES).
- ▶ Secure Hash Algorithm (SHA-1).
- ▶ Advanced Encryption Standard (AES) for 128-bit keys.
- ▶ Pseudo Random Number Generation (PRNG). Note that PRNG is also a standard function supported on the Crypto Express2 feature.
- ▶ SHA-256.

SHA-1 and SHA-256 are shipped enabled on all servers and do not require the CPACF enablement feature. The CPACF functions are supported by z/OS, z/VM, and Linux on System z.

## Remote loading of ATM keys and key exchange with non-CCA systems

Remote Key Loading refers to the process of loading Data Encryption Standard (DES) keys to Automated Teller Machines (ATMs) from a central administrative site without the need for personnel to visit each machine to manually load DES keys.

Remote Keyload for ATM is based on the following standards:

- ▶ ISO/IEC 11770-3: Information Technology, Security Techniques, Key Management, Part 3: Mechanisms Using Asymmetric Techniques.
- ▶ ANS X9.24-2 (Draft): Retail Financial Services, Symmetric Key Management, Part 2: Using Asymmetric Techniques for the Distribution of Symmetric Keys.

Remote Key Loading Benefits:

- ▶ Provides a mechanism to load initial ATM keys without the need to send technical staff to ATMs.
- ▶ Reduces downtime due to key entry errors.
- ▶ Reduces service call and key management costs.
- ▶ Improves the ability to manage ATM conversions.
- ▶ Improved key Exchange with Non-CCA Cryptographic systems.

## IBM Common Cryptographic Architecture (CCA)

The features in CCA provide the ability to exchange keys between CCA systems, and systems that do not use Control Vectors, by allowing the CCA system owner to define the permitted types of key import and export while preventing uncontrolled key exchange that can open the system to an increased threat of attack.

## ICSF Callable Services

Integrated Cryptographic Service Facility (ICSF), together with Crypto Express 2, support the basic mechanisms in Remote Key Loading: The implementation offers a security-rich bridge between the CCA environment and the various formats and encryption schemes offered by the ATM vendors.

## Configurable Crypto Express2

The Crypto Express2 feature has two PCI-X adapters. On the z9 EC, each of the PCI-X adapters can be configured as either a Coprocessor or an Accelerator.

- ▶ Crypto Express2 Coprocessor - For secure key encrypted transactions (default)
  - Designed to support security-rich cryptographic functions, use of secure encrypted key values, and User Defined Extensions (UDX)
  - Designed for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification
- ▶ Crypto Express2 Accelerator - For Secure Sockets Layer (SSL) acceleration
  - Designed to support clear key RSA operations
  - Offloads compute-intensive RSA public-key and private-key cryptographic operations employed in the SSL protocol

Since the features are implemented in Licensed Internal Code, current Crypto Express2 features carried forward from z990 may take advantage of the increased SSL performance and new configuration options on z9 EC.

The configurable Crypto Express2 feature is supported by z/OS, z/VM, z/VSE, and Linux on System z.

### **TKE 5.0 workstation**

The Trusted Key Entry (TKE) workstation (FC 0839) and the TKE 5.1 level of Licensed Internal Code (FC 0856) are optional features. The TKE workstation offers security-rich local and remote key management, providing authorized persons a method of operational and master key entry, identification, exchange, separation, and update. The TKE workstation supports connectivity to an Ethernet Local Area Network operating at 10, 100, or 1000 Mbps.

An optional Smart Card Reader attached to the TKE 5.0 workstation allows for the use of smart cards that contain an embedded microprocessor and associated memory for data storage. Access to and the use of confidential data on the smart cards is protected by a user-defined Personal Identification Number (PIN).

## **1.3.4 Performance**

The performance design of the z/Architecture enables the entire server to support a new standard of performance for applications by expanding on a balanced system approach.

As CMOS technology has been enhanced to support not only additional processing power, but also more PUs, the entire server is modified to support the increase in processing power. The I/O subsystem supports a greater amount of bandwidth through internal changes, providing for larger and quicker data movement into and out of the server. Support of larger amounts of data within the server required improved management of storage configurations made available through integration of the operating system and hardware support of 64-bit addressing.

The combined balanced system design allows for increases in performance across a broad spectrum of work. However, due to the increased flexibility in the z9 EC model structure and resource management in the system, it is expected that there will be larger performance variability than has been previously seen by our traditional customer set. This variability may be observed in several ways.

The Large System Performance Reference (LSPR) should be referenced when considering performance on the z9 EC. The range of performance ratings across the individual LSPR workloads is likely to have a large spread. There will also be more performance variation of individual logical partitions as the impact of fluctuating resource requirements of other partitions can be more pronounced with the increased number of partitions and additional PUs available.

The impact of this increased variability is expected to be seen as increased deviations of workloads from single-number-metric based factors, such as MIPS, MSUs, and CPU time charge back algorithms. It is important to realize the z9 EC has been optimized to run many workloads at high utilization rates.

With a modular book design, the z9 EC is designed to provide up to 95 percent more total system capacity than the z990 Model D32, and has up to double the available memory. The performance of the z9 EC (2094) 701 is 1.35 times the z990 (2084) 301 (LSPR mixed workload).<sup>1</sup>

---

<sup>1</sup> This is a comparison of the z9 EC 54-way and the z990 D32 and is based on the LSPR mixed workload average.

The LSPR contains the Internal Throughput Rate Ratios (ITRRs) for the z9 EC and the previous generation processor families based upon measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user may experience will vary depending upon such considerations as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated.

For more detailed performance information, consult the Large Systems Performance Reference (LSPR) available at:

<http://www.ibm.com/servers/eserver/zseries/lspr/>

The MSU ratings are available on the Web:

<http://www.ibm.com/servers/eserver/zseries/library/swpriceinfo>

It is important to notice that the LSPR workloads have been updated to reflect more closely current and growth workloads. The traditional Commercial Batch Short Job Steps (CB-S) workload (formerly CB84) is dropped and a new Java batch (CB-J) workload is added. The remainder of the LSPR workloads are the same as the ones used for the z990 LSPR.

The new LSPR provides two tables:

- ▶ The single image z/OS from 1-way to 32-way.
- ▶ The typical logical partition configuration from 1-way to 54-way, based on customer profiles. This logical partition configuration table is used to establish single-number metrics.

Figure 1-5 shows an overview of the performance comparison between z9 EC and z990.

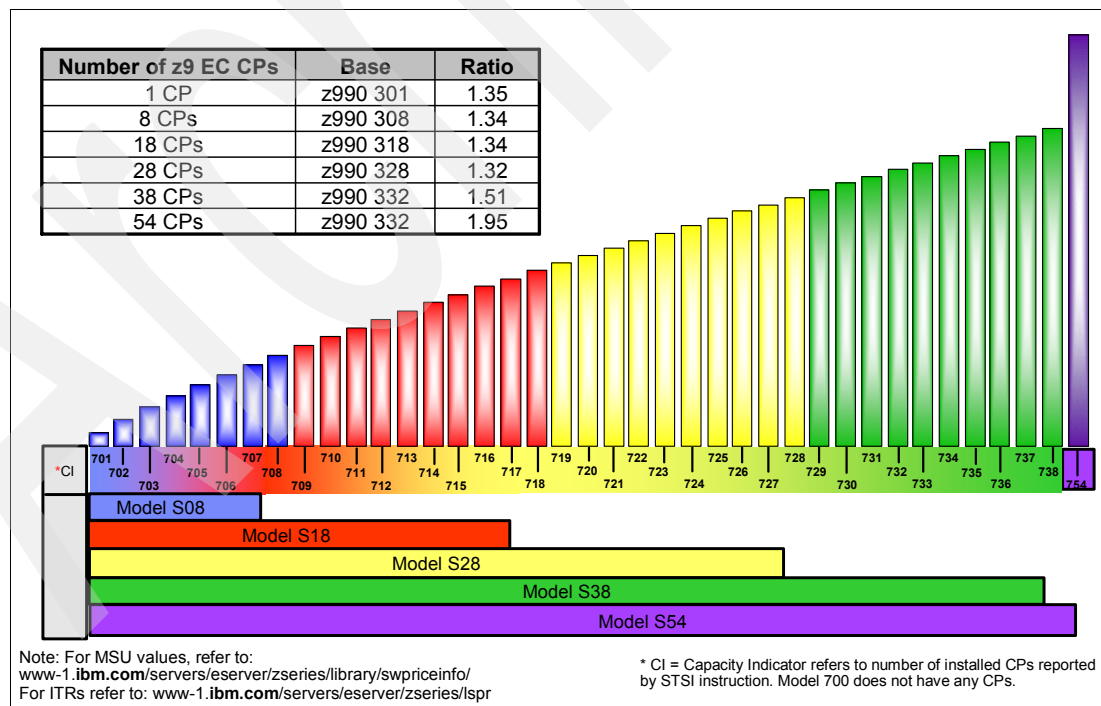


Figure 1-5 z9 EC to z990 performance comparison

Figure 1-6 shows the z9 EC granular capacity for up to eight CPs.

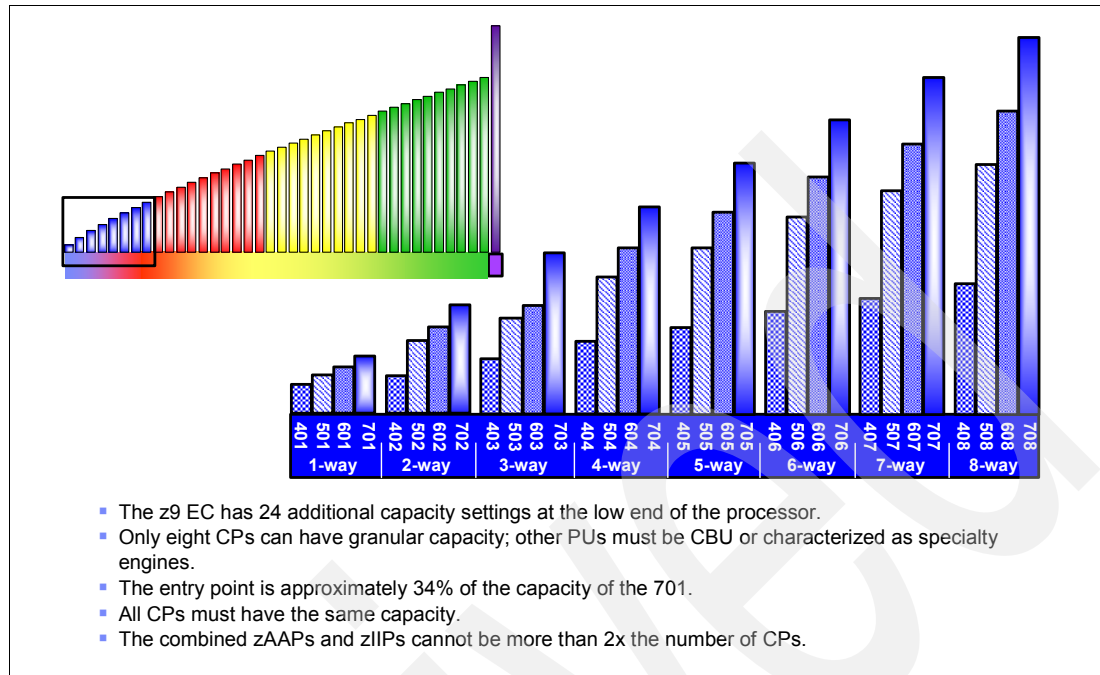


Figure 1-6 z9 EC granular capacity for up to eight CPs

The z9 EC LSPR rates all z/Architecture processors running in LPAR mode and 64-bit mode.

The actual throughput that any user will experience will vary, depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios shown.

### 1.3.5 Parallel Sysplex support

This section lists the connectivity options supported for Parallel Sysplex.

#### ISC-3

InterSystem Coupling Facility-3 links provide the connectivity required for data sharing between the Coupling Facility and the CPCs directly attached to it. The ISC-3 feature is available in Peer mode only and can be used to connect to other System z servers.

#### ICB-3

The Integrated Cluster Bus-3 link is like the ISC-3 link used by coupled systems to pass information back and forth over high speed links in a Parallel Sysplex environment. ICB-3 are used to connect z900 and z800 servers to a z9 EC. An ICB-3 connection is made at an extender card, called STI-3, in the I/O cage. ICB-4 links (2 gigabytes per second) should be used to provide coupling communication between z9 EC, z9 BC, z990, and z890 servers, because they deliver improved performance over ICB-3 links (1 gigabyte per second).

## ICB-4

An ICB-4 connection consists of one link that attaches directly to an STI port in the system, and does not require connectivity to a card in the I/O cage. Even though it is possible to connect and use a ICB-3 link between two z9 ECs or between a z9 EC, and a z9 BC, z990, or z890, the preferred option is to use the ICB-4 link, as it provides a faster link speed.

## Internal Coupling (IC)

The Internal Coupling-3 channel emulates the Coupling Facility functions in LIC between images within a single system. No hardware is required.

## System-Managed CF Structure Duplexing

System-Managed Coupling Facility (CF) Structure Duplexing provides a general purpose, hardware-assisted, and easy-to-exploit mechanism for duplexing CF structure data. This provides a robust recovery mechanism for failures (such as loss of a single structure or CF or loss of connectivity to a single CF) through rapid failover to the other structure instance of the duplex pair. Customers interested in deploying System-Managed CF Structure Duplexing should read the technical paper *System-Managed CF Structure Duplexing*, ZSW01975USEN. See the Parallel Sysplex Web site at:

<http://www.ibm.com/systems/z/psa/index.html>

## 1.3.6 Server Time Protocol

Server Time Protocol is a server-wide facility that is implemented in the Licensed Internal Code (LIC) of z9 EC, z9 BC, z990, z890 servers, and Coupling Facilities. STP presents a single view of time to PR/SM and provides the capability for multiple servers and CFs to maintain time synchronization with each other. A z9 EC, z9 BC, z990, or z890 server or CF may be enabled for STP by installing the STP feature. Each server and CF planned to be configured in a Coordinated Timing Network (CTN) must be STP-enabled.

The Server Time Protocol (STP) feature is designed to be the supported method for maintaining time synchronization between System z9, z990, z890 servers, and Coupling Facilities (CFs). The STP design uses a new concept called Coordinated Timing Network (CTN). A Coordinated Timing Network (CTN) is a collection of servers and Coupling Facilities that are time synchronized to a time value called Coordinated Server Time.

Prior to the STP, a Sysplex Timer® was used to synchronize the time of attached servers in an ETR network.

STP is for servers that have been configured to be in a Parallel Sysplex or a sysplex (without a Coupling Facility), as well as servers that are not in a sysplex, but need to be time-synchronized. STP is a message-based protocol in which timekeeping information is passed over data links between servers. The timekeeping information is transmitted over externally defined coupling links. The following coupling link types are supported:

- ▶ ISC-3 (standard or RPQ 892197 version) defined in peer mode (CFP)
- ▶ ICB-3 or ICB-4 links defined in peer mode (CBP)

STP provides the following additional value over the Sysplex Timer:

- ▶ STP supports a multisite timing network of up to 100 km without requiring an intermediate site. The fiber distance between Sysplex Timers cannot exceed 40 km.
- ▶ The STP design allows more stringent synchronization between servers and CFs using short communication links, such as ICB-4 and ICB-3 links, compared to servers and CFs using long ISC-3 links across sites.

- ▶ STP helps eliminate infrastructure requirements, such as power and space, needed to support the Sysplex Timers.
- ▶ STP helps eliminate maintenance costs associated with the Sysplex Timers.
- ▶ STP may reduce the fiber optic infrastructure requirements in a multisite configuration. Dedicated links may not be required to transmit timing information.

The Coordinated Timing Network (CTN) concept is used in order to help meet two key goals of System z customers:

- ▶ Concurrent migration from an existing External Time Reference (ETR) network to a timing network using STP.
- ▶ Capability of servers that cannot support STP to be synchronized in the same network as servers that support STP (z9, z990, and z890).

STP supports a multi-site timing network of up to 100 km (62 miles) over fiber optic cabling, allowing a Parallel Sysplex to span these distances and reducing the cross-site connectivity required for a multi-site Parallel Sysplex.

STP supports dial-out time services to set the time to an international time standard, such as *Coordinated Universal Time* (UTC), as well as adjust to the time standard on a periodic basis. In addition, setting of local time parameters, such as time zone and Daylight Saving Time (DST), and automatic updates of Daylight Saving Time, are supported.

STP is available as a charged feature on the z9 EC, z9 BC, z990, and z890 and is supported by z/OS V1.7 (PTFs are required to enable STP support) and above.

### 1.3.7 Intelligent Resource Director (IRD)

Exclusive to the IBM z/Architecture is Intelligent Resource Director (IRD), a function that optimizes processor and channel resource utilization across logical partitions based on workload priorities. IRD combines the strengths of the PR/SM, Parallel Sysplex clustering, and z/OS Workload Manager.

Intelligent Resource Director uses the concept of an *logical partition cluster*, the subset of z/OS systems in a Parallel Sysplex cluster that are running as logical partitions on the same server. In a Parallel Sysplex environment, Workload Manager directs work to the appropriate resources, based on business policy. With IRD, resources are directed to the priority work. Together, Parallel Sysplex technology and IRD provide flexibility and responsiveness to e-business workloads that are unrivaled in the industry.

IRD has three major functions:

- ▶ Channel Subsystem Priority Queuing - Allows priority queuing of I/O requests within the Channel Subsystem, and the specification of relative priority among logical partitions. WLM in goal mode sets priorities for a logical partition, and coordinates this activity among clustered logical partitions.
- ▶ Dynamic Channel Path Management - Enables customers to have channel paths that dynamically and automatically move to those ESCON I/O devices that have a need for additional bandwidth due to high I/O activity. The benefits are enhanced by the use of goal mode and clustered logical partitions.
- ▶ LPAR CPU Management - Workload Manager (WLM) dynamically adjusts the number of logical processors within a logical partition and the processor weight, based on the WLM policy. The ability to move the CPU weights across a logical partition cluster provides processing power to where it is most needed, based on WLM goal mode policy.



### 1.3.8 Capacity On Demand

The z9 EC servers have *concurrent* upgrade capability through the Capacity Upgrade on Demand (CUoD) function. This function is also used by the Capacity BackUp (CBU) feature, Customer Initiated Upgrades (CIUs), and the On/Off Capacity Upgrade on Demand implementation.

#### Capacity Upgrade on Demand (CUoD)

Capacity Upgrade on Demand offers server upgrades through Licensed Internal Code (LIC) enabling. CUoD can concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs) and memory to an existing configuration when no hardware changes are required, resulting in an upgraded server. Also, I/O features can be added concurrently.

#### Capacity BackUp (CBU)

Capacity BackUp (CBU) is the temporary activation of CPs, IFLs, ICFs, zAAPs, and zIIPs for robust disaster recovery. The CBU features provide the ability to concurrently increment the CP or specialty engine capacity of System z9 server, using LIC-CC, in the event of an unforeseen loss of substantial System z9 computing capacity at one or more sites. The CBU features contain additional resources and alter the target server to an agreed upon configuration for up to a 90-day period of time. CBU CP, IFL, ICF, zIIP, and zAAP activations are mutually exclusive with On/Off CoD activation. CBU test upgrades for the sole purpose of checking the ability in the event of an emergency are permitted. The CBU offering allows for up to five, 10-day tests over five years.

#### Customer Initiated Upgrade (CIU)

CIU is designed to allow to respond to sudden increased capacity requirements by requesting a System z9 Processor Unit (PU) or memory upgrade through the Web, using IBM Resource Link™, and downloading and applying it to the System z9 server using the system's Remote Support connection. Further, with the Express option on CIU, an upgrade may be made available for installation within a few hours after order submission.

#### On/Off Capacity on Demand

On/Off Capacity on Demand (On/Off CoD) is designed to temporarily turn on Central Processors (CPs), Internal Coupling Facilities (ICFs), Integrated Facilities for Linux (IFLs), System z9 Integrated Information Processors (zIIPs), or System z Application Assist Processors (zAAPs). On/Off CoD is delivered through the function of Customer Initiated Upgrade (CIU).

Both On/Off CoD and CBU can reside on the server, but only one can be activated at a time.

#### On/Off CoD test

On/Off CoD allows for a no-charge test. No IBM charges are assessed for the test, including IBM charges associated with temporary hardware capacity, IBM software, or IBM maintenance. This test can be used to validate the processes to download, activate, and deactivate On/Off CoD capacity nondisruptively.

An additional test offering is available on the z9 EC. The Administrative On/Off Capacity on Demand (On/Off CoD) Test enables customers to order zero capacity PU features through Resource Link. This test allows a customer to thoroughly rehearse the entire On/Off CoD process without incurring any cost. There is an unlimited number of tests for Administrative On/Off CoD Test and no time period restrictions.

### 1.3.9 Reliability, Availability, and Serviceability (RAS)

The z9 EC RAS strategy is a building-block approach developed to meet the customer's stringent requirements of achieving Continuous Reliable Operation. Those building blocks are Error Prevention, Error Detection, Recovery, Problem Determination, Service Structure, Change Management, and Measurement and Analysis.

The initial focus is on preventing failures from occurring in the first place. This is accomplished by using *Hi-Rel* (highest reliability) components, using screening, sorting, burn-in, run-in, and by taking advantage of technology integration. For Licensed Internal Code and hardware design, failures are eliminated through rigorous design rules, design walk-through, peer reviews, element, subsystem and system simulation, and extensive engineering and manufacturing testing.

The z9 EC RAS strategy is focused on a recovery design that is necessary to mask errors and make them “transparent” to customer operations. There is an extensive hardware recovery design implemented to be able to detect and correct array faults. In cases where total transparency cannot be achieved, the capability exists for the customer to restart the server with the maximum possible capacity.

#### **Enhanced Book Availability to help reduce the impact of outages**

The z9 EC is designed to allow a single book, in a multibook server, to be concurrently removed from the server and reinstalled during an upgrade or repair action. To help minimize the impact on current workloads and applications, it is necessary to ensure that there is sufficient inactive physical resources on the remaining books to complete a book removal. For a maximum availability configuration, review recommendations in 8.3, “Enhanced Book Availability (EBA)” on page 232 or consider the purchase of one additional book.

To help ensure that there is the appropriate level of memory, it is wise to consider the selection of the flexible memory option to provide additional resources when replacing a book or when considering plan ahead options for the future.

Enhanced Book Availability is an extension of the support for Concurrent Book Add (CBA). CBA is designed to allow to concurrently upgrade a z9 EC by integrating a second, third, or fourth book into the server without affecting application processing.

#### **Redundant I/O Interconnect**

The z9 EC is designed to allow a single book, in a multibook server, to be concurrently removed from the server and reinstalled during an upgrade or repair, while continuing to provide connectivity to the server I/O resources using a second path from a different book. Redundant I/O Interconnect is exclusive to System z9.

#### **Enhanced driver maintenance**

One of the contributors to downtime during planned outages is Licensed Internal Code (LIC) updates performed in support of new features and functions. When properly configured, the z9 EC is designed to support activating a selected new LIC level concurrently. Concurrent activation of the selected new LIC level is only supported at specific sync points (points in the maintenance process when LIC may be applied concurrently - MCL service level). Sync points may exist throughout the life of the current LIC level. Once a sync point has passed, it will be required to wait until the next sync point supporting concurrent activation of a new LIC level. Certain LIC updates will not be supported by this function.

### Dynamic oscillator switch-over

The z9 EC has two oscillator cards, a primary and a backup. In the event of a failure of the primary oscillator card, the backup is designed to detect the failure, switch over, and provide the clock signal to the server transparently. Dynamic oscillator switch-over is exclusive to the z9 EC.

### MBA fanout card hot-plug

A Memory Bus Adapter (MBA) fanout card is designed to provide the path for data between memory and I/O using Self-Timed Interconnect (STI) cables. A hot-pluggable MBA fanout card is available in the z9 EC. Up to eight MBA fanout cards are available per book for a total of up to 32 MBA fanout cards on the z9 EC when four books are installed. In the event of an outage, an MBA fanout card, used for I/O, may be concurrently repaired using Redundant I/O Interconnect.

## 1.3.10 Software

Software support of the System z9 EC requires:

- ▶ Any in-service z/OS release
- ▶ Any in-service z/VM release
- ▶ z/VSE V3R1 and later
- ▶ TPF V4.1 and z/TPF V1.1
- ▶ Linux on System z: the currently available distributions of SUSE SLES and Red Hat RHEL

## 1.4 Service-Oriented Architecture (SOA)

A Service-Oriented Architecture (SOA) is an architectural style. It supports a business integrating within itself, with customers, partners, and suppliers. SOA is a way to create an On Demand Operating Environment (ODOE) that supports becoming an On Demand Business.

An On Demand Business uses business processes that are supported by an On Demand Operating Environment. The characteristics of an On Demand Operating Environment deal with application flexibility, IT optimization, security, and many more capabilities. An ODOE is not an architecture or architectural style. It is whatever the customer chooses to do to create an IT environment in support of becoming an On Demand Business.

There is a growing recognition within the IT industry of the potential benefits of a Service-Oriented Architecture (SOA). A well architected SOA can bring many business benefits, including the following:

- ▶ Better reuse existing investment in core business applications
- ▶ A common platform that uses enterprise-wide standards to provide utility services across geographies
- ▶ Reduced time to market for new products and services and reduced operational costs by providing a flexible infrastructure and IT delivery environment
- ▶ The ability to more easily integrate acquisitions and to provide *white labeling* both inside, and external to the enterprise

**Note:** White labeling is the mechanism where an enterprise provides the business functionality and operating environment under the brand name of another organization. An enterprise offers this as a service to other organizations that do not want to support the technical infrastructure of their applications themselves. This may be less expensive to the white label organization since they benefit from the economies of scale of the provider organization.

Service-Oriented Architecture (SOA) is an architectural style to create an On Demand Operating Environment. The core of an ODOE are services. These service are:

- ▶ Process Services
- ▶ Business Application Services
- ▶ Information Services
- ▶ Access Services
- ▶ Interaction Services
- ▶ Partner services

These services are embedded infrastructure services, taking care of optimization of throughput, availability, and performance, IT service management, to manage and secure services, applications and resources, and development services, offering an integrated environment for design, and creation of solutions.

The communication between services in a SOA is built on the Enterprise Service Bus (ESB), which is a common distributing network for services to work with. The ESB is the heart of the SOA environment and must be reliable, available, and secure. This architectural construct delivers all connectivity capabilities required to use services implemented across the entire architecture. The ESB provides the following fundamental services:

- ▶ Transport services providing the fundamental connection layer.
- ▶ Event services that allow the system to respond to specific events that are part of a business process.
- ▶ Mediation services like transformation and validation services that allow loose-coupling between interacting services in the system.

SOA defines an application development architecture model, independent of platform, technology and vendor, but when the decision needs to be made about where to deploy SOA applications and infrastructure, different aspects and quality of services (QoS) provided by each platform and each vendor should be considered.

The IBM Mainframe has successfully been running the core IT systems of many businesses, from medium to very large size, for more than 40 years. During these years, IBM has consistently invested in the evolution of the mainframe's unparalleled technology. Mainframes have incorporated new technologies and computing models, from the centralized model to the Web model, from Assembler and COBOL languages to Java. Today, the mainframe provides all capabilities to form the backbone in the enterprise Service-Oriented Architecture, and more than ever offers a unique proposition to become a key component of a SOA deployment.

Mainframes are in use by thousands of enterprises world wide, with trillions of dollars invested in applications and skills. They run by far the most of the total business transactions worldwide, providing real 24x7 availability. When considering the value propositions of the SOA, and at the same time the value of the existing IT assets, the qualities of the mainframe, the potential of the combination of the mainframe strengths and the SOA concepts becomes clear. The combination clearly manifests itself when transitioning to an ODOE, using SOA as a reference.

The generally recognized IBM mainframe qualities fall into a number of categories:

- ▶ Security
- ▶ Manageability
- ▶ Virtualization and Workload Management
- ▶ Reliability
- ▶ Scalability
- ▶ Availability
- ▶ Transaction processing
- ▶ Batch processing

The SOA philosophy proclaims advantages through strong corporate IT governance and reuse. The mainframe was made to support these tasks.

The Enterprise Service Bus is the intermediary through which all service communication runs. It therefore requires the highest levels of availability, scalability, security, and performance. Bundling the highest qualities of service in the industry provided by z/OS, and the additional qualities gained from co-locating the business processes with the back-end service, is an attractive proposition.

### ***z/OS is the platform for core application services***

New applications based on J2EE™, which require mainframe qualities of service, can run on WebSphere Application Server for z/OS, taking advantage of the cost-effectiveness of zAAP engines for Java. IMS™ and CICS® are also well positioned for the development of new functionality and can participate fully in an SOA. This is an attractive option for those customers that have deep investments in mainframe applications, and want to reuse these assets.

### ***Leverage existing skills***

SOA brings new opportunities for existing z/OS skills. Since the Java popularity in the commercial business, COBOL programmers were challenged to acquire this new skill in order to stay competitive in the marketplace. Now, moving to a SOA, Java is no longer a must, the SOA programming model does not require a specific programming language, COBOL, and PL/I can be part of it.

### ***z/OS provides the qualities a SOA requires***

Deploying the SOA on z/OS not only brings the SOA functionality to the mainframe, and the mainframe qualities to the SOA, but the deployment on z/OS extends the functionality of the SOA with components like WebSphere Application Server, WebSphere Message Broker, DB2, and MQ, and brings functional options no other platform can provide.

It is generally accepted that a Service-Oriented Architecture offers many business benefits, including better reuse of existing assets, a more flexible IT infrastructure, and reduced costs. Many customers have invested heavily in business applications running on z/OS, using a combination of CICS, IMS, and DB2 to deliver the majority of transactions. z/OS together with WebSphere Application Server form the center of a service based architecture, exposing the business functionality of transactions running on CICS and IMS.

## 1.5 Summary

IBM mainframes provide an advanced combination of reliability, availability, security, scalability, and virtualization. The IBM System z9, delivering excellence in large-scale enterprise computing, is designed and optimized for On Demand Business. The IBM System z9 Enterprise Class has been designed to deliver:

- ▶ Great granularity with subcapacity engines and high scalability with up to 54 engines on a single server.
- ▶ Fast and robust connectivity.
  - The FICON Express4 and FICON Express 2 cards enable up to 336 FICON channels.
  - The Modified Indirect Data Address Word (MIDAW) facility is designed to help improve performance for native FICON applications that use data chaining for extended format data sets by reducing channel, director, and control unit overhead.
  - Resource sharing in the open environment. N\_Port ID virtualization (NPIV) allows the sharing of FCP channels among operating system images in logical partitions or virtual machines.
  - Networking with the OSA-Express2 1000Base-T Ethernet feature supporting large send (offloading TCP segmentation processing), 640 TCP/IP stacks for improved virtualization, and concurrent LIC update to minimize traffic disruption. The OSA-Express2 OSN (OSA for NCP) feature provides channel connectivity from System z9 operating systems to the communications controller for Linux (CCL), which can help protect investments in traditional SNA applications and data.
  - OSA-Express2 Link Aggregation for z/VM.
  - OSA-Express2 Network Traffic Analyzer, QDIO Diagnostic Synchronization, and Layer 3 virtual MAC support.
- ▶ Strong security with many hashing algorithms and strong encryption.
- ▶ Extended virtualization capabilities with the ability to handle up to 60 logical partitions. Virtualization helps to reduce management complexity and can facilitate a more efficient use of system resources.

Application of the z9 EC advanced virtualization technologies provides accelerated application development and integration with applications across various platforms. With support for IBM WebSphere software, full support for SOA, Web services, J2EE, Linux, and Open Standards, the z9 EC is intended to be a platform of choice for integration of a new generation of applications with existing applications and data.

- ▶ Customer Initiated Upgrade can be used to allow for the permanent nondisruptive addition of one or more CPs, IFLs, zAAPs, zIIPs, and ICFs.

On/Off Capacity on Demand allows an installation to turn on additional, temporary system resources to meet the demands of business cycles or unexpected demand throughout the year.

The Capacity BackUp (CBU) feature gives extra capacity to operate in emergency situations. The z9 EC allows CBU for specialized processing units, such as IFLs, ICFs, zAAPs, and zIIPs.

- ▶ Reduced scheduled downtime with concurrent book and memory add/repair/replace, and reduced planned outages, too, with concurrent microcode upgrades.

With proper planning, and when properly configured, the z9 EC is designed to allow a single book, in a multi-book server, to be concurrently removed from the server and reinstalled during an upgrade or repair. Redundant I/O Interconnect (RII) provides

connectivity to I/O resources on other books during the removal in support of this capability.

I/O availability features include assists for isolation of ESCON and FICON fiber optic cabling problems and FICON link incident reporting.

Archived

Archived





## System structure and design

This chapter introduces the IBM System z9 system structure. Significant functions and features are described, along with their characteristics and options.

The goal of this chapter is to explain how the z9 EC is structured, what its main components are, and how these components interconnect from a physical and logical point of view. This information is useful for planning purposes and will help to define the configuration that best fits given requirements.

The following topics are included:

- ▶ 2.1, “System structure” on page 28
- ▶ 2.2, “System design” on page 45
- ▶ 2.3, “Model configurations” on page 71
- ▶ 2.4, “Logical partitioning” on page 79
- ▶ 2.5, “Storage operations” on page 84

## 2.1 System structure

The z9 EC structure is the result of the continuous evolution of S/390® to zSeries®, and System z9 since CMOS servers were introduced in 1994. Its structure and also its design have been continuously improved, adding more capacity, performance, functionality, and connectivity, keeping in mind the balanced system approach where memory sizes, internal bandwidth, processing capacity, and connectivity are in balance with each other.

The objective of the z9 EC system structure and design is to offer a flexible infrastructure to accommodate a wide range of operating systems and applications, whether they be traditional or emerging e-business applications based on WebSphere, Java, and Linux, for integration and deployment in heterogeneous business solutions.

For that purpose, the z9 EC introduces improved uniprocessor performance, an increase in the number of usable processors per system, an increased number of logical partitions, separate handling of characterized PUs, additional instructions to speed up the processing of new workload types, and an improved memory structure on top of the already available superscalar microprocessor architecture. In order to keep a balanced system, the I/O bandwidth and available memory sizes have been increased accordingly.

### 2.1.1 Book concept

The z9 EC Central Processor Complex (CPC) uses packaging concept based on books. A book contains processors (PUs), memory, and connectors to I/O cages and ICB-4 links. Books are located in the CEC cage in Frame A. A z9 EC server has at least one book, but may have up to four books installed.

A book and its components are shown in Figure 2-1. Each book contains:

- ▶ 12 or 16 Processor Units (PUs). The PUs reside on microprocessor chips located on a Multi-Chip Module (MCM).
- ▶ 16 GB to 128 GB physical memory. At least four memory cards are present, each containing 4, 8, or 16 GB. Four additional memory cards are added when the configured memory size goes beyond 64 GB.
- ▶ Up to eight Memory Bus Adapter fanout cards (MBAs), supporting up to 16 Self-Timed Interconnects (STIs) to the I/O cages or ICB channels.

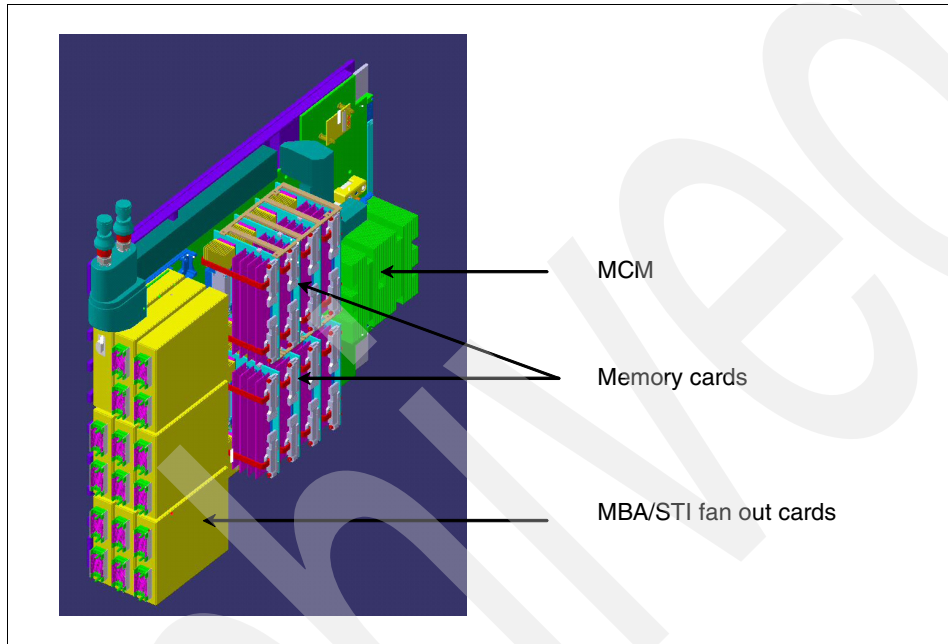


Figure 2-1 Book structure and components

Up to four books can reside in the CEC cage. Books plug into cards, which plug into slots of the CEC cage board.

### Power

Each book get its power from two Distributed Converter Assemblies (DCA) that reside on the opposite side of the processor board. The DCAs provide the required power for the book. Each book is supported by two DCAs. The N+1 power supply design means that there is more DCA capacity than is required for the book. If one DCA fails, the power requirement for a book can still be satisfied from the remaining DCA. The DCAs can be concurrently maintained, which means that replacement of one DCA can be done without taking the book down.

There is the location of two oscillator cards (OSC) and the two optional external time reference cards (ETR) between two sets of DCAs. If installed, there are two ETR ports to which an optional Sysplex Timer can be connected.

The two oscillator cards act as a primary and a backup. In case the primary oscillator card would fail, the backup card detects the failure and continues to provide the clock signal so that no outage due to an oscillator failure is taken.

Seen from the top, the packaging of a four-book system appears as shown (schematically) in Figure 2-2.

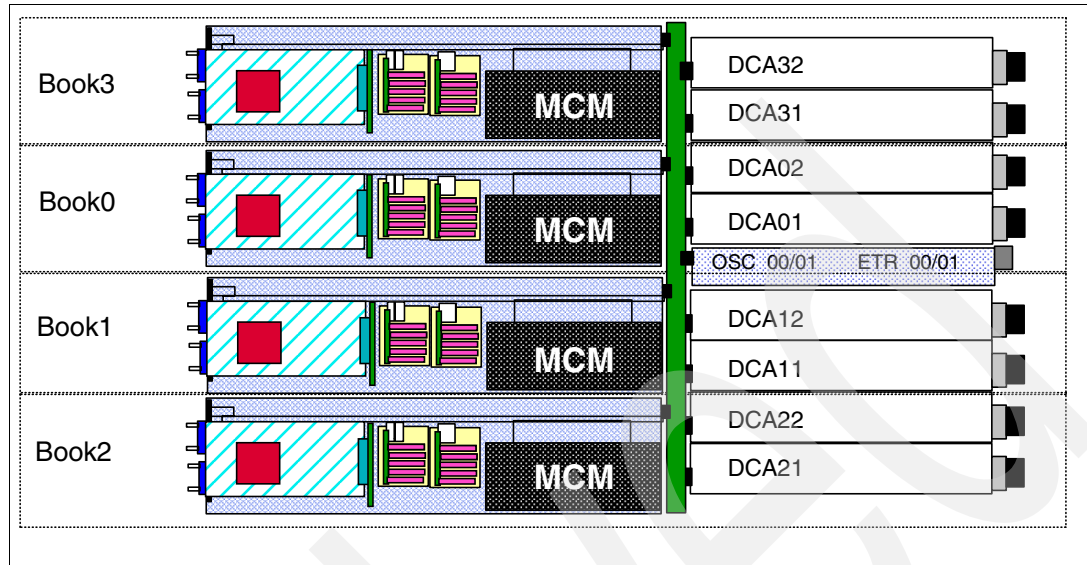


Figure 2-2 Book and power packaging (top view)

Located within each book are nine cards, eight of which are Memory Bus Adapter/STI fanout cards. Each fanout card drives two STIs, for a total of 16 STI connections to the I/O cages or ICB-4 channels (see Figure 2-6 on page 37).

The order of book installation in Figure 2-2 is:

- ▶ In a one-book model, only book 0 is present.
- ▶ A two-book model has books 0 and 1.
- ▶ A three-book model has books 0, 1, and 2.
- ▶ A four-book model has books 0, 1, 2, and 3.

Book installation from one to up to four books is concurrent.

## Cooling

The z9 EC is an air-cooled system assisted by refrigeration. Refrigeration is provided by a closed-loop liquid cooling subsystem. The entire cooling subsystem has a modular construction. Its components and functions are found throughout the cages, and are made up of three subsystems:

- ▶ The Modular Refrigeration Units (MRU)
  - One or two MRUs (MRU0 and MRU1), located in the front of the A-cage above the books, provide refrigeration to the content of the books together with Motor Drive Assemblies in (MDAs) in the rear.
  - A one-book system has MRU0 installed. Upgrading to a two-book system causes MRU1 to be installed, providing all refrigeration needs for a four-book system. Concurrent repair of an MRU is possible by taking advantage of the hybrid cooling implementation described in the next section.
- ▶ The Motor Scroll Assembly (MSA)

- ▶ The Motor Drive Assembly (MDA)

MDAs are found throughout the frames to provide air cooling where required. They are located at the bottom front of each cage, and in between the CEC cage and I/O cage, one in combination with the MSAs.

### Hybrid cooling system

The z9 EC has a hybrid cooling system that is designed to lower power consumption. Normal cooling is provided by one or two MRUs connected to the heat sinks of all MCMs in all books.

If one of the MRUs fails, backup MSAs are switched in to compensate for the lost refrigeration capability with additional air cooling. At the same time, the oscillator card is set to a slower cycle time, slowing the system down by up to 10 percent of its maximum capacity, to allow the degraded cooling capacity to maintain the proper temperature range. Running at a slower cycle time, the MCMs produce less heat. The slowdown process is done in steps, based on the temperature in the books.

Figure 2-3 shows the refrigeration scope of MRU0 and MRU1.

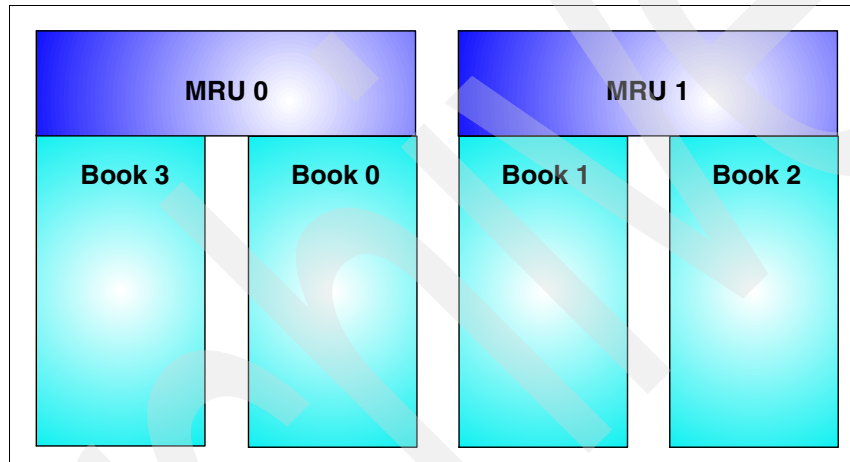


Figure 2-3 MRU scope

### 2.1.2 Models

The z9 EC has five orderable models. The model numbers are directly related to the maximum number of PUs that can be characterized by the installation. For customer use, PUs can be characterized as CPs, IFLs, ICFs, zAAPs, zIIPs, or if need be, additional SAPs.

- ▶ The z9 EC Model S08 has one book with 12 PUs, of which eight can be characterized. The four remaining PUs are two system assist processors (SAPs) and two spares.
- ▶ The z9 EC Model S18 has two books with 12 PUs in each book for a total of 24 PUs, of which 18 can be characterized. The six remaining PUs are four system assist processors (SAPs), two in each book, and two spares.
- ▶ The z9 EC Model S28 has three books with 12 PUs in each book for a total of 36 PUs, of which 28 can be characterized. The eight remaining PUs are six system assist processors (SAPs), two in each book, and two spares.
- ▶ The z9 EC Model S38 has four books with 12 PUs in each book for a total of 48 PUs, of which 38 can be characterized. The 10 remaining PUs are eight system assist processors (SAPs), two in each book, and two spares.

- ▶ The z9 EC Model S54 has four books with 16 PUs in each book for a total of 64 PUs, of which 54 can be characterized. The 10 remaining PUs are eight system assist processors (SAPs), two in each book, and two spares.

The last two digits of the model number reflect the maximum number of PUs that can be characterized for installation use. The PUs can be characterized as CPs, IFLs, ICFs, zAAPs, zIIPs, or additional SAPs.

Whether one, two, three, or four books are present, to the user, all books together appear as one Symmetric Multi Processor (SMP) with a certain number of CPs, IFL, ICFs, and zAAPs, zIIPs, and a certain amount of memory and bandwidth to drive the I/O channels and devices. The packaging is designed to scale to a 54-PU Symmetric Multi-Processor (SMP) server in four books. Scaling from a Model S08, S18, S28, or S38 to up to 54 PUs in a Model S54 is disruptive. If the starting point is a Model S54, scaling to 54 PUs characterized as CPs, or specialty engines, is nondisruptive.

### 2.1.3 Memory

Maximum physical memory sizes are directly related to the number of books in the system. Each book may contain a maximum of 128 GB of physical memory. Physical memory is organized in two banks of four memory cards each. One bank of four memory cards in each book is always populated. The memory size per bank per book may differ. Also, memory sizes in each book do not have to be similar; different books may contain different amounts of memory. The minimum orderable amount of memory is 16 GB, system-wide.

- ▶ A one-book system (z9 EC Model S08) may contain 16 GB, 32 GB, 64 GB, or 128 GB of physical memory. Memory is orderable in 16 GB increments for customer use.
- ▶ A two-book system (z9 EC Model S18) may contain up to a maximum of 256 GB of physical memory. For memory card distribution variations in newly built two-book systems, refer to Table 2-2 on page 33. Memory is orderable in 16 GB increments for customer use.
- ▶ A three-book system (z9 EC Model S28) may contain up to a maximum of 384 GB of physical memory. For memory card distribution variation in a newly built three-book system, refer to Table 2-2 on page 33. Memory is orderable in 16 GB increments for customer use.
- ▶ A four-book system (z9 EC Model S38 or z9 EC S54) may contain up to a maximum of 512 GB of physical memory. For memory card distribution variation in a newly built four-book system, refer to Table 2-2 on page 33. Memory is orderable in 16 GB increments for customer use.

The system physical memory is the sum of all book memories. Not all books need to contain the same amount of memory, and not all installed memory is necessarily configured for use.

#### Memory sizes

The minimum orderable amount of usable memory for all models is 16 GB. Memory upgrades are available in 16 GB increments:

- ▶ z9 EC Model S08, from 16 to 128 GB
- ▶ z9 EC Model S18, from 16 to 256 GB
- ▶ z9 EC Model S28, from 16 to 384 GB
- ▶ z9 EC Model S38, from 16 to 512 GB
- ▶ z9 EC Model S54, from 16 to 512 GB

Physically, the memory cards are organized as follows:

- ▶ Within a book, memory is organized in two rows (banks) of four memory cards. Since book memory is organized in up to four Processor Memory Arrays (PMAs) and since one memory card only encompasses half a PMA, eight memory cards are needed for a full book with 128 GB.
- ▶ A book always contains a minimum of four memory cards. Memory cards come in three sizes:
  - 4 GB (512 Mb DRAMs)
  - 8 GB (512 Mb, or 1 Gb DRAMs)
  - 16 GB (1 Gb DRAMs)
- ▶ Within a book, different memory card sizes can be plugged as long as the DRAM sizes are the same. Not all books necessarily need to have the same amount of physical memory installed.
- ▶ A book may have more memory installed than enabled. The excess amount of memory can be installed by a Licensed Internal Code code load (sometimes called *dial-a-Gig*) when required by the installation.
- ▶ On initial installation, the amount of physical memory in a given model is nearest to the smallest possible size. An example of this, for a z9 EC Model S08, is shown in Table 2-1.
- ▶ Memory upgrades are satisfied from already installed unused memory capacity until exhausted. When no more unused memory is available from the installed memory cards, cards have to be upgraded to a higher capacity, or the addition of a book with additional memory is necessary.

Table 2-1 Model S08 memory configuration

Purchased Memory: Model S08	16 GB	32 GB	48 GB	64 GB	80 GB	96 GB	112 GB	128 GB
Memory card configuration (cards x size)	4 x 4	8 x 4	8 x 8	8 x 8	8 x 16	8 x 16	8 x 16	8 x 16

Table 2-2 shows examples of memory configurations (all possible combinations up to 128 GB are shown). It shows that an z9 EC Model S08 may have 16 GB of usable memory out of a minimum of 16 GB physically installed, and that an z9 EC Model S38, though unlikely, may have 16 GB of usable memory out of a minimum of 64 GB physical memory.

Table 2-2 New build z9 EC physical memory card distribution

Purchased capacity	z9 EC mod S08 Physical Cards	z9 EC mod S18 Physical Cards	z9 EC mod S28 Physical Cards	z9 EC mod S38 or Model S54 Physical Cards
16 GB	Book 0: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
32 GB	Book 0: 8 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB

Purchased capacity	z9 EC mod S08 Physical Cards	z9 EC mod S18 Physical Cards	z9 EC mod S28 Physical Cards	z9 EC mod S38 or Model S54 Physical Cards
48 GB	Book 0: 8 x 8 GB	Book 0: 8 x 4 GB Book 1: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
64 GB	Book 0: 8 x 8 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
80 GB	Book 0: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 4 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 4 x 4 GB	Book 0: 8 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
96 GB	Book 0: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
112 GB	Book 0: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB	Book 0: 8 x 8 GB Book 1: 8 x 4 GB Book 2: 4 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB Book 3: 4 x 4 GB
128 GB	Book 0: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB	Book 0: 8 x 8 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB Book 3: 8 x 4 GB

**Note:** The amount of memory available for use is the sum of all enabled memory on all memory cards in all books.

When activated, a logical partition can use memory resources located in any book. No matter in which book the memory resides, a logical partition has access to that memory if so allocated. Despite the book structure, the z9 EC is still a Symmetric Multi-Processor (SMP).

Memory upgrade is concurrent when it requires no change of the physical memory cards. A memory card change is disruptive when no use is made of Enhanced Book Replacement. See “Enhanced Book Availability” on page 39.

### Chip sparing

Chip sparing is implemented by the use of X4 DRAMs across eight DIMMs where eight DIMMs make up one PMA. This results in four spares per PMA with Chipkill™. Chipkill is an advanced error correction code that corrects multi-bit memory errors.

### Memory upgrades

For a model upgrade that results in the addition of a book, the minimum memory increment is added to the system. Remember, the minimum physical memory size in a book is 16 GB. During a model upgrade, the addition of a book is a concurrent operation. The addition of the physical memory that is in the added book is also concurrent.



If all or part of the additional memory is enabled for installation use, it becomes available to an active logical partition if this partition has reserved storage defined (see 2.5.1, “Reserved storage” on page 86 for more detailed information). Or, it may be used by an already defined logical partition that is activated after the memory addition.

### Book replacement and memory

With Enhanced Book Availability as supported for z9 EC (see “Enhanced Book Availability” on page 39), having sufficient resources available to accommodate resources that are lost when a book is removed for upgrade or repair is needed. Removal of a book most of the time results in the removal of active memory. With the Flexible Memory option (see “Flexible Memory option” on page 35) it is possible to evacuate the affected memory and reallocate its use elsewhere in the system. This requires additional available memory to compensate for the memory lost with the removal of the book.

### Flexible Memory option

With the Flexible Memory option, sufficient inactive memory resources are made available for use when replacing a book. When ordering memory for a z9 EC, additional flexible memory can be specified. For example, on a z9 EC Model S18, this results in doubling the purchased amount for normal use. This ensures that the content of the memory in the book to be removed can be moved to the excess memory in the other book. Flexible memory must be purchased but cannot be used for normal every day use. For that reason, a different purchase price for the flexible memory is offered to increase the overall availability of the system.

In Table 2-3, the physical memory requirements for high availability for a two-book system are shown. The table does not show a complete list of all options. Card combinations up to 128 GB are shown.

Table 2-3 New build z9 EC physical memory card distribution for maximum availability

Purchased capacity	z9 EC mod S08 Physical Cards	z9 EC mod S18 Physical Cards	z9 EC mod S28 Physical Cards	z9 EC mod S38 or Model S54 Physical Cards
16 GB		Book 0: 4 x 4 GB Book 1: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
32 GB		Book 0: 8 x 4 GB Book 1: 8 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
48 GB		Book 0: 8 x 8 GB Book 1: 8 x 8 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 4 x 4 GB	Book 0: 4 x 4 GB Book 1: 4 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
64 GB		Book 0: 8 x 8 GB Book 1: 8 x 8 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 4 x 4 GB Book 3: 4 x 4 GB
80 GB		Book 0: 8 x 16 GB Book 1: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB Book 2: 4 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB Book 3: 4 x 4 GB

Purchased capacity	z9 EC mod S08 Physical Cards	z9 EC mod S18 Physical Cards	z9 EC mod S28 Physical Cards	z9 EC mod S38 or Model S54 Physical Cards
96 GB		Book 0: 8 x 16 GB Book 1: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB Book 2: 8 x 4 GB	Book 0: 8 x 4 GB Book 1: 8 x 4 GB Book 2: 8 x 4 GB Book 3: 8 x 4 GB
112 GB		Book 0: 8 x 16 GB Book 1: 8 x 16 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB Book 2: 8 x 8 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB Book 2: 8 x 4 GB Book 3: 4 x 4 GB
128 GB		Book 0: 8 x 16 GB Book 1: 8 x 16 GB	Book 0: 8 x 16 GB Book 1: 8 x 16 GB Book 2: 4 x 4 GB	Book 0: 8 x 8 GB Book 1: 8 x 8 GB Book 2: 8 x 4 GB Book 3: 8 x 4 GB

### 2.1.4 Ring topology

Two concentric loops or rings are constructed (one flowing clock-wise, and one flowing counter clock-wise) such that in a four-book system each book only is connected to two others, which means that only data transfers or data transactions to the third book require passing through one of the other books.

Book-to-book communications are organized as shown in Figure 2-4. Book 0 communicates with book 2 and book 3; communication to book 1 must go through another book (either book 2 or book 3). In a two or three book configuration, jumper books complete the ring.

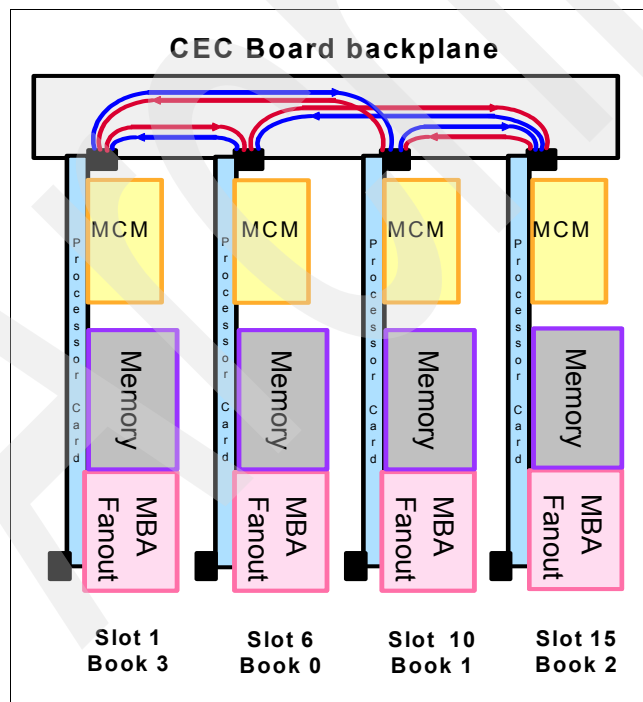


Figure 2-4 Concentric ring structure

A memory-coherent director optimizes ring traffic and filters out cache traffic by not looking on the ring for cache hits in other books if it is certain that the resources for a given logical partition exist in the same book.

The Level 2 (L2) cache is implemented on four cache (SD) chips. Each SD chip holds 10 MB, resulting in a 40 MB L2 cache per book. The L2 cache is shared by all PUs in the book and has a store-in buffer design. The connection to processor memory is done through four high-speed memory buses.

There is a ring structure within which the books maintain interbook communication at the L2 cache level. Additional books extend the function of the ring structure for interbook communication. A simplified ring topology for 2, 3, and book systems is shown in Figure 2-5. A book jumper completes the ring in order to be able to insert additional books into the ring nondisruptively.

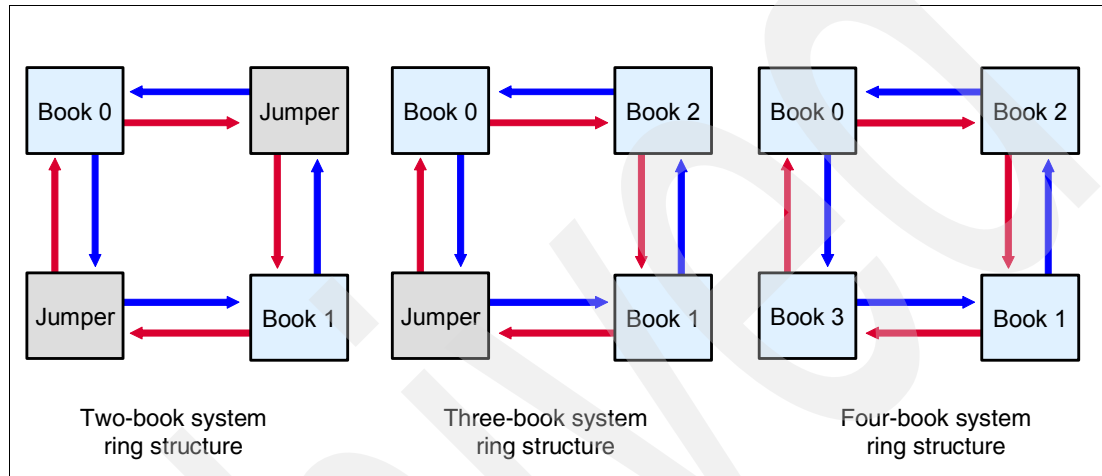


Figure 2-5 Multi-book system ring structure

### 2.1.5 Connectivity

STI connections to I/O cages, STI-3 extender cards, and ICB-4 links are driven from the Memory Bus Adapters (MBAs) fanout cards that are located on the front of the book. Figure 2-6 shows the location of the STI connectors and the MBA fanout cards.

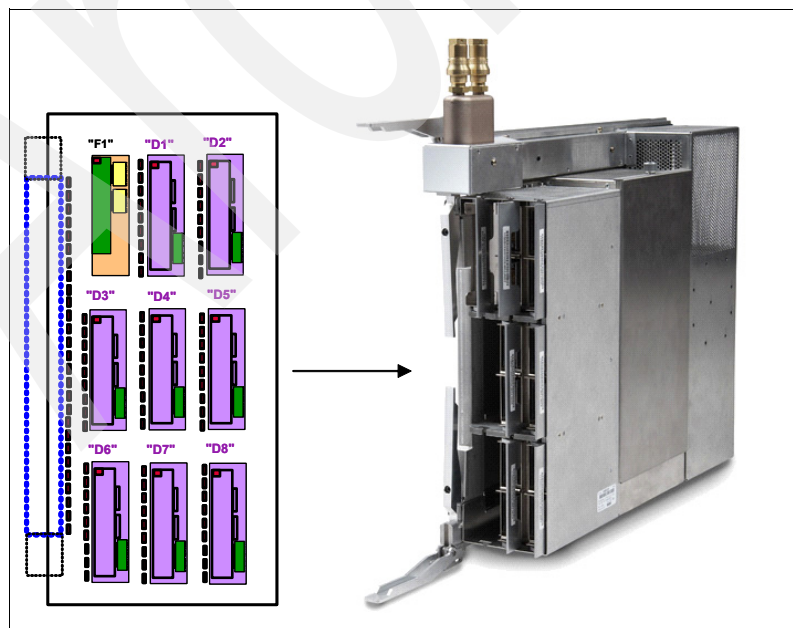


Figure 2-6 STI connectors and MBA fanout cards

Each MBA fanout card connects to up to two STI cables. There are up to eight MBA fanout cards (numbered D1 to D8) per book, each driving two STIs, resulting in 16 STIs per book. An MBA fanout card can be repaired concurrently with the use of Redundant I/O Interconnect. See “Redundant I/O Interconnect” on page 38.

All 16 STIs in a book have a data rate of 2.7 GB per second. Depending on the channel types installed, a maximum of 1024 channels per server is supported.

Two STIs are related to an MBA fanout card. When configuring for availability, channels, links, and OSAs across books, MBAs and STIs should be balanced. In a system configured for maximum availability, alternate paths will maintain access to critical I/O devices, such as disks, networks, and so on.

Enhanced Book Availability allows a single book in a multi-book server to be concurrently removed and reinstalled for an upgrade or a repair. Removing a book would also mean that the connectivity to the I/O connected to that book is lost. To prevent connectivity loss, the z9 EC Redundant I/O Interconnect feature allows you to maintain full connection to critical devices (except ICBs) when a book is removed.

### Redundant I/O Interconnect

Redundant I/O Interconnect is accomplished by the facilities of the Self-Timed Interconnect Multiplexer (STI-MP) card. Each STI-MP card is connected to an STI jack located in the MBA fanout card of a book. STI-MP cards are half-high cards and are interconnected with cards called STI-A8 and STI-A4, allowing Redundant I/O Interconnect in case the STI connection coming from a book ceases to function, as is the case when, for example, a book is removed. A conceptual view of how Redundant I/O Interconnect is accomplished is shown in Figure 2-7.

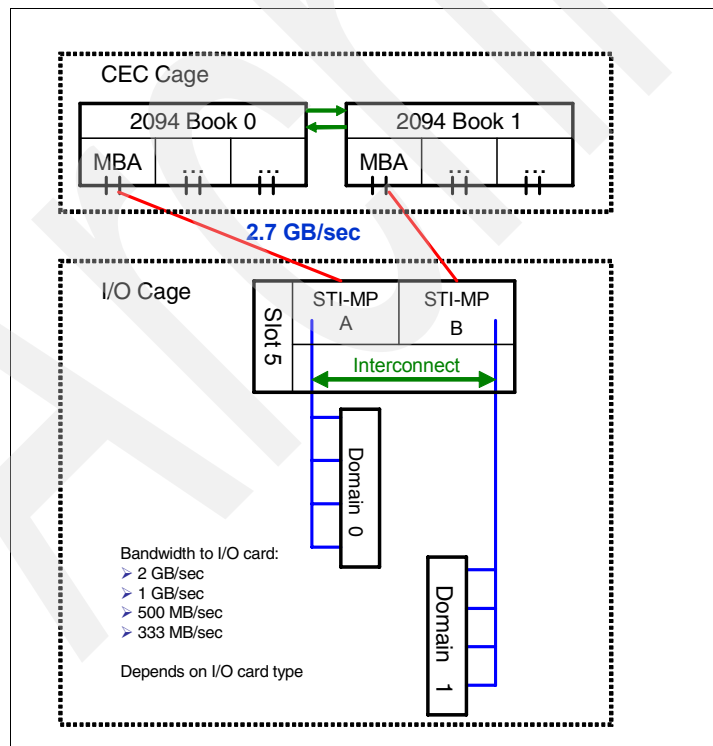


Figure 2-7 Redundant I/O Interconnect concept

Normally, book 0 MBA/STI connects to the STI-MP (A) card and services domain 0 I/O connections (slots 01, 03, 06, and 08). In the same fashion, book 1 MBA/STI connects to the STI-MP (B) card and services domain 1 (slots 02, 04, 07, and 09). If book 1 is removed, or the connections from book 1 to the cage are removed, connectivity to domain 1 is maintained by guiding the I/O to domain 1 through the interconnect between STI-MP (A) and STI-MP (B).

In configuration reports, books are numbered 0, 1, 2, and 3, MBAs are numbered from D1 to D8, and the STIs are identified as jacks numbered J00 and J01 for each MBA.

### **Enhanced Book Availability**

With Enhanced Book Availability, the impact of book replacement is minimized. In a multi-book system, a single book can be concurrently removed and reinstalled for an upgrade or repair. To be able to do so requires sufficient resources in the remaining books to remove a book without impacting the workload. CPs and memory from the book must be relocated before the book can be removed. Not only additional PUs need to be available on the remaining books to replace the deactivated book, but also sufficient redundant memory must be available if it is required that no degradation of applications is allowed. The “Flexible Memory option” on page 35, should be considered to ensure that the server configuration supports removal of a book with minimal impact to the workload. Any book can be replaced, including book 0, that initially contains the HSA.

Removal of a book also removes the book connectivity to the I/O cages. The impact of the removal of the book on the system is limited by the use of Redundant I/O Interconnect, as described in “Redundant I/O Interconnect” on page 38; however, all ICBs on the removed book have to be configured offline.

When Enhanced Book Availability and the Flexible Memory option are *not* used and when a book must be replaced, for example, due to an upgrade or a repair action, the memory in the failing book is removed as well. Until the removed book is replaced, a Power-on Reset of the server with the remaining books is supported.

### **Book upgrade**

All MBA fanout cards used for I/O are concurrently rebalanced as part of the book addition. However, to have MBA fanouts used for ICBs rebalanced, the STI rebalance feature (FC 2400) needs to be ordered. The STI rebalance feature is disruptive, but is highly recommended when upgrading from a S08 to a larger model. When going from a multi-book to a larger model, individual evaluation is also recommended. See “STI Rebalance feature” on page 96 for more information.

## **2.1.6 Frames and cages**

The z9 EC frames are enclosures built to Electronic Industry Association (EIA) standards. The server always has two frames that are composed of two 40 EIA frames. The A and Z frames are bolted together and have two cage positions (top and bottom).

- ▶ Frame A has the CEC cage at the top and I/O cage 1 at the bottom.
- ▶ Frame Z can be one of the following configurations:
  - Without I/O cage
  - With one I/O cage, I/O cage 2 at the bottom
  - With two I/O cages, I/O cage 2 at the bottom and I/O cage 3 on top

All books, the DCAs for the books, and the cooling components are located in the CEC cage in the top half of the A frame of the z9 EC. In Figure 2-8, the arrows point to the front view of the CEC cage in which four books are shown as being installed.

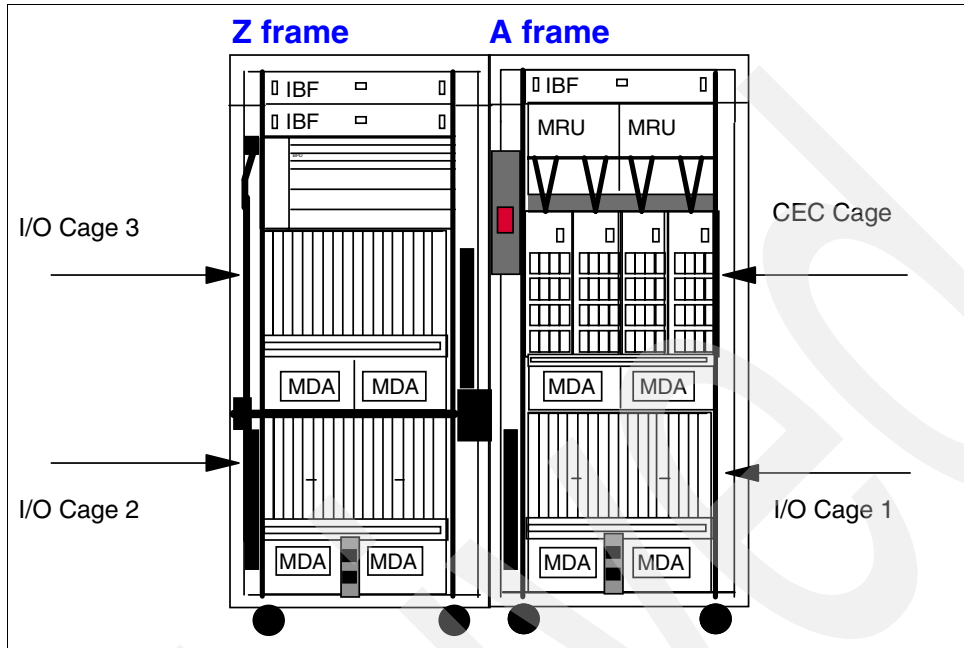


Figure 2-8 CEC cage and I/O cage locations

### A frame

As shown in Figure 2-8, the main components in the A frame are:

- ▶ Two optional Internal Battery Features (IBFs). The optional Internal Battery Feature provides the function of a local uninterrupted power source.

The IBF further enhances the robustness of the power design, increasing Power Line Disturbance immunity. It provides battery power to preserve processor data in case of a loss of power on both of the AC feeders from the utility company. The IBF can hold power briefly over a *brownout*, or for orderly shutdown in case of a longer outage. The IBF provides up to 13 minutes of full power, depending on I/O configuration.

Table 2-4 IBF estimated power time

z9 EC model	I/O configuration		
	One I/O cage	Two I/O cages	Three I/O cages
S08	7 minutes	11 minutes	7.3 minutes
S18	12 minutes	7.5 minutes	10 minutes
S28	8.5 minutes	11 minutes	8 minutes
S38	13 minutes	9.5 minutes	7.3 minutes
S54	13 minutes	9.5 minutes	7.3 minutes

- ▶ One or two Modular Refrigeration Units (MRUs) that are air-cooled by their own internal cooling fans.
- ▶ The CEC cage, containing up to four books, each with two insulated refrigeration lines to an MRU.

- ▶ I/O cages can house all supported types of channel cards. An I/O cage accommodates up to 420 ESCON channels or up to 112 FICON Express4 channels in the absence of any other card. Up to three I/O cages are supported.
- ▶ Air-moving devices (AMD) providing N+1 cooling for the MBAs, memory, and DCAs.

### **Z frame**

As shown in Figure 2-8 on page 40, the main components in the Z-frame are:

- ▶ Two optional Internal Battery Features (IBFs).
- ▶ The Bulk Power Assemblies (BPAs).
- ▶ I/O cage 2 (bottom) and I/O cage 3 (top). Note that both I/O cages are the same as the one in the A frame, and can house all supported types of channel cards.  
  
The Z frame can hold only the bottom cage (I/O cage 2), or both the bottom and top I/O cages (I/O cage 2 and I/O cage 3).
- ▶ The Support Element (SE) tray, located in front of I/O cage 2, contains the two SEs (not shown in Figure 2-8 on page 40).

### **I/O cages**

There are up to 16 STI buses per book to transfer data, with a bi-directional bandwidth of 2.7 GB per second each. An STI is driven off an MBA fanout card. There are eight MBA fanout cards per book, each driving two STIs, providing an aggregated bandwidth of 43.2 GB per second per book.

The STIs connect to I/O cages that may contain a variety of channel, Coupling Link, OSA-Express, and Cryptographic feature cards:

- ▶ ESCON channels (16 port cards, 15 usable ports and one spare).
- ▶ FICON channels (FICON or FCP modes).
  - FICON Express channels (two port cards) - carried forward during an upgrade only
  - FICON Express2 channels (four port cards) - carried forward during an upgrade only
  - FICON Express4 channels (four port cards)
- ▶ ISC-3 links (up to four Coupling Links, two links per daughter card). Two daughter cards (ISC-D) plug into one mother card (ISC-M).
- ▶ ICB-4 channels do not require a slot in the I/O cage and attach directly to the STI of the communicating server with a bandwidth of 2.0 GBps.
- ▶ ICB-3 channels require an STI-3 extender card in the I/O cage. The STI-3 extender card provides two output ports to support the ICB-3 links. The STI-3 card converts the 2 GBps input from the MBA fanout into two 1 GBps ICB-3 links.
- ▶ OSA-Express channels:
  - OSA-Express Gb Ethernet - carried forward during an upgrade only
  - OSA-Express2 Gb Ethernet
  - OSA-Express2 10 Gb Ethernet LR
  - OSA-Express 1000BASE-T Ethernet - carried forward during an upgrade only
  - OSA-Express2 1000BASE-T Ethernet
- ▶ Crypto Express2, with two PCI-X adapters per feature. A PCI-X adapter can be configured as a cryptographic coprocessor for secure key operations or as accelerator for clear key operations.

## 2.1.7 The MCM

The z9 EC MultiChip Module (MCM) contains 16 chips: Eight are processor chips (supporting 12 or 16 PUs), four are System Data cache (SD) chips, one is the Storage Control (SC) chip, two chips carry the Memory Subsystem Control function (MSC), and there is one chip for the clock (CLK) and ETR receiver function.

The 95 x 95 mm glass ceramic substrate on which the chips are mounted has 102 interconnect layers in the substrate and 545 meters of internal wiring. The total number of transistors on all chips on the MCM amounts to more than 4.5 billion.

The MCM plugs into a card that is part of the book packaging, as shown in Figure 2-9. The book itself is plugged into the processor board to provide inter connectivity between the books, so that a multibook system appears as a Symmetric Multi Processor (SMP). The MCM is connected to its environment by 5184 Land Grid Arrays (LGA) connectors. Figure 2-10 on page 43 shows the chip locations.

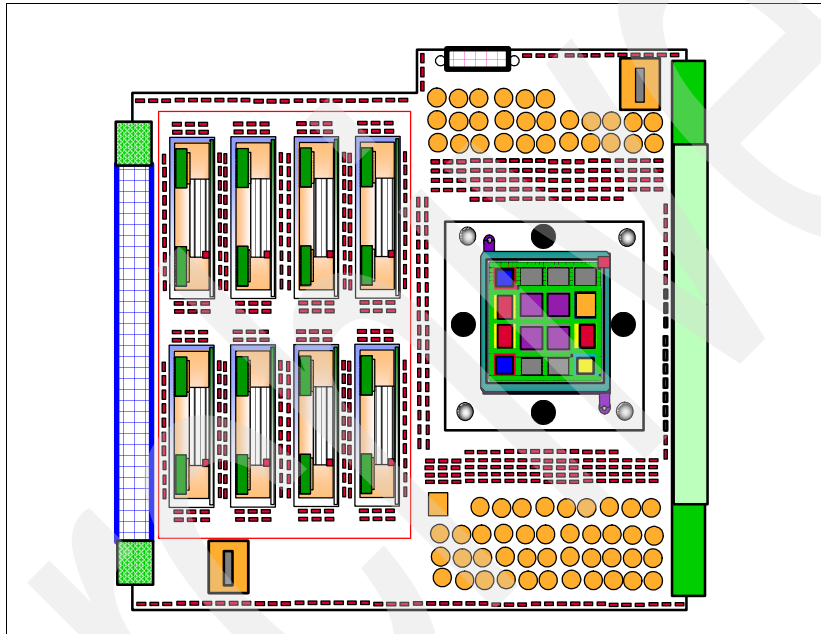


Figure 2-9 MCM card

## 2.1.8 The PU, SC, SD, and MSC chips

All chips use CMOS 10S chip technology, except for the clock chip (CMOS 8S). CMOS 10S is state-of-the-art microprocessor technology based on ten-layer Copper Interconnections and Silicon-On Insulator technologies. The chip lithography line width is 0.125 micron.

The eight PU chips come in two versions. For the z9 EC models S08, S18, S28, and S38, the Processor Units (PUs) on the MCM in each book are implemented with a mix of single-core and dual-core PU chips. Four single-core and four dual-core chips are used, resulting in 12 PUs per MCM.

For the z9 EC Model S54, the Processor Units (PUs) on the MCMs in all books are implemented with eight dual-core PU chips, resulting in 16 PUs per MCM.



Eight to ten PUs of the 12 PU version may be characterized for customer use. The two standard SAPs are initially allocated to the dual-core processor chips. Optionally, up to two spare chips may be allocated on an MCM. System-wide, two spare chips are available that may be allocated on any MCM of the system. Each core on the chip runs at a cycle time of 0.58 nanoseconds. Each dual-core PU chip measures 15.78 x 11.84 mm and has 121 million transistors.

In the 16 PU version, used in all MCMs of a z9 EC Model S54, 12 to 14 PUs are available for customer use. Optionally, up to two spare chips may be allocated on an MCM. System-wide, two spare chips are available that may be allocated on any MCM of the system.

Each PU has a 512 KB on chip Level 1 cache (L1) that is split into a 256 KB L1 cache for instructions and a 256 KB L1 cache for data, providing large bandwidth.

Figure 2-10 shows the MCM chip layout.

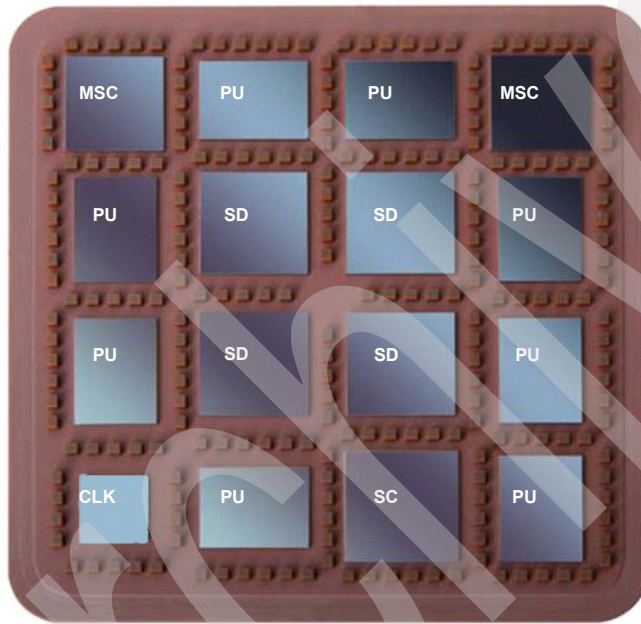


Figure 2-10 MCM chip layout

### SC chip

The L1 caches communicate with the L2 caches (SD chips) by two bi-directional 16-byte data buses. There is a 2:1 bus/clock ratio between the L2 cache and the PU, controlled by the Storage Controller (SC chip), that also acts as an L2 cache cross-point switch for L2-to-L2 ring traffic, L2-to-MSB traffic, and L2-to-MBA traffic. The L1-to-L2 interface is shared by two PU cores on a dual core PU chip. The SC chip measures 16.41 x 16.41 mm and has 162 million transistors.

### SD chip

The Level 2 cache (L2) is implemented on the four System Data (SD) cache chips, each with a capacity of 10 MB, providing a cache size of 40 MB. These chips measure 15.66 x 15.40 mm and carry 660 million transistors, making it one of the world's densest chips.

## MSC chip

Two Memory Storage Control (MSC) chips police traffic between memory (PMAs) and the Level 2 cache. The MSC chips measures 14.31 x 14.31 mm and each have 24 million transistors.

The dual-core PU chips share the path to the MSC chip (L2 control) and the clock chip (CLK).

## 2.1.9 Summary

Table 2-5 summarizes all aspects of the z9 EC system structure.

Table 2-5 System structure summary

	<b>z9 EC Model S08</b>	<b>z9 EC Model S18</b>	<b>z9 EC Model S28</b>	<b>z9 EC Model S38</b>	<b>z9 EC Model S54</b>
Number of MCMs	1	2	3	4	4
Total number of PUs	12	24	36	48	64
Maximum number of characterized PUs	8	18	28	38	54
Number of CPs	0–8	0–18	0–28	0–38	0– 54
Number of IFLs	0–8	0–18	0–28	0–38	0–54
Number of ICFs	0–8	0–16	0–16	0–16	0–16
Number of zAAPs	0–4	0–9	0–14	0–19	0– 27
Number of zIIPs	0–4	0–9	0–14	0–19	0– 27
Standard SAPs	2	4	6	8	8
Additional SAPs	0–5	0–13	0–21	0–24	0–24
Standard spare PUs	2	2	2	2	2
Number of memory cards	8	16	24	32	32
Enabled Memory Size (multiples of 8 GB)	16–128 GB	16–256 GB	16–384 GB	16–512 GB	16–512 GB
L1 Cache per PU	256/256 KB	256/256 KB	256/256 KB	256/256 KB	256/256 KB
L2 Cache	40 MB	80 MB	120 MB	160 MB	160 MB
Cycle time (ns)	0.58	0.58	0.58	0.58	0.58
Maximum number of STIs	16	32	48	64	64
STI bandwidth	2.7 GB per second	2.7 GB per second	2.7 GB per second	2.7 GB per second	2.7 GB per second
Maximum number of I/O cages	3	3	3	3	3
Number of Support Elements	2	2	2	2	2
External power	3 phase	3 phase	3 phase	3 phase	3 phase

	<b>z9 EC Model S08</b>	<b>z9 EC Model S18</b>	<b>z9 EC Model S28</b>	<b>z9 EC Model S38</b>	<b>z9 EC Model S54</b>
Internal Battery Feature	Optional	Optional	Optional	Optional	Optional

## 2.2 System design

The design of the z9 EC Symmetrical Multi Processor (SMP) is the next step in an evolutionary trajectory stemming from the introduction of CMOS technology back in 1994. Over time, the design has been adapted to the changing requirements dictated by the shift towards e-business applications that customers are becoming more and more dependent on.

The z9 EC offers very high levels of availability, reliability, resilience, and security, and fits in IBM strategy in which mainframes play a central role in realizing an intelligent integrated infrastructure. The z9 EC is designed using a holistic approach. This means not only the server is considered important for the infrastructure, but also everything around it in terms of operating systems, middleware, storage, security, and network technologies supporting open standards, all to help customers achieve their business goals.

The modular book design of the z9 EC aims to reduce planned and unplanned outages by offering concurrent repair, replace, and upgrade functions for processors, memory, and I/O. The z9 EC with its superscalar processor and flexible configuration options is the next implementation to address the ever-changing IT environment.

### 2.2.1 Design highlights

The physical packaging is comparable to the packaging used for z990 systems. Its modular book design creates the opportunity to address the ever-increasing costs related to building systems with ever-increasing capacities. The modular book design is flexible and expandable and may contain even larger capacities in the future.

The main objectives of the z9 EC system design, which are covered in this and subsequent chapters, are:

- ▶ To offer a *flexible infrastructure* to concurrently accommodate a wide range of operating systems and applications, from the traditional systems to the world of Linux and e-business.
- ▶ To have state-of-the-art *integration* capability for server consolidation, offering virtualization techniques, such as:
  - Logical partitioning, which allows up to 60 logical servers
  - z/VM, which can virtualize hundreds of servers as Virtual Machines
  - HiperSockets, which implements virtual LANs between logical and virtual servers within a z9 EC server

This allows logical and virtual server coexistence and maximizes system utilization by sharing hardware resources.

- ▶ To have *high performance* to achieve the outstanding response times required by e-business applications, based on z9 EC superscalar processor technology, architecture, and high bandwidth channels, which offer high data rate connectivity.
- ▶ To offer the *high capacity* and *scalability* required by the most demanding applications, both from single system and clustered systems points of view.

- ▶ To have the capability of *concurrent upgrades* for processors, memory, and I/O connectivity, avoiding server outages in planned situations.
- ▶ To implement a system with *high availability* and *reliability*, from the redundancy of critical elements and sparing components of a single system, to the clustering technology of the Parallel Sysplex environment.
- ▶ To have a broad *connectivity* offering, supporting open standards such as Gigabit Ethernet (GbE) and Fibre Channel Protocol (FCP) for Small Computer System Interface (SCSI).
- ▶ To provide the highest level of *security*, each CP has a CP Assist for Cryptographic Function (CPACF). Optional Crypto Express2 features with Cryptographic Coprocessors and Cryptographic Accelerators for Secure Sockets Layer (SSL) transactions of e-business applications can be added.
- ▶ To be *self-managing*, adjusting itself on workload changes to achieve the best system throughput, through the Intelligent Resource Director and the Workload Manager functions.
- ▶ To have a *balanced system* design, providing large data rate bandwidths for high performance connectivity along with processor and system capacity.

The following sections describe the z9 EC system structure, showing a logical representation of the data flow from PUs, L2 cache, memory cards, and MBAs, which connect to the I/O cage through Self-Timed Interconnects (STI).

## 2.2.2 Book design

A book has 12 or 16 PUs, four or eight memory cards, and up to 16 STIs organized on eight MBA/STI fanout cards, coordinated by the System Controller (SC). Each memory card has a capacity of 4 GB, 8 GB, or 16 GB, resulting in up to 128 GB of memory Level 3 (L3) per book. A four-book z9 EC can have up to 512 GB memory. The Storage Controller, shown as SCC Cntl in Figure 2-11 on page 47, acts as a cross-point switch between Processor Units (PUs), Memory Controllers (MSCs), and Memory Bus Adapters (MBAs).

The SD chips, shown as SCD in Figure 2-11 on page 47, also incorporate a Memory Coherent Controller (MCC) function.

Each PU chip has its own 512 KB Cache Level 1 (L1), split into 256 KB for data and 256 KB for instructions. The L1 cache is designed as a store-through cache, meaning that altered data is also stored to the next level of memory (L2 cache). The z9 EC models S08, S18, S28, S38, and S54 use the CMOS 10KS-SOI PU chips running at 0.58 ns.

The MCC function controls a large 40 MB L2 cache, and is responsible for the interbook communication in a ring topology connecting up to four books through two concentric loops, called the ring structure. The MCC function optimizes cache traffic and will not look for cache hits in other books when it knows that all resources of a given logical partition are available in the same book.

The L2 cache is the aggregate of all cache space on the SD chips, resulting in a 40 MB L2 cache per book. The SC chip (SCC) controls the access and storing of data in the four SD chips. The L2 cache is shared by all PUs within a book and shared across books through the ring topology, providing the communication between L2 caches across books in systems with more than one book installed; the L2 has a store-in buffer design.

The interface between the L2 cache and processor memory (L3) is accomplished by four high-speed memory buses and controlled by the memory controllers (MSC). Storage access is interleaved between the storage cards, which tends to equalize storage activity across the

cards. Each PMA has two ports that each have a maximum bandwidth of 8 GB per second. Each port contains a control and a data bus, in order to further reduce any contention by separating the address and command from the data bus.

The memory cards support store protect key caches to match the key access bandwidth with that of the memory bandwidth.

The logical book structure is shown in Figure 2-11.

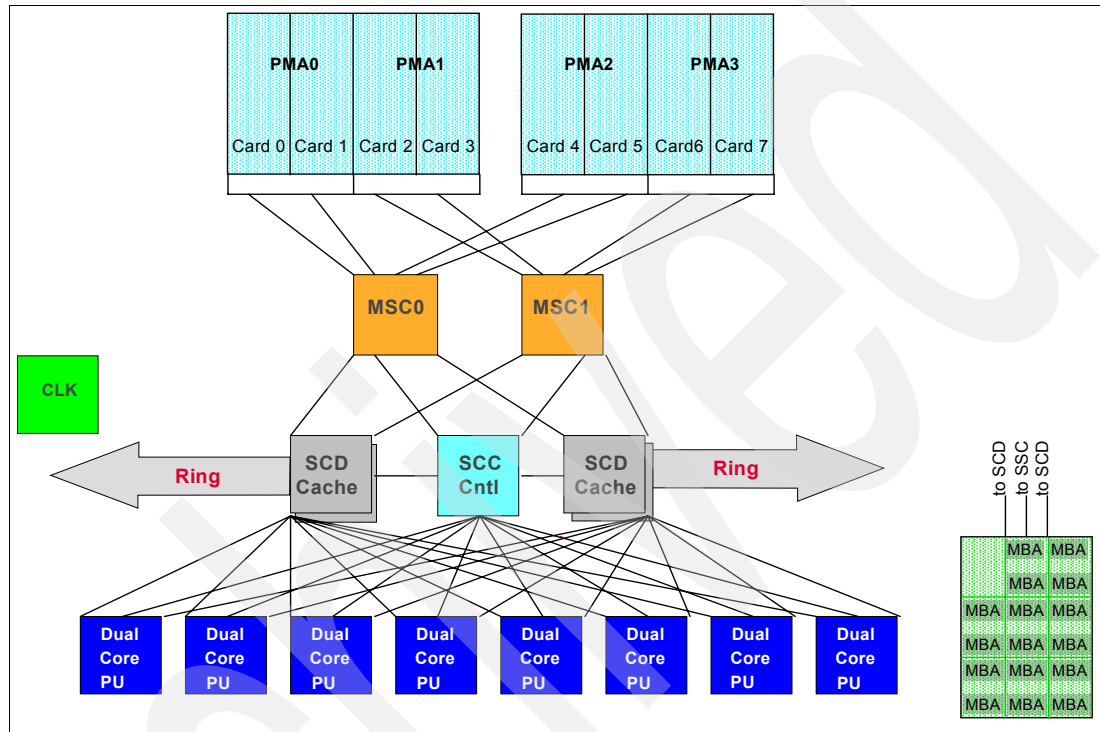


Figure 2-11 Logical book structure

There are up to 16 STI buses per book to transfer data and each STI has a bidirectional bandwidth of 2.7 GB per second. A four-book z9 EC server may have up to 64 STIs.

An STI is an interface from the Memory Bus Adapter (MBA) to:

- ▶ An STI-MP card in an I/O cage, to connect to:
  - ESCON channels (16 port cards).
  - FICON-Express (two port cards), FICON Express2 (four port cards), or FICON Express4 (four port cards), used in FICON or FCP modes.
  - OSA-Express channels:
    - OSA-Express Gb Ethernet
    - OSA-Express2 Gb Ethernet
    - OSA-Express2 10 Gb Ethernet LR
    - OSA-Express 1000BASE-T Ethernet
    - OSA-Express2 1000BASE-T Ethernet
  - ISC-3 links, up to four Coupling Links with two links per daughter card (ISC-D). Two daughter cards plug into one mother card (ISC-M).

- Crypto Express2, minimum two features, each containing two PCI-X cryptographic adapters to be configured as coprocessors or accelerators.
- ▶ An STI-3 extender card in an I/O cage, connecting to ICB-3 channels in z990, z890, z900 or z800. The STI-3 card takes a 2 GBps STI directly from the MBA fanout and provides two 1 GBps ICB-3 links.
- ▶ An ICB-4, directly attached to the STI interface between z9 EC, z9 BC, z990 or z890. The ICB-4 runs at 2.0 GB per second.

Data transfer between processor memory and attached I/O devices or servers is done through the Memory Bus Adapter. The physical path includes the channel card (except for STI connected servers), the Self-Timed Interconnect bus, and possibly an STI extender card, the Storage Control, and the Storage Data chips.

More detailed information about I/O connectivity and each channel type can be found in Chapter 3, “I/O system structure” on page 89.

### Dual External Time Reference

The optional ETR connections, although not part of the book design, are found adjacent to the books on the opposite side of the processor board. The z9 EC implements an Enhanced ETR Attachment Facility (EEAF) designed to provide a dual External Time Reference (ETR) attachment facility. Two ETR cards are automatically shipped when Coupling Links are ordered and provide a dual path interface to the IBM Sysplex Timers, which are used for timing synchronization between systems in a Sysplex environment. This allows continued operation even if a single ETR card fails. This redundant design also allows concurrent maintenance.

The External Time Reference connections will be replaced over time by the implementation of Server Time Protocol (STP), which makes use of coupling links to pass timing messages to the servers. Transition to STP makes it possible to have a Mixed Coordinated Network configuration. The Sysplex Timer provides the timekeeping information in a Mixed CTN. Once an STP-only configuration is established, the ETR connections are no longer needed.

**Note:** Server Time Protocol (STP) is available as a charged feature (F/C 1021) that is implemented in the Licensed Internal Code of the IBM System z9 servers, and is designed for multiple servers to maintain time synchronization with each other. Refer to *Server Time Protocol Planning Guide*, SG24-7280, and *Server Time Protocol Implementation Guide*, SG24-7281 for detailed information.

### Oscillator

The z9 EC has two oscillator cards, a primary and a backup. If the primary fails, the secondary detects the failure, takes over transparently, and continues to provide the clock signal to the server.

## 2.2.3 Processor Unit design

Each Processor Unit (PU) is optimized to meet the demands of e-business workloads, without compromising the performance characteristics of traditional workloads. The PUs in the z9 EC have a superscalar design.

## Superscalar processor

A scalar processor is a processor that is based on a single issue architecture, which means that only a single instruction is executed at a time. A superscalar processor allows concurrent execution of instructions by adding additional resources onto the microprocessor to achieve more parallelism by creating multiple pipelines, each working on their own set of instructions.

A superscalar processor is based on a multi-issue architecture. In such a processor, where multiple instructions can be executed at each cycle, a higher level of complexity is reached, because an operation in one pipeline may depend on data in another pipeline. A superscalar design therefore demands careful consideration of which instruction sequences can successfully operate in a multi-pipeline environment.

As an example, consider the following: If the branch prediction logic of the microprocessor makes the wrong prediction, it might be necessary to remove all instructions in the parallel pipelines also (refer to “Processor Branch History Table (BHT)” on page 51 for more details).

There are challenges in creating an efficient superscalar processor. The superscalar design of the z9 EC PU has made big strides in avoiding address generation interlock situations. Instructions requiring information from memory locations may suffer multi-cycle delays to get the memory content. The superscalar design of the z9 EC PU tries to overcome these delays by continuing to execute (single cycle) instructions that do not cause delays. The technique used is called *out-of-order operand fetching*. This means that some instructions in the instruction stream are already underway, while earlier instructions in the instruction stream that cause delays due to storage references take longer. Eventually, the delayed instructions catch up with the already fetched instructions and all are executed in the designated order. The z9 EC PU gets much of its superscalar performance benefits from avoiding address generation interlocks.

It is not only the processor that contributes to the capability of the successful execution of instructions in parallel. Given a superscalar design, compilers and interpreters must create code that benefits optimally from the particular superscalar processor implementation. The C++ compiler and Java Virtual Machine for z/OS exploit the z9 EC microprocessor superscalar implementation. The intent is to improve the performance advantage for e-business workloads, such as WebSphere and Java applications.

In order to create instruction sequences that are least affected by interlock situations, instruction grouping rules are enforced to create instruction streams that benefit most from the superscalar processor. It is expected that e-business workloads will primarily benefit from this design since they tend to use more computational instructions.

A WebSphere Application Server workload environment that runs a mix of Java and DB2 code will greatly benefit from the superscalar processor design of the z9 EC. Measurements show additional improvement for these types of workloads, on top of the improvements attributed to the cycle time decrease from 0.83 ns on a z990 model to 0.58 ns on a z9 EC.

The superscalar design of the z9 EC microprocessor means that some instructions are processed immediately and that processing steps of other instructions may occur out of the normal sequential order, called *pipelining*. The superscalar design of the z9 EC offers:

- ▶ Decoding of two instructions per cycle
- ▶ Execution of three instructions per cycle (given that the oldest instruction is a branch)
- ▶ In-order execution
- ▶ Out-of-order operand fetching

Other features of the microprocessor, aimed at improving the performance of the emerging e-business application environment, are:

- ▶ Floating point performance for IEEE Binary Floating Point arithmetic is improved to assist further exploitation of Java application environments.
- ▶ A secondary cache for Dynamic Address Translation, called the Secondary level Translation Look aside Buffer (TLB), is provided for both L2 instruction and data caches, increasing the number of buffer entries by a factor of eight.
- ▶ The CP Assist for Cryptographic Function (CPACF) accelerates the encrypting and decrypting of SSL transactions and VPN encrypted data transfers. The assist function uses a special instruction set for symmetrical clear key cryptographic encryption and encryption operations.

### Asymmetric mirroring for error detection

Each PU in the z9 EC servers uses mirrored instruction execution as a simple error detection mechanism. The mirroring is dependent on a dual instruction processor design with dual I-units, E-units, and floating point function. It is asymmetric because the mirrored execution is delayed from the actual operation. The benefit of the asymmetric design is that the mirrored units do not have to be closely located to the units where the actual operation takes place, thus allowing for optimization for performance (see Figure 2-12).

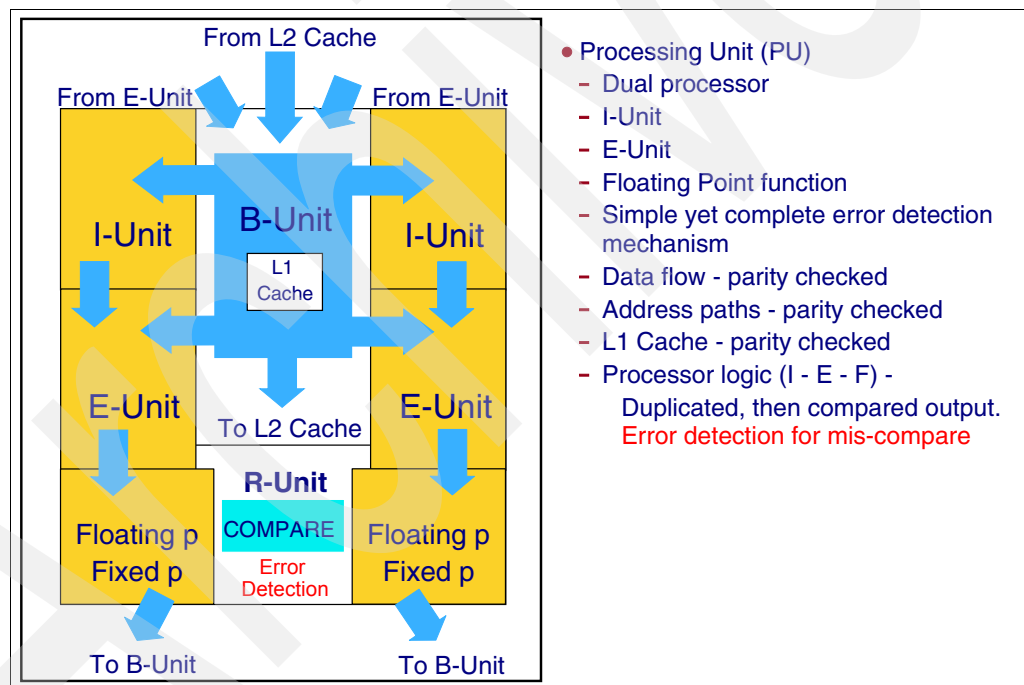


Figure 2-12 Dual (asymmetric) processor design

Each PU has a dual processor and each processor has its own Instruction Unit (I-Unit) and Execution Unit (E-Unit), which includes the floating point function. The instructions are executed asymmetrically (not exactly in parallel) on each processor and compared after processing.

This design simplifies error detection during instruction execution, saving additional circuits and extra logic required to do this checking. The z9 EC servers also contain error-checking circuits for data flow parity checking, address path parity checking, and L1 cache parity checking.



## Compression Unit on a chip

Each z9 EC PU has a Compression Unit on the chip, providing excellent hardware compression performance. The Compression Unit is integrated with the CP Assist for Cryptographic Function, benefiting from combining the use of buffers and interfaces.

## CP Assist for Cryptographic Function

Each PU has a CP Assist for Cryptographic Function (CPACF) on the chip. The assist provides high performance hardware encrypting and decrypting support for clear key operations. Five special instructions are used with the cryptographic assist function.

CPACF offers a set of symmetric cryptographic functions for high encrypting and decrypting performance of clear key operations for SSL, VPN, and data storing applications that do not require FIPS 140-2 level 4 security. The cryptographic architecture includes support for:

- ▶ Data Encryption Standard (DES) data encryption and decrypting
- ▶ Triple Data Encryption Standard (TDES) data encryption and decrypting
- ▶ Advanced Encryption Standard (AES) for 128-bit keys
- ▶ Pseudo Random Number Generation (PRNG)
- ▶ MAC message authorization
- ▶ Secure Hash Algorithm (SHA-1) hashing
- ▶ Secure Hash Algorithm (SHA-256) hashing

The CPACF complements public key (RSA) functions and the secure cryptographic operations provided by the Crypto Express2 feature. See Chapter 5, “Cryptography” on page 149, for more information about the cryptographic features on the z9 EC.

## Processor Branch History Table (BHT)

The Branch History Table implementation on processors has a key performance improvement effect. The BHT was originally introduced on the IBM ES/9000® 9021 in 1990 and has been improved ever since.

The z9 EC server BHT offers significant branch performance benefits. The BHT allows each CP to take instruction branches based on a stored BHT, which improves processing times for calculation routines. Using a 100-iteration calculation routine as an example (Figure 2-13), the hardware pre-processes the branch incorrectly 99 times without a BHT. With a BHT, it pre-processes branch correctly 98 times out of 100.

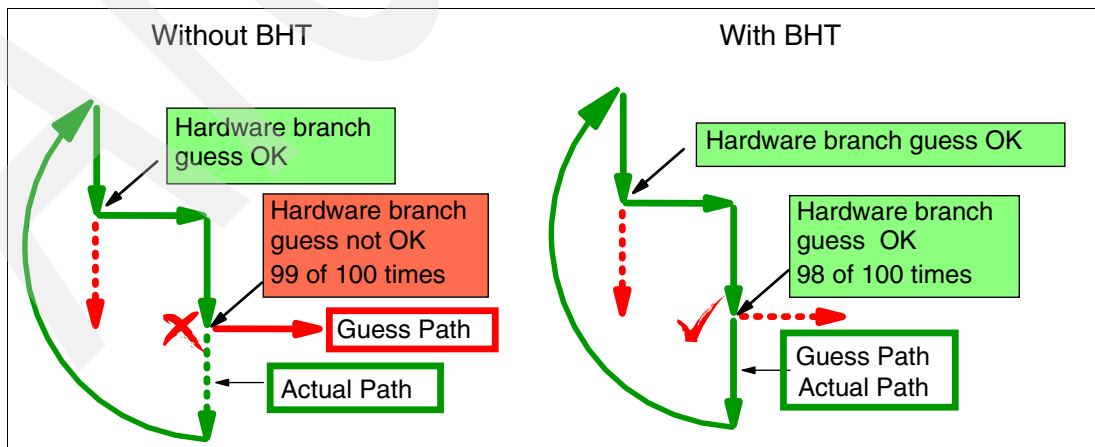


Figure 2-13 Branch History Table (BHT)

- ▶ Without BHT, the processor:
  - Makes an incorrect branch guess the first time through the loop (at the second branch point in Figure 2-13 on page 51).
  - Preprocesses instructions for the guessed branch path.
  - Starts preprocessing a new path if the branch is not equal to the guess.
  - Repeats this 98 more times until the last time, when the guess matches the actual branch taken.
- ▶ With BHT, the processor:
  - Makes an incorrect branch guess the first time through the loop (at the second branch point in Figure 2-13 on page 51).
  - Preprocesses instructions for the guessed branch path.
  - Starts preprocessing a new path if the branch is not equal to the guess.
  - Updates the BHT to indicate the last branch action taken at this address.
  - The next 98 times, the branch path comes from the BHT.
  - The last time, the guess is wrong.

The key point is that, with the BHT, the table is updated to indicate the last branch action taken at branch addresses. Using the BHT, if a hardware branch at an address matches a BHT entry, the branch direction is taken from the BHT. Therefore, in the diagram, the branches are correct for the remainder of the loop through the program routine, except for the last one.

The success rate of the BHT design contributes a great deal to the superscalar aspects of the z9 EC, given the fact that the architecture rules prescribe that for successful parallel execution of an instruction stream, the correctly predicted result of the branch is essential.

### **Wild branch**

In instances where a bad pointer is used or when code overlays a data area containing a pointer to some code, this results in a random branch causing a 0C1 or 0C4 abend. Random branches are very hard to diagnose since there is no clue about how the system got there.

With the wild branch hardware facility of the z9 EC, the last address from which a successful branch instruction was executed is kept. z/OSV1.7 uses this information in conjunction with debugging aids, like the SLIP command, to determine where wild branch came from and may collect data from that storage location. This will decrease the many debugging steps to go through in finding out where the branch came from.

### **IEEE Floating Point**

Over 130 binary and hexadecimal floating-point instructions are present in System z9. They incorporate IEEE Standards into the platform.

The key point is that Java and C/C++ applications tend to use IEEE Binary Floating Point operations more frequently than earlier applications. This means that the better the hardware implementation of this set of instructions, the better the performance of e-business applications will be.

## Hardware Decimal Floating Point

Base 10 arithmetic is used for most business and financial computation. Floating point computation used for work typically done in decimal arithmetic has involved frequent necessary data conversions and approximation to represent decimal numbers. This has made floating point arithmetic complex and error prone for programmers using it in applications where the data is typically decimal data.

Hardware decimal floating point computational instructions provide 4, 8, and 16 byte data formats, an encoded decimal (base 10) representation for data, instructions for performing decimal floating point computations, and an instruction that performs data conversions to and from the decimal floating point representation. Instructions are added in support of the Draft Standard for Floating-Point Arithmetic P754, which is intended to supersede the ANSI/IEEE Std 754-1985.

## Translation Look aside Buffer

The Translation Look aside Buffer (TLB) in the Instruction and Data L1 caches have a secondary TLB to enhance performance. In addition, a translator unit is added to translate misses in the secondary TLB.

## Instruction fetching and instruction decode

The superscalar design of the z9 EC microprocessor allows for the decoding of up to two instructions per cycle and the execution of three instructions per cycle. Execution takes place in order, but storage accesses for instruction and operand fetching may occur out of sequence.

### *Instruction fetching*

Instruction fetch in pre-z990 models tries to get as far ahead of instruction decode and execution as possible because of the relatively large instruction buffers available. In the z9 EC (and also z990) microprocessor, smaller instruction buffers are used. The operation code is fetched from the I-cache and put in instruction buffers that hold pre-fetched data awaiting decode.

### *Instruction decoding*

The processor can decode one or two instructions per cycle. The result of the decoding process is queued and subsequently used to form a group.

## Instruction grouping

From the instruction queue, one simple branch instruction and up to two general instructions can be issued every cycle. The instructions are taken from the instruction queue and grouped together. The instructions are assembled according to instruction grouping rules. A complete description of the rules is beyond the scope of this book.

It is the compiler's responsibility to select instructions that best fit with the z9 EC superscalar microprocessor and abide by the grouping rules to create code that best exploits the superscalar implementation.

## Extended Translation Facility

Instructions have been added to the z/Architecture instruction set in support of the Extended Translation Facility adds instructions. They are used in data conversion operations for data encoded in Unicode, making applications enabled for Unicode or globalization more efficient. These data encoding formats are used in Web Services, Grid, and on demand environments where XML and SOAP technologies are used. The High Level Assembler supports the Extended Translation Facility instructions.

## Instruction set extensions

A large number of instructions added to the z/Architecture instruction set are introduced in z9 EC. The added instruction supports functions like the ones listed here:

- ▶ Hexadecimal floating point instructions for various un-normalized multiply instructions and multiply and add instructions.
- ▶ Immediate instructions, including various add, compare, exclusive OR, OR, subtract, load, and insert formats. Use of these instructions improves performance and reduces the need for a database register.
- ▶ Load instructions for handling unsigned half words (such as those used for Unicode).
- ▶ Cryptographic instructions have been extended with AES, SHA-256, and random number generation functions.
- ▶ Extended Translate Facility-3 instructions have been enhanced to conform with the current Unicode 4.0 standard.
- ▶ Assist instructions to help eliminate hypervisor overhead.

## 2.2.4 Processor Unit functions

One of the key components of the z9 EC server is the Processor Unit (PU). This is the microprocessor chip where instructions are executed and the related data resides. The instructions and the data are stored in the PUs high-speed buffer, called the Level 1 cache. Each PU has its own 512 KB Level 1 cache, split into 256 KB for data and 256 KB for instructions.

The L1 cache is designed as a store-through cache, which means that altered data is synchronously stored into the next level, the L2 cache. Each PU has multiple processors inside and instructions are executed twice, asynchronously, on both processors.

This asymmetric mirroring of instruction execution runs one cycle behind the actual operation. This allows the circuitry on the chip to be optimized for performance and does not compromise the simplified error detection process that is inherent to a mirrored execution unit design.

All PUs in a z9 EC server are physically identical. When the system is initialized, PUs can be characterized to specific functions: CP, IFL, ICF, zAAP, zIIP, or SAP. The function assigned to a PU is set by the Licensed Internal Code loaded when the system is initialized (Power-on Reset) and the PU is *characterized*. Only characterized PUs have a designated function; non-characterized PUs are considered spares.

This design brings outstanding flexibility to the z9 EC server, as any PU can assume any available characterization. This also plays an essential role in system availability, because PU characterization can be done dynamically, with no server outage, allowing the actions discussed in the following sections.

### Concurrent upgrades

Except on a fully configured model, concurrent upgrades can be done by the Licensed Internal Code, which assigns a PU function to a previously non-characterized PU. Within the book boundary or boundary of multiple books, no hardware changes are required, and the upgrade can be done through Capacity Upgrade on Demand (CUoD), Customer Initiated Upgrade (CIU), On/Off Capacity on Demand (On/Off CoD), or Capacity BackUp (CUB). More information about capacity upgrades is provided in Chapter 8, “Concurrent upgrades and availability” on page 205.

## PU sparing

In the rare event of a PU failure, the failed PU's characterization is dynamically and transparently reassigned to a spare PU. More information about PU sparing is provided in "Sparing rules" on page 67.

A minimum of one PU per z9 EC server must be ordered as one of the following:

- ▶ A Central Processor (CP)
- ▶ An Integrated Facility for Linux (IFL)
- ▶ An Internal Coupling Facility (ICF)

The number of CPs, IFLs, ICFs, zAAPs, zIIPs, or SAPs assigned to particular models depends on the configuration. The z9 EC 12-PU and 16-PU MCMs have two SAPs as standard. The standard number of SAPs is two in a Model S08, four in an S18, six in an S28, and eight in a S38 and S54. Optional additional SAPs may be purchased, as shown in Table 2-5 on page 44.

The z9 EC has two spare PUs that may reside in any of the MCMs. A Model S08 has two spare PUs in its MCM. A Model S18 may have its spare PUs in the MCM of book 0 or in the MCM of book 1, or the spares may be straddled across both MCMs. Non-characterized PUs act as spares. The number of these additional spare PUs is dependent on the number of books in the configuration and how many PUs are non-characterized.

## PU pools

PUs defined as CPs, IFLs, ICFs, zIIPs, and zAAPs are grouped together in their own pool, from where they can be managed separately. This simplifies capacity planning and management for logical partition significantly. The separation also has an effect on weight management since CP, zAAP, and zIIP weights can be managed separately. See "PU weighting" on page 56 for more.

All assigned PUs of a z9 EC are grouped together in the PU pool. These PUs are dispatched to online logical PUs. As an example, consider a z9 EC with 10 CPs, three zAAPs, two IFLs, two zIIPs, and one ICF. The system has a PU pool of 18 PUs, called the *pool width*.

Subdivision of the PU pool defines:

- ▶ A CP pool of 10 CPs
- ▶ An ICF pool of one ICF
- ▶ An IFL pool of two IFLs
- ▶ A zAAP pool of three zAAPs
- ▶ A zIIP pool of two zIIPs

PUs are placed in these pools:

- ▶ When the server is Power-On Reset
- ▶ At a time of a concurrent upgrade
- ▶ As a result of an addition of PUs during a CBU
- ▶ Following a Capacity on Demand upgrade, through On/Off Capacity on Demand or Customer Initiated Upgrade

Also, when a dedicated logical partition is deactivated or logically deconfigures a logical PU, its PUs are returned to the proper pool.

PUs are removed from their pools when a concurrent downgrade takes place as the result of removal of a CBU, and through On/Off Capacity on Demand and conversion of a PU. Also, when a dedicated logical partition is activated, its PUs are taken from the proper pools, as is the case when a logical partition logically configures a PU on, if the width of the pool allows.

By having different pools, a weight distinction can be made between CPs, zAAPs, and zIIPs, where in the past specialty engines like a zAAP automatically received the weight of the initial CP.

For a logical partition, logical PUs are dispatched from the supporting pool only. This means that logical CPs are dispatched from the CP pool, logical zAAPs are dispatched from the zAAP pool, logical zIIPs from the zIIP pool, logical IFLs from the IFL pool, and the logical ICFs from the ICF pool.

### PU weighting

Since zAAPs, zIIPs, IFLs, and ICFs on a z9 EC have their own pools from where they are dispatched, they can be given their own weights. zAAPs and zIIPs do not get a weight assigned based on the logical partition CP weight, but based on their own weight specification.

The following paragraphs explain the difference between the implementation of PU weighting on a z990 and a z9 EC.

The z990 has a PU pool with 10 CPs, and a PU pool with six specialty engines (zAAPs, IFLs, and ICFs, all in one pool). The total pool weight is 1000 for the CPs and 1500 for the specialty engines.

On a z990, as demonstrated in Figure 2-14, the logical partition PU share is calculated as:

Pool PUs x (logical partition pool weight / total pool weight)

The PU share for the ZOS1 logical partition CPs is therefore  $10 \times (250/1000) = 2.5$ . The two zAAPs in this logical partition get their weight, calculated as  $6 \times (250/1500) = 1$ .

LPAR Name	LPAR Weight	Shared Logical PUs ON				PU Share	
		CP	zAAP	IFL	ICF	CP	Specialty
ZOS1	250c / 250z	10	2	NA	NA	2.5	1
ZOS2	750c / 750z	10	3	NA	NA	7.5	3
CF1	50	0	NA	NA	1	0	.2
CF2	50	0	NA	NA	1	0	.2
ZVM1	100	0	NA	2	NA	0	.4
LINUX1	300	0	NA	2	NA	0	1.2
Pool Weight >		1000	1500				
Total PUs (Physical) >						10	6

Figure 2-14 z990 PU weighting

The z9 EC has a PU pool of 10 CPs, three zAAPs, two IFLs, and one ICF. The total pool weight for CPs is 1000, for zAAPs it is 200, for IFLs 400, and for ICFs 100.

On a z9 EC, as demonstrated in Figure 2-15, the logical partition PU share is calculated as:

Pool PUs x (logical partition pool weight / total pool weight)

The PU share for the ZOS1 logical partition CPs is therefore  $10 \times (250/1000) = 2.5$ . The two zAAPs in this logical partition get their own weight, calculated as  $3 \times (100/200) = 1.5$ .

LPAR Name	LPAR Weight	Shared Logical PUs On				PU Share			
		CP	zAAP	IFL	ICF	CP	zAAP	IFL	ICF
ZOS1	250c / 100z	10	2	NA	NA	2.5	1.5	NA	NA
ZOS2	750c / 100z	10	3	NA	NA	7.5	1.5	NA	NA
CF1	50 - ICF	0	NA	NA	1	0	NA	NA	.5
CF2	50 - ICF	0	NA	NA	1	0	NA	NA	.5
ZVM1	100 - IFL	0	NA	2	NA	0	NA	.5	NA
LINUX1	300 - IFL	0	NA	2	NA	0	NA	1.5	NA
Pool Weight >		1000	200	400	100				
Total PUs (Physical) >						10	3	2	1

Figure 2-15 z9 EC PU weighting

For more information about PU pools and processing weights, refer to *IBM System z9 Processor Resource/Systems Manager™ Planning Guide*, SB10-7041.

## Central Processors

A Central Processor is a PU that has the z/Architecture instruction sets. It can run z/Architecture based operating systems (z/OS, z/VM, z/TPF, z/VSE, Linux), and the Coupling Facility Control Code (CFCC).

The z9 EC can only be initialized in LPAR mode. In LPAR mode, CPs can be defined as dedicated or shared to a logical partition. Reserved CPs can be defined to a logical partition, to allow for nondisruptive *image* upgrades. A logical partition can have up to 54 logical CPs defined. However, we recommend defining no more CPs than the operating system supports.

All PUs characterized as CPs within a configuration are grouped into the CP pool. The CP pool can be seen on the hardware management console workplace. Any z/Architecture operating systems and CFCCs can run on CPs that are assigned from the CP pool.

Within the limit of all non-characterized PUs available in the installed configuration, CPs can be concurrently assigned to an existing configuration through Capacity Upgrade on Demand (CUoD), Customer Initiated Upgrade (CIU), On/Off Capacity on Demand (On/Off CoD), or Capacity BackUp (CBU). More information about all forms of concurrent CP additions can be found in Chapter 8, “Concurrent upgrades and availability” on page 205.

If the MCMs in the installed books have no available PUs left, the assignment of the next CP will result in the need for a model upgrade and the installation of an additional book. Book installation is nondisruptive, but will take more time than a simple Licensed Internal Code upgrade. Only if reserved processors have been defined to a logical partition, and when the

operating system supports the function, additional CP capacity can be allocated to the logical partition dynamically.

### ***Granular capacity***

The z9 EC recognizes four distinct capacity settings for CPs. Full capacity CPs are identified as CP7. Up to 54 CPs can be configured. Beside full capacity CPs, three sub-capacity settings (CP6, CP5, and CP4) each for up to eight CPs are offered. The four capacity settings appear in hardware descriptions as shown below:

- ▶ CP7 feature code 7810
- ▶ CP6 feature code 7809
- ▶ CP5 feature code 7808
- ▶ CP4 feature code 7807

Granular capacity adds 24 sub-capacity settings to the 54 capacity settings that are available with full capacity CPs (CP7). Each of the 24 sub-capacity settings only apply to up to eight CPs, independent of the z9 EC model installed.

Information about CPs in the remainder of this chapter applies to all CP capacity settings, CP7, CP6, CP5, and CP4 unless indicated otherwise. For more details on granular capacity, see 2.3, “Model configurations” on page 71

### ***Capacity marker***

A capacity marker is how you remember the presence of purchased but unused capacity. For example, a z9 EC Model S08 could be configured with four full capacity CPs (CP7) and one CP purchased but not being used. The 4-core processor is identified with FC 4504 and a model capacity identifier 705 (FC 5705) is added to the configuration to remember the capacity purchased.

The same applies to the sub-capacity models; for example, when a z9 EC Model S18 is configured with four sub-capacity CPs (say CP5) and one CP purchased but not being used. The 4-core processor is identified with FC 4704 and a capacity marker identifying capacity indicator 505 (FC 5505) is added to the configuration to remember the capacity in store.

**Note:** Capacity settings smaller than the full capacity setting (CP6, CP5, and CP4) only apply to up to eight PUs characterized as CPs. Specialty engines such as IFLs, ICFs, zAAPs, and zIIPs always run at full capacity.

## **Integrated Facilities for Linux**

An Integrated Facility for Linux (IFL) is a PU that can be used to run Linux on System z or Linux guests on z/VM operating systems. Up to 54 PUs may be characterized as IFLs, depending on the z9 EC configuration. IFLs can be dedicated to a Linux or a z/VM logical partition, or be shared by multiple Linux guests or z/VM logical partitions running on the same z9 EC. Only z/VM and Linux on System z operating systems can run on IFLs.

All PUs characterized as IFLs within a configuration are grouped into the IFL pool. The IFL pool can be seen on the hardware management console workplace.

IFLs do not change the model capacity identifier of the z9 EC. Software product license charges based on the model capacity identifier are not affected by the addition of IFLs.



Within the limit of all non-characterized PUs available in the installed configuration, IFLs can be concurrently added to an existing configuration through Capacity Upgrade on Demand (CUoD), Customer Initiated Upgrade (CIU), or On/Off Capacity on Demand (On/Off CoD). An IFL CBU may have been purchased to provide IFL backup capacity for lost IFLs elsewhere. If the installed books have no unassigned PUs left, the assignment of the next IFL may require the installation of an additional book.

For more information about CUoD, CIU, or On/Off CoD, see Chapter 8, “Concurrent upgrades and availability” on page 205.

### ***Unassigned IFLs***

IFLs purchased but not put to use on a z9 EC are registered with an *unassigned IFL* feature. When the system is upgraded with an additional IFL, it is remembered that an already purchased IFL is present.

### **Internal Coupling Facilities**

An Internal Coupling Facility (ICF) is a PU used to run the IBM Coupling Facility Control Code (CFCC) for Parallel Sysplex environments. Within the capacity of the sum of all unassigned PUs in up to four books, up to 16 ICFs can be characterized, depending on the z9 EC model. At least a z9 EC configuration S18 is needed to assign 16 ICFs.

The ICF processors can only be used by Coupling Facility logical partitions. ICF processors can be dedicated to a CF logical partition, or shared by multiple CF logical partitions running in the same z9 EC server.

All ICF processors within a configuration are grouped into the ICF pool. The ICF pool can be seen on the hardware console.

Only Coupling Facility Control Code (CFCC) can run on ICF processors; ICFs do not change the model capacity identifier of the z9 EC. Software product license charges based on the model capacity identifier are not affected by the addition of ICFs.

ICFs can be concurrently assigned to an existing configuration through Capacity Upgrade on Demand (CUoD), On/Off Capacity on Demand (On/Off CoD), or Customer Initiated Upgrade (CIU). If the installed books have no non-characterized PUs left, the assignment of the next ICF may require the installation of an additional book. An ICF CBU may have been purchased to provide ICF backup capacity for lost ICFs elsewhere. For more information about CUoD, CIU, or On/Off CoD, see Chapter 8, “Concurrent upgrades and availability” on page 205.

### ***Dynamic ICF expansion***

Dynamic ICF expansion is a function that allows a CF logical partition running on dedicated ICFs to acquire additional capacity from the LPAR pool of shared CPs or shared ICFs. The trade-off between using ICF features or CPs in the LPAR shared pool is the exemption from software license fees for ICFs. Dynamic ICF expansion is available on any z9 EC that has at least one ICF.

Dynamic ICF expansion requires that the Dynamic Coupling Facility Dispatching function be turned on.

### ***Dynamic Coupling Facility Dispatching***

The Dynamic Coupling Facility Dispatching function has a dispatching algorithm that lets you define a backup Coupling Facility in a logical partition on the system. While this logical partition is in backup mode, it uses very little processor resources. When the backup CF becomes active, only the resource necessary to provide coupling is allocated.

The CFCC command DYNDISP controls the Dynamic CF Dispatching (use DYNDISP ON to enable the function). For more information, see 7.2.6, “Dynamic CF dispatching and Dynamic ICF expansion” on page 193.

## System z Application Assist Processors

A System z Application Assist Processor (zAAP) reduces the standard processor (CP) capacity requirements for Java applications, freeing up capacity for other workload requirements. zAAPs do not increase the MSU value of the processor and therefore do not affect the software license fee. Support for the System z Application Assist Processor is available in z/OS V1.6 and later.

The System z Application Assist Processor (zAAP) is a PU that is used exclusively for running Java application workloads under z/OS. zAAPs only run Java code. The IBM SDK for z/OS Java 2 Technology Edition (the Java Virtual Machine), in cooperation with z/OS and PR/SM, directs JVM™ processing from CPs to zAAPs. Apart from the cost savings this may realize, the integration of Java-based applications with their associated database systems (such as DB2, IMS, or CICS) may simplify the infrastructure, for example, by reducing the number of TCP/IP programming stacks and server interconnect links. Furthermore, processing latencies that would occur if Java application servers and their database servers were deployed on separate server platforms are prevented.

z/VM V5R3 and later supports zAAP for guest exploitation.

One CP must be installed with or prior to any zAAP being installed. The number of zAAPs in a server cannot exceed the number of CPs plus unassigned CPs in that server. Within the capacity of the sum of all unassigned PUs in up to four books, up to 27 zAAPs can be characterized. This is on a z9 EC configuration S54. Table 2-6 shows the maximum number of zAAPs per model.

Table 2-6 Maximum number of zAAPs per model

	z9 EC S08	z9 EC S18	z9 EC S28	z9 EC S38	z9 EC S54
Max zAAPs	4	9	14	19	27

Within the limit of all non-characterized PUs available in the installed configuration, zAAPs can be concurrently added to an existing configuration through Capacity Upgrade on Demand (CUoD), Customer Initiated Upgrade (CIU), and On/Off Capacity on Demand (On/Off CoD). A zAAP CBU may have been purchased to provide zAAP backup capacity for lost zAAPs elsewhere.

With On/Off CoD, temporary zAAP capacity may be concurrently installed by ordering On/Off CoD Active zAAP features up to the number of current zAAPs that are permanently purchased. The total number of On/Off CoD Active zAAPs plus zAAPs cannot exceed the number of On/Off Active CPs plus the number of CPs plus the number unassigned CPs on a z9 EC server.

For more information about CUoD, CIU, or On/Off CoD, see Chapter 8, “Concurrent upgrades and availability” on page 205. If the installed books have no unassigned PUs left, the assignment of the next zAAP may require the installation of an additional book.

PUs characterized as zAAPs within a configuration are grouped into the zAAP pool. This allows zAAPs to have their own processing weights, independent from the weight of parent CPs. The zAAP pool can be seen on the hardware console.

zAAPs are orderable by feature code (FC 7814). Up to one zAAP can be ordered for each CP or marked CP configured in the server.

**Important:** The zAAP is a specific example of an assist processor that is known generically as an Integrated Facility for Applications (IFA). The generic term IFA appears in panels, messages, and other online information relating to the zAAP.

### **zAAPs and logical partition definitions**

zAAP processors can be defined as dedicated or shared processors in a logical partition and are always related to CPs of the same partition. In a logical partition, logical CPs, and zAAPs are either dedicated or shared.

### **Purpose of a zAAP**

zAAPs are designed for z/OS Java code execution. When Java code must be executed (that is, under control of WebSphere), the z/OS Java Virtual Machine (JVM) calls the function of the zAAP. The z/OS dispatcher then suspends the JVM task on the CP it is running on and dispatches it on an available zAAP. After the Java application code execution is finished, the z/OS dispatcher redispaches the JVM task on an available CP, after which normal processing is resumed. This reduces the CP time needed to run WebSphere applications, freeing capacity for other workloads.

Figure 2-16 shows the logical flow of Java code running on a z9 EC server that has a zAAP available. The Java Virtual Machine (JVM), when it starts execution of a Java program, passes control to the z/OS dispatcher that will verify the availability of a zAAP:

- ▶ If a zAAP is available (not busy), the dispatcher will suspend the JVM task on the CP, and assign the Java task to the zAAP. When the task returns control to the JVM, it passes control back to the dispatcher that will reassign the JVM code execution to a CP.
- ▶ If there is no zAAP available at that time, the z/OS dispatcher may allow a Java task to run on a standard CP (depending on the option used in the OPT statement in the IEAOPTxx member of SYS1.PARMLIB).

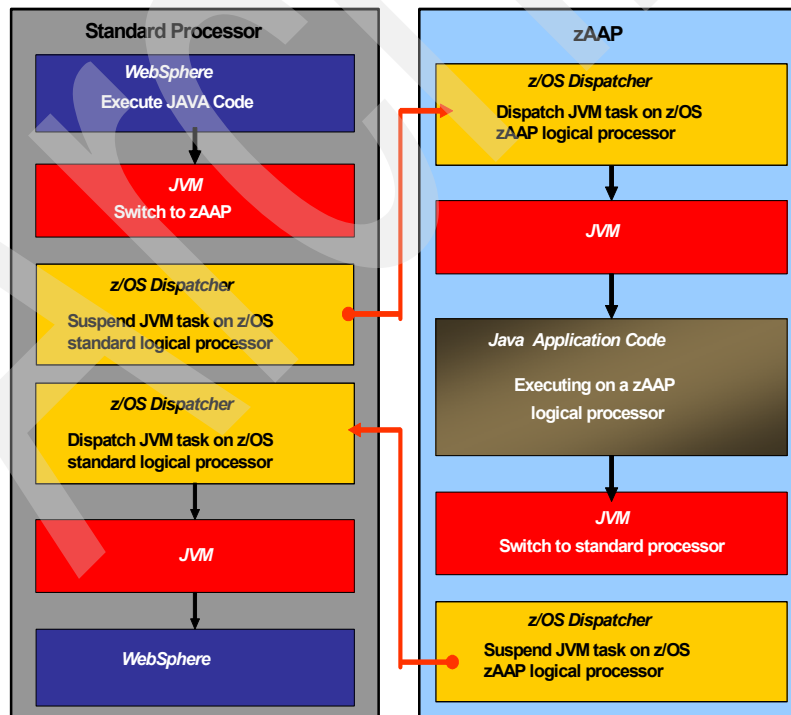


Figure 2-16 Logical flow of Java code execution on a zAAP

## **Software support**

zAAPs do not change the model capacity identifier of the z9 EC. IBM software product license charges based on the model capacity identifier are not affected by the addition of zAAPs. z/OS Version 1.6 is the minimum level for supporting zAAPs, together with IBM SDK for z/OS Java 2 Technology Edition V1.4.1.

Exploiters of zAAPs include:

- ▶ WebSphere Application Server V5.1
- ▶ CICS/TS V2.3
- ▶ DB2 UDB for z/OS Version 8
- ▶ IMS Version 8
- ▶ WebSphere WBI for z/OS

The functioning of a zAAP is transparent to all Java programming on JVM V1.4.1 and later.

A zAAP only executes Java Virtual Machine (JVM) code. JVM is the only authorized user of a zAAP in association with some parts of system code, such as the z/OS dispatcher and supervisor services. A zAAP is not able to process I/O or clock comparator interruptions and does not support operator controls like IPL.

Java application code can either run on a CP or a zAAP. The installation can manage the use of CPs such that Java application code runs only on a CP, only on a zAAP, or on both.

Four execution options for Java code execution are available.<sup>1</sup> These options are user specified in IEAOPTxx and can be dynamically altered by the SET OPT command.

- ▶ Option 1: Java dispatching by priority (IFAHONORPRIORITY=YES)

The default option, which specifies that standard processors run both zAAP eligible and non-zAAP eligible work in priority order when the zAAPs indicate the need for help from standard CPs. The need for help is determined by the Alternate Wait Management (AWM) function of SRM for both standard CPs and zAAPs. Only standard CPs help each other, but they can also help zAAPs if option YES is specified.

When zAAPs are configured and online, they only execute Java work in priority order while the CPs execute normal tasks and JVM tasks in priority order. This option is oriented towards servicing work with the highest priority first, regardless of the type of work.

- ▶ Option 2: Java dispatching by priority (IFAHONORPRIORITY=NO)

Standard CPs will not look for zAAP eligible work regardless of the demand for zAAP eligible work as long as there is non zAAP work available. zAAP eligible work can execute on standard CPs, but at a lower priority than non-Java work, depending on the setting of the IFACROSSOVER parameter.

- ▶ Option 3: Java discretionary crossover (IFACROSSOVER=YES)

Specifies that zAAP eligible work should be examined by standard CPs when the CP would otherwise enter a wait state. zAAP eligible work is treated as though it has a lower priority than the WLM discretionary priority and standard work is considered before any zAAP eligible work.

This option is oriented towards environments where not enough zAAP capacity may be available and the Java workload has no need for priority over non-Java work.

---

<sup>1</sup> With the fix for APARs OA14131 and OA13953

► Option 4: No Java crossover (IFACROSSOVER=NO)

This option specifies that a standard CP will not look for Java work before entering the wait state. This option assures that all Java work is done on a zAAP as long as one is available. If, for example, Sub Capacity Workload License Charging is applicable, Java work that is executed on a CP will increase CP utilization and consequently may increase the software charges.

If zAAPs are defined to the logical partition but are not online, the zAAP eligible work units are processed by standard CPs in priority order. The system ignores the IFACROSSOVER and IFAHONORPRIORITY parameters in this case and handles the work as though it had no eligibility to zAAPs.

Table 2-7 summarizes the possible options.

Table 2-7 Java code execution options

IFACROSSOVER	IFAHONORPRIORITY	Java workload dispatching
No	No	No zAAP eligible work is dispatched to CPs. <sup>a</sup>
No	Yes	zAAP eligible work is dispatched on CP only when help is requested (Alternate Wait Management).
Yes	No	zAAP eligible work is dispatched on CPs only when no non-zAAP eligible work is ready.
Yes	Yes	zAAP eligible work is dispatched on CPs when help is requested (Alternate Wait Management) or when no non-zAAP eligible work is ready.

a. There are cases where zAAP eligible work will run on CPs when it is necessary to handle conflicts for system resources between zAAP eligible work and other work.

### System z9 Integrated Information Processor (zIIP)

A System z9 Integrated Information Processor (zIIP) is designed so that eligible database workloads can work with z/OS (Version 6 and above) to have a portion of its enclave SRB work directed to it. zIIPs do not increase the MSU value of the processor and therefore do not affect the software license fee.

z/VM V5R3 and later supports zIIP for guest exploitation.

z/OS, acting on the direction of the program running in SRB mode, controls the distribution of the work between the general purpose processor (CP) and the zIIP.

DB2 UDB for z/OS Version 8 exploits the zIIP by indicating to z/OS which portions of the work are eligible to be routed to a zIIP. Types of eligible DB2 UDB for z/OS V8 workloads executing in SRB mode that can be sent to the zIIP include:

Three types of DB2 work have been identified that can benefit from the zIIP:

- Query processing of network connected applications that access the DB2 data base over a TCP/IP connection using DRDA®.

DRDA enables relational data to be distributed among multiple platforms. It is native to DB2 for z/OS, thus reducing the need for additional gateway products that may affect performance and availability. The application uses the DRDA requester or server to access a remote data base (DB2 Connect™ is an example of a DRDA application requester.)

- ▶ Star schema query processing mostly used in Business Intelligence (BI) work.  
A star schema is a relational data base schema for representing multidimensional data. It stores data in a central fact table and is surrounded by additional dimension tables holding information about each perspective of the data. A star schema query (as an example) joins several dimensions of a star schema data set.
- ▶ DB2 utilities that are used for index maintenance like LOAD, REORG, and REBUILD.  
Indices allow quick access to table rows but over time as data in large data bases is manipulated they become less efficient. They need to be maintained.

The zIIP runs portions of eligible database workloads and in doing so helps to free up computer capacity and lower software costs. Not all DB2 workloads are eligible for zIIP processing. DB2 UDB for z/OS V8 gives z/OS the necessary information to direct portions of the work to the zIIP. This result is that in every user situation, different variables determine how much work is actually redirected to the zIIP.

To exploit a zIIP, the following is required:

- ▶ IBM System z9.
- ▶ z/OS Version 1.6 or higher with PTF.
- ▶ DB2 UDB for z/OS Version 8 with PTF.

One CP must be installed with or prior to any zIIP being installed. The number of zIIPs in a server cannot exceed the number of CPs plus unassigned CPs in that server. Within the capacity of the sum of all unassigned PUs in up to four books, up to 27 zIIPs on a configured S54, can be characterized. Table 2-8 shows the maximum number of zIIPs per model.

Table 2-8 Maximum number of zIIPs per model

	z9 EC S08	z9 EC S18	z9 EC S28	z9 EC S38	z9 EC S54
Maximum zIIPs	4	9	14	19	27

**Note:** For each assigned, and unassigned CP in the server, one zIIP and one zAAP may be configured.

zIIPs are orderable by feature code (FC 7815). Up to one zIIP can be ordered for each CP or marked CP configured in the server. If the installed books have no unassigned PUs left, the assignment of the next zIIP may require the installation of an additional book.

PUs characterized as zIIPs within a configuration are grouped into the zIIP pool. This allows zIIPs to have their own processing weights, independent from the weight of parent CPs. The zIIP pool can be seen on the hardware console.

Within the limit of all non-characterized PUs available in the installed configuration, zIIPs can be concurrently added to an existing configuration through Capacity Upgrade on Demand (CUoD), Customer Initiated Upgrade (CIU), and On/Off Capacity on Demand (On/Off CoD). zIIP CBU capacity can be purchased to provide zIIP backup capacity.

With On/Off CoD, temporary zIIP capacity may be concurrently installed by ordering On/Off CoD Active zIIP features up to the number of current zIIPs that are permanently purchased. The total number of On/Off CoD Active zIIPs plus zIIPs cannot exceed the number of On/Off active CPs plus the number of CPs plus the number unassigned CPs on a z9 EC server.

For more information about CUoD, CIU, or On/Off CoD, see Chapter 8, “Concurrent upgrades and availability” on page 205.

### ***zIIPs and logical partition definitions***

zIIP processors can be defined as dedicated or shared in a logical partition. The number of zIIPs in a logical partition can be larger than the number of CPs in that logical partition.

### **System Assist Processors**

A System Assist Processor (SAP) is a PU that runs the Channel Subsystem Licensed Internal Code to control I/O operations.

All SAPs perform I/O operations for all logical partitions. All z9 EC models have standard SAPs configured. The z9 EC Model S08 has two SAPs, the Model S18 has four SAPs, the Model S28 has six SAPs, and the models S32 and S54 have eight SAPs as the standard configuration.

A standard SAP configuration provides a very well-balanced system for most environments. However, there are application environments with very high I/O rates (typically some TPF environments). In this case, optional additional SAPs can be ordered. Assignment of additional SAPs can increase the capability of the Channel Subsystem to perform I/O operations. In z9 EC servers, the number of SAPs can be greater than the number of CPs.

### ***Optional additional orderable SAPs***

An option available on all models is additional orderable SAPs. These additional SAPs increase the capacity of the Channel Subsystem to perform I/O operations, usually suggested for TPF environments. The maximum number of optional additional orderable SAPs depends on the configuration and the number of available uncharacterized PUs:

- ▶ z9 EC Model S08: Maximum additional orderable SAPs is 5.
- ▶ z9 EC Model S18: Maximum additional orderable SAPs is 13.
- ▶ z9 EC Model S28: Maximum additional orderable SAPs is 21.
- ▶ z9 EC Model S38: Maximum additional orderable SAPs is 24.
- ▶ z9 EC Model S54: Maximum additional orderable SAPs is 24.

### ***Optionally assignable SAPs***

Assigned CPs may be optionally reassigned as SAPs instead of CPs, using the Reset Profile on the Hardware Management Console (HMC). This reassignment increases the capacity of the Channel Subsystem to perform I/O operations, usually for some specific workloads or I/O intensive testing environments.

If it is intended to activate a modified server configuration with a modified SAP configuration, a reduction in the number of CPs available will reduce the number of logical processors that can be activated. Activation of a logical partition will fail if the number of logical processors attempted to activate exceeds the number of CPs available. To avoid a logical partition activation failure, it should be verified that the number of logical processors assigned to a logical partition does not exceed the number of CPs available.

**Note:** Concurrent upgrades are not supported with CPs defined as additional SAPs.

### **Reserved processors**

Reserved processors can be defined to a logical partition. Reserved processors are defined by the Processor Resource/System Manager (PR/SM) to allow for a nondisruptive *capacity* upgrade. Reserved processors are like spare *logical* processors. They can be defined as shared or dedicated.

Reserved processors can be dynamically configured online by an operating system that supports this function if there are enough unassigned PUs available to satisfy this request. The PR/SM rules regarding logical processor activation remain unchanged.

Reserved processors also provide the capability of defining to a logical partition more logical processors than the number of available CPs, IFLs, ICFs, zAAPs, and zIIPs in the configuration. This makes it possible to configure online, nondisruptively, more logical processors after additional CPs, IFLs, ICFs, zAAPs, and zIIPs have been made available concurrently with one of the Capacity on Demand options (CUoD, CIU, On/Off CoD for CPs, IFLs, ICFs, zAAPs, and zIIPs, or CBU for CPs). See Chapter 8, “Concurrent upgrades and availability” on page 205 for more details.

When no reserved processors are defined to a logical partition, a processor upgrade in that logical partition is disruptive, requiring the following tasks:

- ▶ Partition deactivation
- ▶ A logical processor definition change
- ▶ Partition activation

The maximum number of reserved processors that can be defined to a logical partition depends on the number of logical processors that are already defined.

One should not define more active plus reserved processors than the operating system for the logical partition can support. z/OS V1R8, z/OS V1R7, and z/OS V1R6 support up to 32 processors, including CPs, zAAPs, and zIIPs. z/VM V5R1 and V5R2 support up to 24 processors, either all CPs or all IFLs. z/VM V5R3 supports up to 32 processors. For more information about logical processors and reserved processors definition, see 2.4, “Logical partitioning” on page 79.

### **Processor Unit characterization**

Processor Unit (PU) characterization is done at Power-on Reset time when the server is initialized. The z9 EC is always initialized in LPAR mode, and it is the PR/SM hypervisor that has responsibility for the PU assignment.

Additional SAPs are characterized first, then CPs, followed by IFLs, ICFs, zAAPs, and zIIPs. For performance reasons, CPs for a logical partition are grouped together as much as possible. Having all CPs grouped in as few books as possible limits memory and cache interference to a minimum.

When an additional book is added concurrently after Power-on Reset and new logical partitions are activated, or processor capacity for active partitions is dynamically expanded, the additional PU capacity may be assigned from the new book. It is only after the next Power-on Reset that the Processor Unit allocation rules take into consideration the newly installed book.

### **Transparent CP, IFL, ICF, zAAP, zIIP, and SAP sparing**

Characterized PUs, whether CPs, IFLs, ICFs, zAAPs, zIIPs, or SAPs, are transparently spared, following distinct rules.

The z9 EC server comes with two spare PUs system wide. Depending on the model, CP, IFL, ICF, zAAP, zIIP, and SAP sparing is completely transparent and requires no operating system or operator intervention.



With transparent sparing, the status of the application that was running on the failed processor is preserved and will continue processing on a newly assigned CP, IFL, ICF, zAAP, zIIP, or SAP (allocated to one of the spare PUs) without customer intervention. If no spare PU is available, application preservation is invoked.

### **Application preservation**

Application preservation is used in the case where a processor fails and there is no spare PU available. The state of the failing processor is passed to another active processor used by the operating system and, through operating system recovery services, the task is resumed successfully (in most cases, without customer intervention).

### **Dynamic SAP sparing and reassignment**

Dynamic recovery is provided in case of failure of the System Assist Processor (SAP). In the event of an SAP failure, if a spare PU is available, the spare PU will be dynamically assigned as a new SAP. If there is no spare PU available, and more than one CP is characterized, a characterized CP is reassigned as an SAP. In either case, there is no customer intervention required. This capability eliminates an unplanned outage and permits a service action to be deferred to a more convenient time.

### **Sparing rules**

Two PUs are reserved as spares on the z9 EC. The reserved spares are available to replace a failing two PU chip or two failing single PU chips. The spare PUs can be used for sparing any characterization, be it a CP, IFL, ICF, zAAP, zIIP, or SAP. On a z9 EC Model S08, two spares are allocated in one book (book 0). In multi book systems, it is the decision of PR/SM to allocate the spares anywhere in the system, even such that book 0 does not contain a spare at all.

Systems with a failed PU for which no spare is available will *call home* for a replacement. A system with a failed PU that has been spared and in fact requires an MCM to be replaced (called a *pending repair*) can still be upgraded when sufficient PUs are available.

The two standard SAPs are initially allocated to dual core processor chips. On a single-book configuration S08:

- ▶ When a PU failure occurs on a dual-core chip, the two standard spare PUs are used to recover the failing chip, even though only one of the PUs has failed.
- ▶ When a failure occurs on a PU on a single-core chip, one standard spare PU is used.
- ▶ When there are no spares left, non-characterized PUs are used for sparing.

The system does not issue an RSF call in either of the above circumstances.

On a multi-book configuration, models S18, S28, S38, or S54:

- ▶ In a first step, the standard spare PU is assigned as the spare, in the same manner as for a one-book system.
- ▶ In a second step, when there are no spares left, non-characterized PUs are used for sparing.

When non-characterized PU are used for sparing and might be needed to satisfy a On/Off CoD request, an RSF call occurs to request a book repair.

## 2.2.5 Memory design

As for PUs and the I/O subsystem designs, the z9 EC memory design equally provides great flexibility and high availability, allowing:

- ▶ Concurrent Memory upgrades (If the physically installed capacity is not yet reached.)

The z9 EC servers may have more physically installed memory than the initial available capacity. Memory upgrades within the physically installed capacity can be done concurrently by the Licensed Internal Code, and no hardware changes are required. Concurrent memory upgrades can be done through Capacity Upgrade on Demand or Customer Initiated Upgrade. Note that memory upgrades *cannot* be done through Capacity BackUp (CBU); see Table 8-1 on page 210 for more information.

- ▶ Concurrent Memory upgrades (If the physically installed capacity is reached.)

Physical memory upgrades require a book to be removed and re-installed after having replaced the memory cards in the book. Except for a Model S08, the combination of Enhanced Book Availability and flexible memory option allow you to concurrently add memory to the system. See “Book replacement and memory” on page 35 and “Flexible Memory option” on page 35 for more information.

- ▶ Memory sparing

Memory sparing is done by the use of X4 DRAMs across eight DIMMs (eight DIMMs equal one PMA), resulting in four spares per PMA with Chipkill.

- ▶ Partial Memory Restart

In the rare event of a memory card failure, Partial Memory Restart enables the system to be restarted with only part of the original memory. In a one-book system, the memory cards that make up PMA0 and PMA1 or PMA2 and PMA3 (depending on where the failure resides) are deactivated, after which the system can be restarted with the memory on the remaining memory cards.

In a system with more than one book, all physical memory in the book containing the failing memory card is taken offline, which allows you to bring up the system with the remaining physical memory in the other books. In this way, processing can be resumed until a replacement memory card is installed.

Memory error-checking and correction code detects and corrects single-bit errors, or 2-bit errors from a Chipkill failure, using the Error Correction Code (ECC). Also, because of the memory structure design, errors due to a single memory chip failure are corrected.

Memory background scrubbing provides continuous monitoring of storage for the correction of detected faults before the storage is used.

The memory cards use the latest fast 512 Mb and 1 Gb synchronous DRAMs. Memory access is interleaved between the memory cards to equalize memory activity across the cards.

Memory cards have 4 GB, 8 GB, or 16 GB of capacity. Memory cards installed in a book do not necessarily have the same capacity (as long as the DRAM sizes are the same). Books may contain different memory sizes.

The total capacity installed may have more usable memory than required for a configuration, and Licensed Internal Code Configuration Control (LIC-CC) will determine how much memory is used from each card. The sum of the LIC-CC provided memory from each card is the amount available for use in the system.

## Memory allocation

Memory assignment or allocation is done at Power-on Reset (POR) when the system is initialized. Actually, PR/SM is responsible for the memory assignments; it is PR/SM that controls the resource allocation of the server. Table 2-2 on page 33 shows the distribution of physical memory across books when a system is initially installed with the amounts of memory shown in the first column. However, the table gives no indication of *where* the initial memory is allocated. Memory allocation is done as evenly as possible across all installed books.

PR/SM has knowledge of the amount of purchased memory and how it relates to the available physical memory in each of the installed books. PR/SM has control over all physical memory and therefore is able to make physical memory available to the configuration when a book is nondisruptively added. PR/SM also controls the reassignment of the content of a specific physical memory array in one book to a memory array in another book. This is known as the Memory Copy/Reassign function, used to reallocate the memory content from the memory in a book to another memory location when Enhanced Book Availability is applied to concurrently remove and re-install a book in case of an upgrade or repair action.

Due to the memory allocation algorithm, systems that undergo a number of MES upgrades for memory can have a variety of memory card mixes in all books of the system. If, however unlikely, memory should fail, it is technically feasible to Power-on Reset the system with the remaining memory resources (see “Partial Memory Restart” on page 68). After Power-on Reset, the memory distribution across the books is now different, as is the amount of available memory.

Capacity Upgrade on Demand (CUoD) for memory can be used to order more memory than needed on the initial model, but that is required on the target model; see “Memory upgrades” on page 34. For more information about CUoD for memory, refer to “CUoD for memory” on page 213.

Processor memory, even though physically the same, can be configured as both Central Storage and Expanded Storage.

## Central Storage (CS)

Central Storage (CS) consists of main storage, addressable by programs, and storage not directly addressable by programs. Non-addressable storage includes the Hardware System Area (HSA). Central Storage provides:

- ▶ Data storage and retrieval for PUs and I/O
- ▶ Communication with PUs and I/O
- ▶ Communication with and control of optional Expanded Storage
- ▶ Error checking and correction

Central Storage can be accessed by all processors, but cannot be shared between logical partitions. Any system image (logical partition) must have a Central Storage size defined. This defined Central Storage is allocated exclusively to the logical partition during partition activation.

A logical partition can have more than 2 GB defined as Central Storage, but 31-bit operating systems cannot use Central Storage above 2 GB; refer to 2.5, “Storage operations” on page 84 for more details.

## Expanded Storage (ES)

Expanded Storage can optionally be defined on z9 EC servers. Expanded Storage is physically a section of processor storage. It is controlled by the operating system and transfers 4 KB pages to and from Central Storage.

Except for z/VM, z/Architecture operating systems do *not* use Expanded Storage. As they operate in 64-bit addressing mode, they can have all the required storage capacity allocated as Central Storage. z/VM is an exception since, even when operating in 64-bit mode, it can have guest virtual machines running in 31-bit addressing mode, which can use Expanded Storage.

It is *not* possible to define Expanded Storage to a Coupling Facility image. However, any other image type can have Expanded Storage defined, even if that image runs a 64-bit operating system and does not use Expanded Storage.

The z9 EC only runs in LPAR mode. Storage is placed into a single storage pool called LPAR Single Storage Pool, which can be dynamically converted to Expanded Storage and back to Central Storage as needed when partitions are activated or de-activated.

### ***LPAR single storage pool***

In LPAR mode, storage is not split into Central Storage and Expanded Storage at Power-on Reset. Rather, the storage is placed into a single Central Storage pool that is dynamically assigned to Expanded Storage and back to Central Storage, as needed.

The Storage Assignment function of a Reset Profile on the Hardware Management Console just shows the total *Installed Storage* and the *Customer Storage*, which is the total installed storage minus the Hardware System Area (HSA). Logical partitions are still defined to have Central Storage and optional Expanded Storage. Activation of logical partitions, as well as dynamic storage reconfiguration, will cause the storage to be converted to the type needed.

Activation of logical partitions as well as dynamic storage reconfiguration will cause the storage to be assigned to the type needed (CS or ES). This does not require a Power-on Reset. No software support is required to take advantage of this function.

### **Hardware System Area (HSA)**

The Hardware System Area (HSA) is a non-addressable storage area that contains the server Licensed Internal Code and configuration-dependent control blocks. The HSA size varies according to:

- ▶ The number of defined logical partitions.
- ▶ If dynamic I/O is not enabled, the size and complexity of the system I/O configuration varies. The HSA may hold the configuration information for up to 63.75 K devices, and up to 64 K alias devices per Channel Subsystem defined in IOCDS.
- ▶ If dynamic I/O is enabled, the MAXDEV value specified in HCD or IOCP, in support of dynamic I/O configuration, limits the amount of definable subchannels.

**Note:** The size of the HSA on the z9 EC may be significantly larger than on a z990 or z890 system. It may range from less than 2 GB to up to 4.5 GB for a fully configured system. Especially for systems with relatively small memory sizes, caution is requested. We recommend using the HSA Estimator for z9 EC, available on Resource Link, to be able to plan for a sufficient amount of memory.

When system is activated, the HSA is always allocated in the physical memory of book 0. However it can be moved to another book later.

## 2.3 Model configurations

The z9 EC server model nomenclature is based on the number of PUs available for customer use in each configuration. Five models of the z9 EC server are available:

<b>Model S08</b>	Eight PUs are available for characterization as CPs, IFLs, ICFs, up to four zAAPs or zIIPs, or up to five additional SAPs.
<b>Model S18</b>	Eighteen PUs are available for characterization as CPs, IFLs, up to 16 ICFs, up to nine zAAPs or zIIPs, or up to 13 additional SAPs.
<b>Model S28</b>	Twenty-eight PUs are available for characterization as CPs, IFLs, up to 16 ICFs, up to 14 zAAPs or zIIPs, or up to 21 additional SAPs.
<b>Model S38</b>	Thirty-eight PUs are available for characterization as CPs, IFLs, up to 16 ICFs, up to 19 zAAPs or zIIPs, or up to 24 additional SAPs.
<b>Model S54</b>	Fifty-four PUs are available for characterization as CPs, IFLs, up to 16 ICFs, up to 27 zAAPs or zIIPs, or up to 24 additional SAPs.

When a z9 EC order is configured, PUs are characterized according to their intended usage. They can be ordered as:

<b>CP</b>	The processor purchased and activated supporting the z/OS, z/VSE, VSE/ESA™, z/VM, TPF, and Linux operating systems. Can also run Coupling Facility Control Code (CFCC). A CP can also be configured to run as an SAP.
<b>Capacity marked CP</b>	A processor purchased for future use as a CP is marked as available capacity. It is offline and unavailable for use.
<b>IFL</b>	The Integrated Facility for Linux is a processor that is purchased and activated for use by the z/VM for Linux guests and Linux operating systems.
<b>Unassigned IFL</b>	A processor purchased for future use as an IFL. It is offline and unavailable for use.
<b>ICF</b>	A processor purchased and activated for use by the Coupling Facility Control Code (CFCC).
<b>zAAP</b>	A processor purchased and activated to run Java code under control of z/OS JVM. <sup>2</sup>
<b>zIIP</b>	A processor purchased and activated for z/OS V1R8 and later to run eligible workloads. <sup>2</sup>
<b>Additional SAP</b>	The optional System Assist Processor is a processor that is purchased and activated for use as an SAP.

A Capacity Marker identifies that a certain number of CPs have been purchased. This number of purchased CPs is higher than the number of CPs actively used. The Capacity Marker marks the availability of purchased but unused capacity intended to be used as CPs in the future; they usually have this status for software charging reasons. Unused CPs do not count in establishing the MSU value to be used for MLC software charging, or when charged on a per processor basis.

Unassigned IFLs are purchased IFLs with the intention to be used as IFLs, and usually have this status for software and maintenance charging reasons. Unassigned IFLs do not count in establishing the charge for either z/VM or Linux.

<sup>2</sup> z/VM V5R3 supports zAAP and zIIP processors for guest exploitation.

This method prevents RPQ handling in case a temporary downgrade is required. When the capacity need arises, the marked CPs and unassigned IFLs can be assigned nondisruptively.

## Upgrades

Concurrent CP, IFL, ICF, zAAP, or zIIP upgrades are done within a z9 EC model. Concurrent upgrades require PU spares. PU spares are PUs that are *not* the two standard spares but those PUs that are not characterized as a CPs, IFLs, ICFs, zAAPs, zIIPs, or SAPs.

If the upgrade request cannot be accomplished within the given configuration, an upgrade is required. An upgrade will cause the addition of one or more books to accommodate the desired capacity. Additional books can be installed concurrently.

Upgrades from one z9 EC configuration to another are concurrent and mean that one or more books are added. However, there is an exception. Upgrades from any z9 EC model up to a Model S38 to a z9 EC Model S54 is disruptive since this upgrade requires a full book replacement. Table 2-9 shows the possible upgrades within the z9 EC configuration range.

Table 2-9 z9 EC upgrade paths

from	to	z9 EC S08	z9 EC S18	z9 EC S28	z9 EC S38	z9 EC S54 <sup>a</sup>
Model S08		-	Yes	Yes	Yes	Yes
Model S18		-	-	Yes	Yes	Yes
Model S28		-	-	-	Yes	Yes
Model S38		-	-	-	-	Yes

a. Disruptive upgrade

Upgrades from any z900 to any z9 EC server are supported (with the exception of the z900 model 100, which can only be upgraded to another z900 model). Upgrades from any z990 to any z9 EC server are supported.

There are no upgrade paths to the z9 EC from either z800 models or z890 models. Upgrades from a z9 BC Model xxx to a z9 EC Model S08 are available. Table 2-10 shows the upgrade paths to a z9 EC.

Table 2-10 Upgrade paths to z9 EC

From	To	z9 EC
z900		Yes
z990		Yes
z9 BC Model S07		Yes (S08)

## PU characterization

A minimum of one PU characterized as a CP, IFL, or ICF is required per system. The maximum number of CPs is 54, the maximum number of IFLs is 54, and the maximum number of ICFs is 16. The maximum number of zAAPs amounts to 27, but requires an equal number of characterized CPs. Even so the maximum number of zIIPs amounts to 27 and requires an equal number of characterized CPs. The sum of all zAAPs, and zIIPs cannot be larger than two times the number of characterized CPs.

Not all PUs on a given model are required to be characterized. Only purchased PUs are identified by a feature code.

### Concurrent PU conversions

Assigned CPs, assigned IFLs, unassigned IFLs, ICFs, zAAPs, and zIIPs may be converted to other assigned or unassigned feature codes. Valid conversions are:

- ▶ From a CP to an IFL, ICF, zAAP, or zIIP.
- ▶ From an IFL to a CP, unassigned IFL, ICF, zAAP, or zIIP.
- ▶ From an unassigned IFL to an IFL.
- ▶ From an ICF to a CP, IFL, zAAP, or zIIP.
- ▶ From a zAAP to a CP, IFL, ICF, or zIIP.
- ▶ From a zIIP to a CP, IFL, ICF, or zAAP.

Most listed conversions are nondisruptive. In exceptional cases, the conversion may be disruptive, for example, when a z9 EC Model S08 with eight CPs is converted to an all IFL system. In addition, a logical partition may be disrupted when PUs must be freed before they can be converted. Conversion information is summarized in Table 2-11.

Table 2-11 Concurrent PU conversions

From	To	CP	IFL	Unassigned IFL	ICF	zAAP	zIIP
CP	-	-	Yes	No	Yes	Yes	Yes
IFL	Yes	Yes	-	Yes	Yes	Yes	Yes
Unassigned IFL	No	No	Yes	-	No	No	No
ICF	Yes	Yes	Yes	No	-	Yes	Yes
zAAP	Yes	Yes	Yes	No	Yes	-	Yes
zIIP	Yes	Yes	Yes	No	Yes	Yes	-

### Model capacity identifier (CI)

In order to recognize how many PUs are characterized as a CP, the STSI instruction returns a value that can be seen as a model capacity identifier that determines the number and speed of characterized CPs. Characterization of a PU as an IFL, an ICF, a zAAP, or a zIIP is not reflected in the output of the STSI instruction, since they have no effect on software charging.

Four distinct model capacity identifier ranges are recognized:

- ▶ For full capacity engines, model capacity identifier 701 to 754 are used. They express the 54 possible capacity settings from one to 54 characterized CPs.
- ▶ Three model capacity identifier ranges offer a unique level of granular capacity at the low end. They are available when no more than eight CPs are characterized. These three subcapacity settings applied to up to eight CPs offer 24 additional capacity settings.

### Granular capacity

The z9 EC offers 24 additional capacity settings at the low end of the processor. Only eight CPs can have granular capacity. When subcapacity settings are used, other PUs can only be characterized as specialty engines.

Three ranges of granular subcapacity settings are defined. They have model capacity identifiers numbered from 401 to 408, 501 to 508, and 601 to 608.

**Note:** Within a z9 EC, all CPs have the same capacity identifier. Specialty engines operate at full speed.

Table 2-12 shows that regardless of the number of books, a configuration with one characterized CP is possible; for example, a z9 EC Model S54 may have only one PU characterized as a CP.

Table 2-12 Model capacity identifiers

z9 EC	Model capacity identifier
z9 EC Model S08	701–708, 601–608, 501–508, 401–408
z9 EC Model S18	701–718, 601–608, 501–508, 401–408
z9 EC Model S28	701–728, 601–608, 501–508, 401–408
z9 EC Model S38	701–738, 601–608, 501–508, 401–408
z9 EC Model S54	701–754, 601–608, 501–508, 401–408

**Note:** Model capacity identifier 700 is used for IFL or ICF only configurations.

### Model capacity identifier and MSU values

All model capacity identifiers have a related MSU value that is used to determine the software license charge for MLC software. Table 2-13 and Table 2-14 on page 75 show MSU values for each model capacity identifier.

Table 2-13 Model capacity identifier and MSU values

Model capacity identifier	MSU	Model capacity identifier	MSU	Model capacity identifier	MSU
701	81	719	1077	737	1850
702	158	720	1127	738	1889
703	229	721	1177	739	1927
704	298	722	1226	740	1936
705	363	723	1274	741	1998
706	422	724	1314	742	2033
707	479	725	1353	743	2067
708	532	726	1400	744	2101
709	584	727	1436	745	2135
710	640	728	1481	746	2168
711	690	729	1524	747	2201
712	742	730	1567	748	2233
713	795	731	1609	749	2265



Model capacity identifier	MSU	Model capacity identifier	MSU	Model capacity identifier	MSU
714	843	732	1650	750	2295
715	893	733	1691	751	2324
716	938	734	1732	752	2353
717	985	735	1772	753	2381
718	1032	736	1811	754	2409

Table 2-14 Model capacity identifier and MSU values for subcapacity models

Model capacity identifier	MSU	Model capacity identifier	MSU	Model capacity identifier	MSU
401	28	501	53	601	65
402	54	502	104	602	127
403	78	503	152	603	184
404	102	504	197	604	240
405	124	505	240	605	292
406	144	506	279	606	339
407	164	507	317	607	385
408	182	508	352	608	428

### Capacity BackUp (CBU)

CBU deliver temporary backup capacity on top of what an installation might have installed in numbers of assigned CPs, IFLs, ICFs, zAAPs, zIIPs, and additional SAPs.

When CBU for CP is added within the same capacity setting range as the currently assigned PUs, the total number of active PUs (the sum of all assigned CPs, IFLs, ICFs, zAAPs, zIIPs, and additional SAPs) plus the number of CBU cannot exceed the total number of PUs available in the system.

When CBU for CP capacity is acquired by switching from one capacity setting to another, no more CBU can be requested than the total number of PUs available for that capacity setting.

**Restriction:** Activation of CBU for CPs, IFLs, ICFs, zAAPs, and zIIPs is mutually exclusive with On/Off Capacity on Demand activation. Both facilities may reside on one z9 EC, but cannot be activated simultaneously.

There is a distinction between five CBU types:

1. CBU for CP
  - CBU for CP7, feature code 7820
  - CBU for CP6, feature code 7819
  - CBU for CP5, feature code 7818
  - CBU for CP4, feature code 7817
2. CBU for IFL, feature code 7821

3. CBU for ICF, feature code 7822
4. CBU for zAAP, feature code 7824
5. CBU for zIIP, feature code 7825

**CBU and granular capacity**

Specialty engines (ICFs, IFLs, zAAPs, and zIIPs) always run at full capacity. When CBU for CP is needed, features 7820, 7819, 7818, or 7817 can be ordered to provide backup capacity.

CBU for CP is ordered by specifying both a number of CPs and a model capacity identifier. Table 2-15 shows the MSU values for subcapacity models.

*Table 2-15 Subcapacity Models MSUs*

	1-core	2-core	3-core	4-core	5-core	6-core	7-core	8-core
<b>40x</b>	28	54	78	102	124	144	164	182
<b>50x</b>	53	104	152	197	240	279	317	352
<b>60x</b>	65	127	184	240	292	339	385	428
<b>70x</b>	81	158	229	298	363	422	479	532

**CBU for CP rules**

The following should be taken into account when considering CBU for CP capacity:

- ▶ CBU does not necessarily increase the capacity of the server.
- ▶ CBU cannot decrease the number of CPs on the server.

When the CBU feature matches the model capacity identifier range of the permanent CP feature, the CBU activation results in the total number of CPs active equaling the number of permanent CPs plus the number of CPs added for CBU. For example, when a z9 EC with model capacity identifier 504 is ordered two CPs for CBU, the CBU configuration, when activated, becomes a z9 EC at capacity identifier 506 (see Example 2-1).

When the CBU configuration is activated, the server capacity will grow from 197 to 279 MSUs, as shown in Table 2-15.

*Example 2-1 Granular capacity with CBU for CPs in matching CI range*

---

4704	4-core Processor CP5	1
5504	504 Capacity Marker	1
7808	CP5	4
7818	CBU CP5	2 added to CP5 (becomes 506)

---

When the CBU results in a cross-over from one capacity identifier range to another, the CBU features no longer specifies an addition to the number of CPs, but rather the total number of CPs that will make up the configuration when the CBU is activated. The total number of CPs when the CBU is activated is equal to the number of CPs ordered for CBU.

For example, when a z9 EC with model capacity identifier 504 specifies a CBU with six CP6, the CBU server being activated will have model capacity identifier 606. When the CBU configuration is activated, the server capacity will grow from 197 to 339 MSUs (see Example 2-2).

*Example 2-2 Granular capacity with CP BUs in non-matching CI range*

4704	4-core Processor CP5	1
5504	504 Capacity Marker	1
7808	CP5	4
7819	CBU CP6	6 becomes model 606

A cross-over from one CP range to another does not necessarily imply that the number of CPs will increase for the CBU configuration. For example, a model capacity identifier 504 can have a CBU to model capacity identifier 704. In this case, there is no increase in the number of CPs. When the CBU configuration is activated, the server capacity will grow from 197 to 298 MSUs (see Example 2-3).

*Example 2-3 Granular capacity with CBU for CPs in non-matching CI range and no CP increase*

4704	4-core Processor CP5	1
5504	504 Capacity Marker	1
7808	CP5	4
7820	CBU CP7	4 becomes model 704

**Note:** CBU activation cannot decrease the number of active CPs in the system.

Addition of CBU for CPs beyond the 8-core limitation of each of the granular capacity settings can only make a full capacity server when in the capacity identifier range from 709 to 754. If a model capacity identifier 504 specifies ten CP7 CBU, when the CBU is activated, the server effectively becomes a full capacity 710.

*Example 2-4 Granular capacity with CBU for CPs beyond the eight CP limit*

4704	4-core Processor CP5	1
5504	504 Capacity Marker	1
7808	CP5	4
7820	CBU CP7	10 becomes model 710

**CBU for specialty engines**

Specialty engines, ICFs, IFLs, zAAPs, and zIIPs run at full capacity for all capacity settings. This also applies to CBU for specialty engines.

Table 2-16 shows the minimum and maximum numbers of all types of CBU.

Table 2-16 Capacity BackUp matrix

Model	Total PUs available	CBU CPs min - max	CBU IFLs min - max	CBU ICFs min - max	CBU zAAPs min - max	CBU zIIPs min - max
Model S08	8	1–8	0–7	0–7	0–4	0–4
Model S18	18	1–18	0–17	0–16	0–9	0–9
Model S28	28	1–28	0–27	0–16	0–14	0–14
Model S38	38	1–38	0–37	0–16	0–19	0–19
Model S54	54	1–54	0–53	0–16	0–27	0–27

Unassigned IFLs are ignored. They are considered spares and are available for use as CBU. When an unassigned IFL is converted into an assigned IFL, or when additional PUs are characterized as IFLs, then the number of CBU of any type that can be activated is decreased.

### On/Off Capacity on Demand and CPs

On/Off Capacity on Demand (On/Off CoD) provides temporary capacity for all types of characterized PUs. On/Off CoD for CPs relative to granular capacity is treated similar to the way CBU is handled.

#### On/Off CoD and granular capacity

When the temporary capacity requested by On/Off CoD for CPs matches the model capacity identifier range of the permanent CP feature, the total number of active CP equals the sum of the number of permanent CPs plus the number of temporary CPs ordered. For example, when a model capacity identifier 504 has two CP 5 added temporarily, it becomes a model capacity identifier 506.

When the addition of temporary capacity requested by On/Off CoD for CPs results in a cross-over from one capacity identifier range to another, the total number of CPs active when the temporary CPs are activated is equal to the number of temporary CPs ordered. For example, when a server with model capacity identifier 504 specifies six CP6 temporary CPs through On/Off CoD, the result is a server with model capacity identifier 606. A cross-over does not necessarily mean that the CP count for the additional temporary capacity will increase. The same 504 could temporarily be upgraded to a server with model capacity identifier 704. There is no increase in the number of CPs, but additional temporary capacity is achieved.

#### On/Off Capacity on Demand rules

The following should be taken into account when requesting temporary capacity.

- ▶ Temporary capacity must be greater than permanent capacity.
- ▶ Temporary capacity cannot be more than double the permanent capacity.
- ▶ On/Off CoD cannot decrease the number of engines on the server.
- ▶ It is not possible to add more engines than are currently owned.

Table 8-2 on page 223 shows all possible On/Off CoD CP upgrades for granular capacity models. For more information about temporary capacity increases, see Chapter 8, “Concurrent upgrades and availability” on page 205

## 2.4 Logical partitioning

Logical partitioning is a function implemented by the Processor Resource/Systems Manager (PR/SM), available on all z9 EC servers.

The z9 EC only runs in LPAR mode. This means that virtually all system aspects are controlled by PR/SM functions.

PR/SM is very much aware of the book structure on the z9 EC. However, logical partitions do not have this awareness. Logical partitions have resources allocated to them coming from a variety of physical resources, and have no control over these physical resources from a systems standpoint, but the PR/SM functions do.

PR/SM manages and optimizes allocation and dispatching work on the physical topology. Most physical topology that was previously handled by the operating systems is the responsibility of PR/SM.

PR/SM always attempts to allocate all real storage for a logical partition within one book, and attempts to dispatch a logical PU on a physical PU in a book that also has the Central Storage for that logical partition; if not possible, a PU in an adjacent book is chosen. In general, PR/SM tries to minimize the number of books required to allocate the resources of a given logical partition. In addition, PR/SM always tries to re-dispatch a logical PU on the same physical PU to assure that as much as possible of the L1 cache content can be reused.

PR/SM enables z9 EC servers to be initialized for a logically partitioned operation, supporting up to 60 logical partitions. Each logical partition can run its own operating system image in any image mode, independently from the other logical partitions.

A logical partition can be activated or deactivated at any time, but changing the number of defined or reserved logical partitions is disruptive, as it requires a Power-on Reset (POR). Some facilities may not be available to all operating systems, as they may have software corequisites.

Each logical partition has the same resources as a “real” CPC. They are:

► Processors

Called *logical processors*, they can be defined as CPs, IFLs, ICFs, zAAPs, or zIIPs. They can be dedicated to a logical partition or shared between logical partitions. When shared, a processor weight can be defined to provide the required level of processor resources to a logical partition. Also, the capping option can be turned on, which prevents a logical partition from acquiring more than its defined weight, limiting its processor consumption.

Logical partitions for z/OS can have CP, zAAP, and zIIP logical processors. All three logical processor types can be defined as either all dedicated or all shared. The zAAP and zIIP support is available in z/OS V1.6 and later.

Only Coupling Facility (CF) partitions can have both dedicated *and* shared logical processors defined.

Figure 2-17 on page 80 shows the logical processor assignment window of the *Customize Image Profile* on the Hardware Management Console. The panel allows the definition of:

- Dedicated or shared logical processors, including CPs, zAAPs, and zIIPs; remember that zAAPs appear as *integrated facility for applications (IFA)* on the HMC panels.
- The Initial weight, capping option, Enable workload manager option, and minimum and maximum processing weight for shared CPs, zAAPs, and zIIPs.
- The optional group profile name the logical partition is assigned to.

- The number of initial and optional reserved CPs, zAAPs, and zIIPs.
- The sum of defined and reserved logical processors in a logical partition is limited to 54. z/OS supports up to 32 logical processors in a logical partition; the limit applies to the sum of CP, zAAP, and zIIP logical processors. z/VM V5.3 supports up to 32 processors.

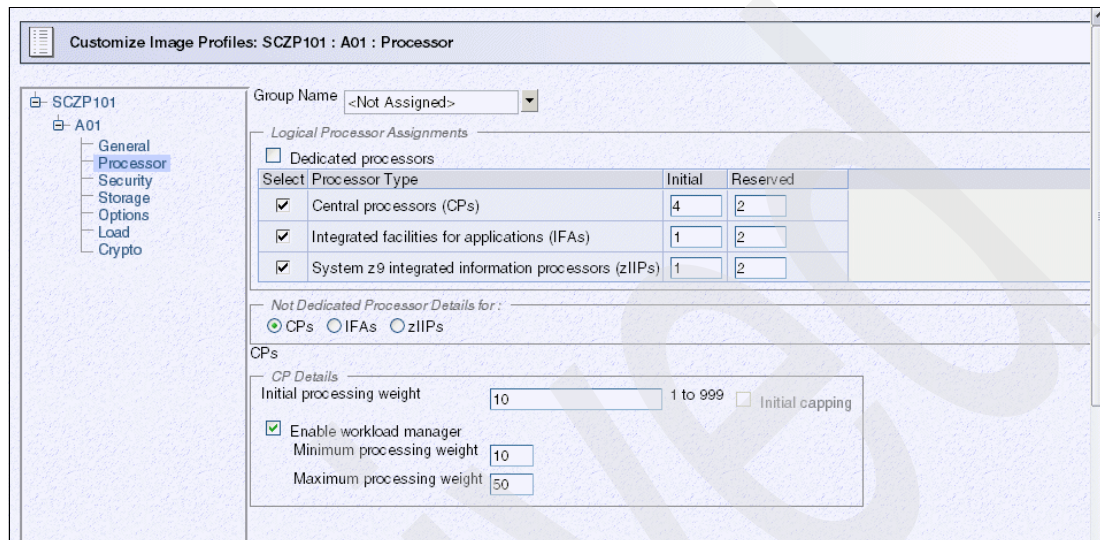


Figure 2-17 Customize Image Profile - Processor page

The weight and the number of online logical processors of a logical partition can be dynamically managed by the LPAR CPU Management function of the Intelligent Resource Director, to achieve the defined goals of this specific partition and of the overall system.

For z/OS Workload License Charge (WLC), a logical partition *Defined Capacity* can be set, enabling the soft capping function. Workload charging introduces the capability to pay software license fees based on the size of the logical partition the product is running in, rather than on the total capacity of the server.

- In support of WLC, the user can specify a defined capacity in millions of service units per hour (MSUs). The Defined capacity sets the capacity of an individual logical partition when soft capping is selected.

The Defined capacity value is specified on the Options tab on the *Customize Image Profile* panel.

- WLM keeps a four-hour rolling average of the CPU usage of the logical partition, and when the 4-hour average CPU consumption exceeds the defined capacity limit, WLM dynamically activates LPAR capping (soft-capping). When the rolling 4-hour average returns below the defined capacity, the soft-cap is removed.

For more information regarding WLM, refer to *System Programmer's Guide to: Workload Manager*, SG24-6472.

#### ► Memory

Memory, either Central Storage or Expanded Storage, must be dedicated to a logical partition. The defined storages must be available during the logical partition activation; otherwise, the activation fails.

*Reserved* storage can be defined to a logical partition, enabling nondisruptive memory add to and removal from a logical partition, using the LPAR Dynamic Storage Reconfiguration. Refer to 2.5.3, “LPAR Dynamic Storage Reconfiguration (DSR)” on page 87 for more information.

► Channels

Channels can be shared between logical partitions by including the partition name in the partition list of a Channel Path ID (CHPID). I/O configurations are defined by the I/O Configuration Program (IOCP) or the Hardware Configuration Dialog (HCD) in conjunction with the CHPID Mapping Tool (CMT). The CMT is an optional, but strongly recommended, tool used to map CHPIDs onto Physical Channel IDs (PCHIDs) that represent the physical location of a port on a card in an I/O cage.

IOCP is available on the z/OS, z/VM, VM/ESA®, z/VSE, and VSE/ESA operating systems, and as a stand-alone program on the z9 EC hardware console. HCD is available on z/OS and z/VM operating systems.

ESCON channels (CHPID type CNC or FCV) can be *managed* by the Dynamic CHPID Management (DCM) function of the Intelligent Resource Director. DCM enables the system to respond to ever-changing channel requirements by moving channels from lesser-used control units to more heavily used control units, as needed.

**Modes of operation**

Table 2-17 shows the z9 EC modes of operation, summarizing all available mode combinations: operating modes and their processor types, operating systems, and addressing modes.

There is no special operating mode for the 64-bit z/Architecture mode, as the architecture mode is not an attribute of the definable images operating mode. The 64-bit operating systems are IPLed in 31-bit mode and, optionally, can change to 64-bit mode during their initialization. It is up to the operating system to take advantage of the addressing capabilities provided by the architectural mode.

Table 2-17 z9 EC modes of operation

Image mode	PU type	Operating system	Addressing mode
ESA/390 mode	CP <i>and</i> zAAP/zIIP (z/OS 1.6 and later)	z/OS z/VM	64-bit Architecture mode
	CP	Linux on System z	64-bit Architecture mode
	CP	z/VSE, VSE/ESA, and Linux on System z	31-bit Architecture mode
ESA/390 TPF mode	CP <i>only</i>	TPF	31-bit Architecture mode
Coupling Facility mode	ICF <i>and/or</i> CP	CFCC	64-bit Architecture mode
Linux only mode	IFL <i>or</i> CP	Linux on System z	64-bit Architecture mode
		z/VM	
		Linux on S/390	31-bit Architecture mode

Information about the operating systems supported on z9 EC is in Chapter 6, “Software support” on page 165.

## Logically partitioned mode

The z9 EC server can only run in LPAR Mode; up to 60 logical partitions can be defined on a z9 EC server. A logical partition can be defined to operate in one of the following image modes:

- ▶ ESA/390 mode, to run:
  - A z/Architecture operating system, on dedicated *or* shared CPs
  - An ESA/390 operating system, on dedicated *or* shared CPs
  - A Linux operating system, on dedicated *or* shared CPs
  - A z/OS V1.6 or later operating system, on any of the following:
    - Dedicated *or* shared CPs
    - Dedicated CPs *and* dedicated zAAPs *or* zIIPs
    - Shared CPs *and* shared zAAPs *or* zIIPs

**Note:** zAAPs and zIIPs can be defined to any ESA/390 mode image (see Table 2-17 on page 81). However, zAAPs, and zIIPs are supported only by z/OS V1.6 and later. Other operating systems cannot use zAAPs or zIIPs, even if they are defined to the logical partition. z/VM V5.3 can provide zAAPs or zIIPs to a guest z/OS.

- ▶ ESA/390 TPF mode, to run a TPF operating system, on dedicated *or* shared CPs
- ▶ Coupling Facility mode, by loading the CFCC code into the logical partition. These can be defined as:
  - Dedicated *or* shared CPs
  - Dedicated *or* shared ICFs
  - Dedicated *and* shared ICFs
  - ICFs dedicated *and* CPs shared
- ▶ Linux-only mode, to run:
  - A Linux operating system, on either:
    - Dedicated *or* shared IFLs
    - Dedicated *or* shared CPs
  - A z/VM operating system, on either:
    - Dedicated *or* shared IFLs
    - Dedicated *or* shared CPs



Table 2-18 shows all LPAR modes, required characterized PUs, and operating systems, and which PU characterizations can be configured to a logical partition image. The available combinations of dedicated (DED) and shared (SHR) processors are also shown. For all combinations, a logical partition can also have Reserved Processors defined, allowing nondisruptive logical partition upgrades.

Table 2-18 LPAR mode and PU usage

LPAR mode	PU type	Operating systems	PUs usage
ESA/390	CPs	z/Architecture operating systems ESA/390 operating systems Linux	CPs DED <i>or</i> CPs SHR
	CPs <i>and</i> zAAPs <i>or</i> zIIPs	z/OS (V1.6 and up) z/VM (V5.3 and up for guest exploitation)	CPs DED <i>and</i> zAAPs DED, <i>and/or</i> zIIPs DED <i>or</i> CPs SHR <i>and</i> zAAPs SHR <i>or</i> zIIPs SHR
ESA/390 TPF	CPs	TPF	CPs DED <i>or</i> CPs SHR
Coupling Facility	ICFs <i>and/or</i> CPs	CFCC	ICFs DED <i>or</i> ICFs SHR, <i>or</i> CPs DED <i>or</i> CPs SHR, <i>or</i> ICFs DED <i>and</i> ICFs SHR, <i>or</i> ICFs DED <i>and</i> CPs SHR
Linux Only	IFLs <i>or</i> CPs	Linux z/VM	IFLs DED <i>or</i> IFLs SHR, <i>or</i> CPs DED <i>or</i> CPs SHR

### Dynamic add or delete of a logical partition name

Dynamic add or delete of a logical partition name is the ability to add meaningful logical partition names or required I/O resources in advance to the configuration without a Power-On Reset. Prior to this support, extra logical partitions were defined by adding reserved names in the Input/Output Configuration Data Set (IOCDs), but one may not have been able to predict what might be meaningful names in advance.

Dynamic add or delete of a logical partition name allows reserved logical partition *slots* to be created in an IOCDs in the form of extra Channel Subsystem (CSS), Multiple Image Facility (MIF) image ID pairs. A reserved partition is defined with the partition name placeholder “\*”, and cannot be assigned to an access or candidate list of channel paths.

By default, reserved logical partitions are included in device candidate lists. So, when defining a new logical partition name, if certain devices should be excluded from the new logical partition, there must be necessary device candidate list changes.

The extra Channel Subsystem, MIF image ID pairs (CSSID/MIFID) can be later assigned a logical partition name for use (or later removed) through dynamic I/O commands using the Hardware Configuration Definition (HCD); at the same time, required channels will need to be defined for the new logical partition. The IOCDs still must have the extra I/O slots defined in advance because many structures are built based upon these major I/O control blocks in the Hardware System Area (HSA).

This support is exclusive to the z9 EC, z990, and z890 and is applicable to z/OS V1.6 and later.

When a logical partition is renamed, its name can be changed from “NAME1” to “\*” and then changed again from “\*” to “NAME2”; the logical partition number and MIFID are retained across the logical partition name change. However, the master keys in a Crypto Express2

feature that were associated with the old logical partition 'NAME1' are retained. There is no explicit action taken against a cryptographic component for this.

**Attention:** Cryptographic coprocessors are not tied to partition numbers or MIF IDs. They are set up with AP numbers and domain indices. These are assigned to a partition profile of a given name. The customer assigns these “lanes” to the partitions and continues to have the responsibility to clear them out if he changes who is using them.

### LPAR group capacity limit

The group capacity limit feature on the System z9 allows the definition of a logical partition group capacity limit on System z9 servers. This function is designed to allow a capacity limit to be defined for each logical partition running z/OS or z/OS.e, and to define a group of logical partitions on a server. This is expected to allow the system to manage the group in such a way that the sum of the LPAR group capacity limits in MSUs per hour will not be exceeded. To take advantage of this, the customer needs to be running z/OS V1.8 and all logical partitions in the group have to be at z/OS V1.8 and higher.

PR/SM and WLM work together to enforce the capacity defined for the group and enforce the capacity optionally defined for each individual logical partition.

## 2.5 Storage operations

In z9 EC servers, memory can be assigned as a combination of Central Storage and Expanded Storage, supporting up to 60 logical partitions.

Before activating a logical partition, Central Storage (and optional Expanded Storage) must be defined to the logical partition. All installed storage can be configured as Central Storage. Each individual logical partition can be defined with a maximum of 128 GB of Central Storage.

Central Storage can be dynamically assigned to Expanded Storage and back to Central Storage as needed without a Power-on Reset (POR); see “LPAR single storage pool” on page 70 for further details.

Memory *cannot* be shared between system images. It is possible to dynamically reallocate storage resources for z/Architecture logical partitions running operating systems that support Dynamic Storage Reconfiguration (DSR). Refer to 2.5.3, “LPAR Dynamic Storage Reconfiguration (DSR)” on page 87 for further details.

Operating systems running under z/VM can exploit the z/VM capability of implementing virtual memory to guest virtual machines. The z/VM dedicated *real* storage can be “shared” between guest operating systems.

Table 2-19 shows the z9 EC storage *allocation* and *usage* possibilities, depending on the image mode.

Table 2-19 Storage definition and usage possibilities

Image mode	Architecture mode (addressability)	Maximum Central Storage		Expanded Storage	
		Architecture	z9 EC definition	z9 EC definable	Operating system usage
ESA/390	z/Architecture (64-bit)	16 EB	128 GB	Yes	Only by z/VM
	ESA/390 (31-bit)	2 GB	128 GB	Yes	Yes
ESA/390 TPF	ESA/390 (31-bit)	2 GB	128 GB	Yes	Yes
Coupling Facility	CFCC (64-bit)	16 EB	128 GB	No	No
Linux Only	z/Architecture (64-bit)	16 EB	128 GB	Yes	Only by z/VM
	ESA/390 (31-bit)	2 GB	128 GB	Yes	Yes

Remember that either a z/Architecture mode or an ESA/390 architecture mode operating system can run in an ESA/390 image on a z9 EC. Any ESA/390 image can be defined with more than 2 GB of Central Storage *and* can have Expanded Storage. These options allow you to configure more storage resources than the operating system is capable of addressing.

### ESA/390 mode

In ESA/390 mode, storage addressing can be 31-bits or 64-bits, depending on the operating system architecture *and* the operating system configuration.

An ESA/390 mode image is always initiated in 31-bit addressing mode. During its initialization, a z/Architecture operating system can change it to 64-bit addressing mode and operate in the z/Architecture mode.

Some z/Architecture operating systems, like z/OS, will *always* change this addressing mode and operate in 64-bit mode. Other z/Architecture operating systems, like z/VM, can be configured to change to 64-bit mode or to stay in 31-bit mode and operate in the ESA/390 architecture mode.

► z/Architecture mode

In z/Architecture mode, storage addressing is 64-bit, allowing for an addressing range of up to 16 exabytes (16 EB). The 64-bit architecture allows a maximum of 16 EB to be used as Central Storage. However, the current z9 EC limit for logical partitions is 128 GB of Central Storage.

Expanded Storage *can* also be configured to an image running an operating system in z/Architecture mode. However, only z/VM is able to use Expanded Storage. Any other operating system running in z/Architecture mode (like a z/OS or a Linux on System z image) *will not* address the configured Expanded Storage. This Expanded Storage remains configured to this image and is *unused*.

► ESA/390 architecture mode

In ESA/390 architecture mode, storage addressing is 31-bit, allowing for an addressing range of up to 2 GB. A maximum of 2 GB can be used for Central Storage. Since the processor storage can be configured as central and Expanded Storage, memory above 2 GB may be configured as Expanded Storage. In addition, this mode permits the use of either 24-bit or 31-bit addressing, under program control.

Since an ESA/390 mode image can be defined with up to 128 GB of Central Storage, the Central Storage above 2 GB will *not* be used, but remains configured to this image.

### **ESA/390 TPF mode**

In ESA/390 TPF mode, storage addressing follows the ESA/390 architecture mode; the TPF/ESA operating system runs in the 31-bit addressing mode.

### **Coupling Facility mode**

In Coupling Facility mode, storage addressing is 64-bit for a Coupling Facility image running CFLEVEL 12 or above (the z9 EC comes with CFLEVEL 14), allowing for an addressing range up to 16 EB. However, the current z9 EC definition limit for logical partitions is 128 GB of storage.

Expanded Storage cannot be defined for a Coupling Facility image. Only IBM Coupling Facility Control Code can run in Coupling Facility mode.

### **Linux Only mode**

In Linux Only mode, storage addressing can be 31-bit or 64-bit, depending on the operating system architecture *and* the operating system configuration, in exactly the same way as in ESA/390 mode.

Only Linux and z/VM operating systems can run in Linux Only mode:

- ▶ Linux on System z uses 64-bit addressing and operates in the z/Architecture mode.
- ▶ Linux for S/390 uses 31-bit addressing and operates in the ESA/390 Architecture mode.
- ▶ z/VM uses 64-bit addressing and operates in the z/Architecture mode.

## **2.5.1 Reserved storage**

Reserved storage can optionally be defined to a logical partition, allowing a nondisruptive image memory upgrade for this partition. Reserved storage can be defined to both central and Expanded Storage, and to any image mode, except in Coupling Facility mode.

A logical partition must define an amount of Central Storage and, optionally (if not a Coupling Facility image), an amount of Expanded Storage. Both central and Expanded Storages can have two storage sizes defined: An initial value and a reserved value.

- ▶ The initial value is the storage size allocated to the partition when it is activated.
- ▶ The reserved value is an additional storage capacity beyond its initial storage size that a logical partition can acquire dynamically. The reserved storage sizes defined to a logical partition do not have to be available when the partition is activated. They are just predefined storage sizes to allow a storage increase from the logical partition point of view.

Without the reserved storage definition, a logical partition storage upgrade is disruptive, requiring:

- ▶ Partition deactivation
- ▶ An initial storage size definition change
- ▶ Partition activation

The additional storage capacity to a logical partition upgrade can come from:

- ▶ Any unused available storage
- ▶ Another partition that has released some storage
- ▶ A concurrent memory upgrade

A concurrent logical partition storage upgrade uses Dynamic Storage Reconfiguration (DSR) and the operating system must use the Reconfigurable Storage Unit (RSU) definition to be able to add or remove storage units in a nondisruptive way. Currently, only z/OS has this support.

## 2.5.2 Logical partition storage granularity

Granularity of Central Storage for a logical partition is dependent on the largest Central Storage amount defined for either initial or reserved Central Storage, as shown in Table 2-20.

Table 2-20 Logical partition storage granularity

Logical partition largest Central Storage amount	Logical partition Central Storage granularity
Central Storage amount <= 32 GB	64 MB
32 GB < Central Storage amount <= 64 GB	128 MB
64 GB < Central Storage amount <= 128 GB	256 MB
128 GB < Central Storage amount <= 256 GB	512 MB
256 GB < Central Storage amount <= 512 GB	1024 MB

The granularity applies across all Central Storage defined, both initial and reserved. For example, for an LP with an initial storage amount of 30 GB and a reserved storage amount of 48 GB, the Central Storage granularity of both initial and reserved Central Storage is 128 MB.

Expanded Storage granularity is fixed at 64 MB.

Logical partition storage granularity information is required for logical partition image setup and for z/OS Reconfigurable Storage Units definition. Remember that logical partitions for z/OS and z/VM are currently limited to a maximum size of 128 GB of Central Storage.

## 2.5.3 LPAR Dynamic Storage Reconfiguration (DSR)

Dynamic Storage Reconfiguration (DSR) on z9 EC servers allows an operating system running in a logical partition to add (nondisruptively) its reserved storage amount to its configuration, if any unused storage exists. This unused storage can be obtained when another logical partition releases some storage, or when a concurrent memory upgrade takes place.

With DSR, the unused storage does not have to be continuous.

When an operating system running in a logical partition assigns a storage increment to its configuration, PR/SM will check if there are any free storage increments and will dynamically bring the storage online.

PR/SM will dynamically take offline a storage increment and will make it available to other partitions when an operating system running in a logical partition releases a storage increment.

LPAR Dynamic Storage Reconfiguration is described in detail in the manual *System z9 Processor Resource/Systems Manager Planning Guide*, SB10-7041.

Archived

## I/O system structure

This chapter describes the I/O system structure, the connectivity, and the cryptographic options available on the IBM System z9.

The z9 EC server I/O and cryptographic features are also discussed, including configuration options for each feature.

The following topics are included:

- ▶ 3.1, “Overview” on page 90
- ▶ 3.2, “I/O cages” on page 91
- ▶ 3.3, “I/O and cryptographic feature cards” on page 97
- ▶ 3.4, “Connectivity” on page 101

## 3.1 Overview

The z9 EC I/O system design provides great flexibility, high availability, and performance, allowing:

- ▶ High bandwidth: Individual channels and ICB-4 Coupling Links can have up to 2.0 GB per second data rate.
- ▶ Wide connectivity: A z9 EC can be connected to an extensive range of interfaces, using protocols such as Fibre Channel Protocol (FCP) for Small Computer System Interface (SCSI), Gigabit Ethernet (GbE), 1000BASE-T Ethernet, 100BASE-T Ethernet, 10BASE-T Ethernet, along with FICON, ESCON, and Coupling Link channels.
- ▶ Cryptographic functions: The z9 EC I/O system also supports optional cryptographic cards to complement the standard CP Assist for Cryptographic Function (CPACF) that is implemented in every PU, enhancing the performance and functionality of cryptographic processing.
- ▶ Concurrent I/O upgrades: It is possible to concurrently add I/O cards to a z9 EC server provided there are unused slot positions in an I/O cage. Additional I/O cages can be installed in advance through CUoD to provide greater capacity for concurrent upgrades. This capability may help eliminate an outage to upgrade the I/O configuration. See more information about concurrent upgrades in Chapter 8, “Concurrent upgrades and availability” on page 205.
- ▶ Dynamic I/O configuration: Dynamic I/O configuration enhances system availability by supporting the dynamic addition, removal, or modification of channel paths, control units, I/O devices, and I/O configuration definitions to both hardware and software (if it has this support) without requiring a planned outage.
- ▶ ESCON port sparing and upgrading: The ESCON 16-port I/O card includes one unused port dedicated for sparing in the event of a port failure on that card. Other unused ports are available for growth of ESCON channels without requiring new hardware, enabling concurrent upgrades through Licensed Internal Code (LIC-CC).

### Summary of I/O feature support

The following I/O feature *ports* are supported on the z9 EC server:

- ▶ Up to 1024 ESCON (up to 960 ESCON on Model S08).
- ▶ Up to 120 FICON Express (up to 64 FICON on Model S08), carried forward on an upgrade only.
- ▶ Up to 336 FICON Express2 (up to 240 FICON on Model S08), carried forward on an upgrade only.
- ▶ Up to 336 FICON Express4 (up to 240 FICON on Model S08).
- ▶ Up to 48 OSA Express, carried forward on an upgrade only.
- ▶ Up to 48 OSA Express2.
- ▶ Up to 16 Integrated Cluster Bus-4 (ICB-4).
- ▶ Up to 16 Integrated Cluster Bus-3 (ICB-3).
- ▶ Up to 48 Inter-System Channel-3 (ISC-3) in peer mode only.
- ▶ Up to two External Time Reference (ETR).
- ▶ Up to eight Crypto Express2 features with two PCI-X cryptographic adapters each. Each adapter can be configured as a cryptographic coprocessor or cryptographic accelerator.



**Note:** The maximum number of Coupling Links combined (IC, ICB-3, ICB-4, and active ISC-3 links) cannot exceed 64 per server.

The z9 EC has two frames. The A frame holds the CEC cage on top and one I/O cage on the bottom. The Z frame holds up to two optional I/O cages, which may be needed to accommodate the I/O configuration requirements.

Additional optional I/O cages may be required to install additional I/O and cryptographic cards during an upgrade. The first optional I/O cage is placed at the bottom of the Z frame, and the second optional I/O cage is at the top, as shown in Figure 2-8 on page 40. An I/O cage installation requires an outage.

### 3.2 I/O cages

As mentioned, the z9 EC server can have up to three I/O cages to host the I/O and cryptographic cards required by a configuration.

Each I/O cage has 28 I/O slots for I/O and cryptographic cards and supports up to seven I/O domains. Each I/O domain is made up of up to four I/O slots, as shown in Figure 3-1.

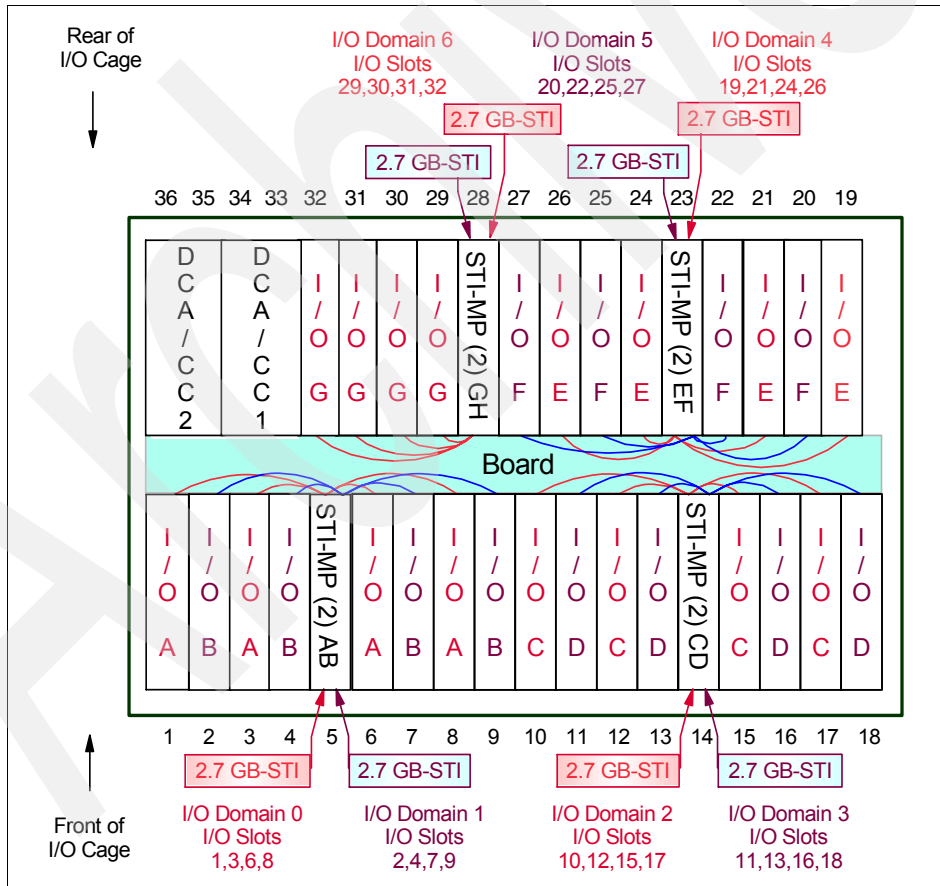


Figure 3-1 z9 EC I/O cage

Each I/O domain requires one Self-Timed Interconnect Multiplexer (STI-MP) card. All I/O cards within an I/O domain are connected to its STI-MP card through the back plane board. A full I/O cage requires eight STI-MP cards, which are half-high cards, using four slots. In addition, two Distributed Converter Assembly-Cage Controller (DCA-CC) cards plug into the I/O cage.

If one I/O domain is fully populated with ESCON cards (each with 15 active ports and one spare per card), up to 60 (four cards x 15 ports) ESCON channels can be installed and used. An I/O cage with six domains fully populated with ESCON cards has 360 (60 x 6 domains) ESCON channels. Table 3-1 lists the I/O domain-to-I/O slot relationships within an I/O cage.

Table 3-1 I/O domain-to-I/O slot relationships

Domain	I/O slots in domain
0	01, 03, 06, 08
1	02, 04, 07, 09
2	10, 12, 15, 17
3	11, 13, 16, 18
4	19, 21, 24, 26
5	20, 22, 25, 27
6	29, 30, 31, 32

Each STI-MP card is connected to an STI jack (J00 and J01) located in a book's Memory Bus Adapter (MBA) fan out card through an STI cable. As each STI-MP card requires one STI, up to eight STIs are required to support one I/O cage.

The configuration process selects which slots are used for I/O cards and supplies the appropriate number of I/O cages and STI cables, either for a new build server, or for a server upgrade.

**Important:** Installing an additional I/O cage to an existing z9 EC server configuration is disruptive. The Plan Ahead process allows avoiding this outage by including, in the initial z9 EC order, the number of additional I/O features needed by a future I/O configuration. This will ensure there are enough I/O slots/cages available.

Domain 6 is not used for I/O cards until all other domains in all three cages are full. A new cage is added when more than 24 or 48 I/O cards need to be installed. STI-3 cards use domain 6, and when more than four STI-3 cards are needed, a new cage is added. STI-3 and PSC24V (always in slot 29) cards are plugged in domain 6.

**Note:** The PSC24V card is a power sequence control card used to turn on, or off, selected control units from the server. The card provides the physical interface between the cage controller and the PSC boxes in the frame. The PSC24V card has two jacks for PSC connection and its installation is disruptive.

### 3.2.1 Self-Timed Interconnect (STI)

There are up to eight Memory Bus Adapters (MBAs) fanout cards on each z9 EC book. The MBA fanout cards are numbered D1 to D8 and each have two Self-Timed Interconnect (STI) jacks, resulting in a total of 16 STI connections on each z9 EC book. Each STI has a bandwidth of 2.7 GB per second full-duplex, resulting in a maximum bandwidth of 43.2 GB per second per book.

Depending on the number of books in the configuration, there can be up to 16, 32, 48, or 64 STIs in a z9 EC server, as shown in Table 3-2. The plugging sequence for all MBA fanout cards is D8, D5, D7, D4, D6, D2, D3, and D1.

Table 3-2 Number of MBAs and STIs

z9 EC Model	Number of books	Maximum number of MBA fanout cards	Number of STI connections
S08	1	8	16
S18	2	16	32
S28	3	24	48
S38	4	32	64
S54	4	32	64

### 3.2.2 STIs and I/O cage connections

Figure 3-2 shows the STI connections from the server CEC cage to an I/O cage, and to an Integrated Cluster Bus-4 (ICB-4) link.

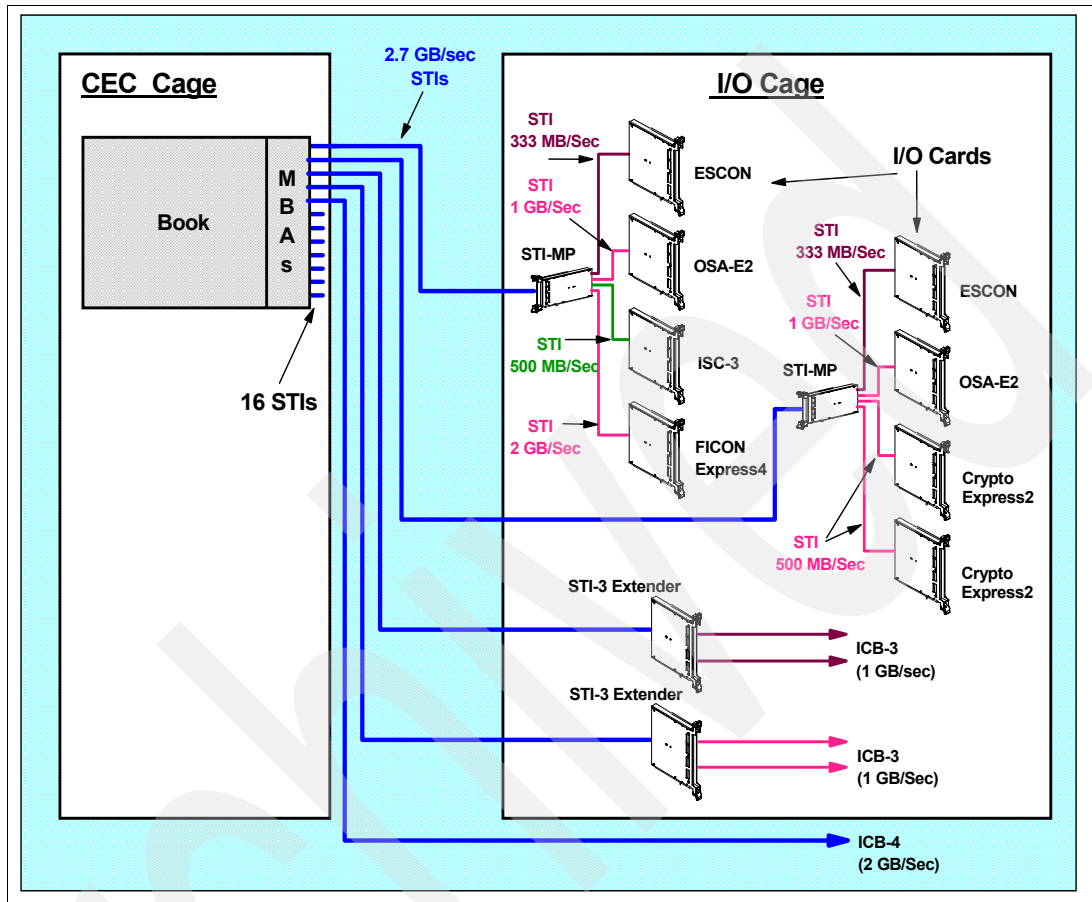


Figure 3-2 STIs and I/O cage connections

A Memory Bus Adapter (MBA) STI connector, located in a book, can be connected to one of the following:

- ▶ An STI-MP card, which creates up to four secondary STI links to connect I/O cards
- ▶ An STI-3 Extender card, which has up to two ICB-3 links
- ▶ An ICB-4 link, which attaches directly to an STI port

#### STI-MP card

For each I/O cage domain, the MBA-to-I/O card connectivity is achieved using STI-MP daughter cards that are plugged into a STI-A8 or STI-A4 mother card. The STI-MP daughter cards are connected to a 2.7 GB per second STI cable. The STI-A8 cards plug into specific slots (5, 14, and 23) in the I/O cage. Slot locations 5, 14, 23, and 28 have two STI-MP cards, while only slot 28 uses the STI-A4 mother card.

Since the STI-MP cards in slot 28 only serve one domain, a slightly different mother card (STI-A4) is used. The STI-A4 mother card hosts two STI-MP cards and only has one connector (serving domain 6) to the planar board of the I/O cage. This way Redundant I/O Interconnect also applies.

This card setup provides the capability for Redundant I/O Interconnect, assuring connection to I/O devices even when a book, including its MBA/STI connections, is removed. Connection to I/O resources that would have been disconnected (after a book or MBA fanout removal for upgrade or repair) is maintained by using the STI connection from a different book or MBA fanout. The I/O interconnect is supported on the STI-A8 and STI-A4 cards, as can be seen in Figure 3-3. This is an exclusive for z9 EC.

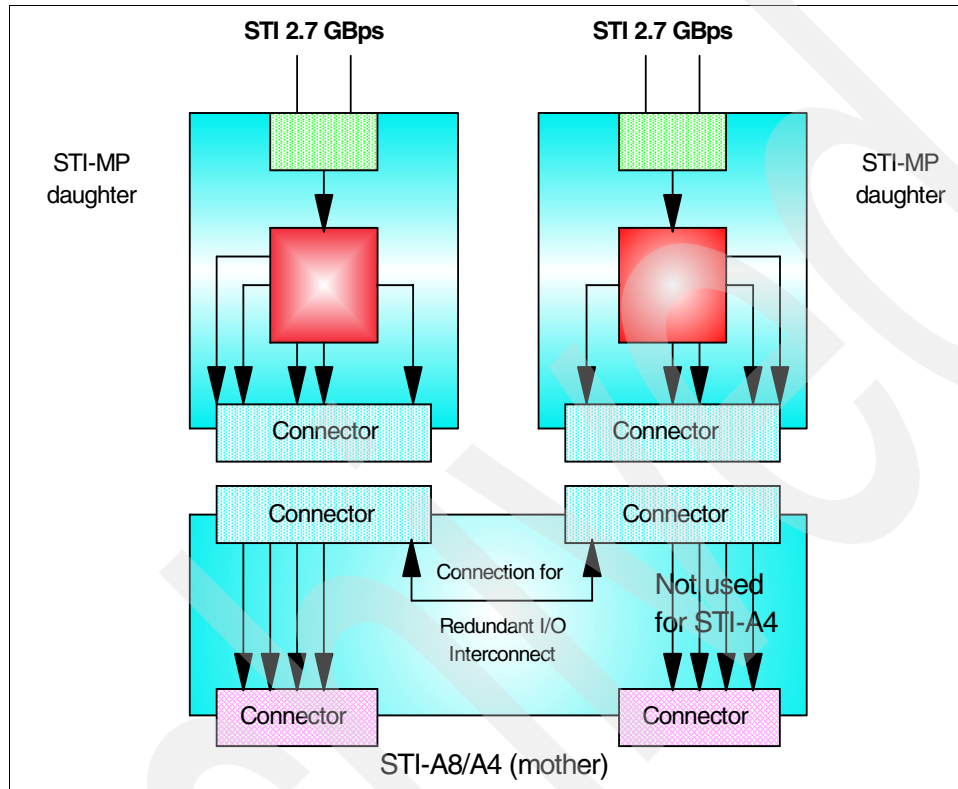


Figure 3-3 STI-MP, STI-A8, and STI-A4 cards

The STI-MP card (FC 0325) takes the 2.7 GB per second link directly from an MBA's STI and creates four secondary STI links, through the STI-A8 or STI-A4 card, which are connected to the I/O and cryptographic cards through the I/O cage board. The bandwidth of the secondary link is determined by the feature card it is attached to:

- ▶ 333 MB per second for ESCON
- ▶ 500 MB per second for Crypto Express2, FICON Express, ISC-3, and OSA-Express
- ▶ 1 GB per second for FICON Express2 and OSA Express2
- ▶ 2 GB per second for FICON Express4

Depending on the number of I/O slots plugged into the cage, there may be from one to eight STI-MP cards plugged into a z9 EC I/O cage. The STI-MP card can be installed or replaced concurrently.

### STI-3 Extender card

The STI-3 Extender card (FC 3993) takes the 2.0 GB per second link directly from an MBA/STI interface and creates two secondary 1 GB per second STI links, which are used to connect ICB-3 links.

The number of STI-3 Extender cards depends on the number of ICB-3 links in a configuration. Usually, the number of STI-3 Extender cards is half the number of ICB-3 links, but for availability reasons, two ICB-3 links are connected to two STI-3 Extender cards, each one having one active ICB-3 link port.

The maximum number of STI-3 Extender cards in a z9 EC server is eight cards, resulting in up to 16 ICB-3 ports. The STI-3 Extender card can be installed and replaced concurrently and is plugged in domain 6 of the I/O cage.

### 3.2.3 Balancing I/O connections

I/O distribution across books, MBAs, STIs, I/O cages, and I/O cards is desirable for both performance and availability purposes.

The STI link balancing across MBAs, I/O cages, and I/O cards is done when a configuration is created. Follow-on upgrades of the initial server configuration, including additional books or I/O cages, may undo the balance of the original STI link distribution.

I/O ports balancing to devices, networks, etc. across I/O cards, I/O cages, STI links, and MBAs is done when the configuration is set up. This is done by either the use of the CHPID Mapping Tool (CMT) to assign CHPIDs to PCHIDs, or manually by assigning installed PCHIDs to CHPIDs. The use of the CHPID Mapping tool is strongly recommended.

#### **STI links balancing across books and MBAs**

A z9 EC Model S18 has two books. The STI links are distributed across books, MBAs, and I/O cages, as a result of the initial server configuration. Nearly the same number of STIs of each book's MBA are used and spread across the redundant domains of each I/O cage, resulting in the best STI links distribution for both performance and availability.

Upgrading the z9 EC Model S18 to a z9 EC Model S38 adds, concurrently, two books. The upgrade automatically includes rebalancing of the MBA/STI connections to the I/O cards across all books, except for ICBs. This is made possible with the Redundant I/O Interconnect facility in the STI-A8 and STI-A4 cards, as shown in Figure 3-3 on page 95. Redundant I/O Interconnect allows maintaining connections to devices, even when an MBA fanout is temporarily disconnected in order to be moved to another slot in an other book to maintain a balanced configuration. See "Redundant I/O Interconnect" on page 38 for more information.

#### **STI Rebalance feature**

To optimize Reliability, Availability, and Serviceability (RAS) characteristics of the server, the STI Rebalance feature (FC 2400) can be ordered to rebalance MBA fanout cards used for ICB connections, on upgrades that include additional books. Use of feature code 2400 is disruptive, requiring a server outage.

The STI rebalance feature is recommended on any model upgrade from a S08, and should be considered on all other model upgrades (one point to evaluate in an imbalanced STI configuration is the risk of losing all ICBs on any of the books for a period of time).

When the STI Rebalance feature is specified when the server is upgraded from a z9 EC Model S18 to a Model S38, the STI links for ICB-4 and STI-3 extender cards (ICB-3) are spread across all books' MBAs, including the two newly installed books. The result is a balanced STI system, as though a new build z9 EC Model S38 server was initially configured.

**Important:** If the z9 EC STI Rebalance feature (FC 2400) is selected when the server is upgraded, this results in STI rebalancing for ICBs. The z9 EC STI Rebalance feature will change the Physical Channel ID (PCHID) number of ICB-4 links requiring a corresponding update on the server I/O definition through HCD/HCM. See 3.3.2, “Physical Channel IDs (PCHIDs)” on page 98.

### I/O port balancing across MBAs and books

When the I/O configuration is defined (using HCD or IOCP), the installation is able to select I/O ports for different paths of a multi-path control unit that come from different I/O cards, different I/O domains (including different STI-MP cards and different STI links), different I/O cages, and different MBAs from different books. This improves I/O throughput and system availability by avoiding single-point-of-failure paths.

Of course, this example assumes that there are enough I/O cards available for such connectivity distribution, and this may not be true for all channel types on a given real configuration. However, the overall goal is to avoid, as much as possible, connectivity single-points-of-failure.

The z9 EC CHPID Mapping Tool (CMT) can help to plan for the best I/O port selection for high availability purposes. For more information about the CMT, see Appendix B, “CHPID mapping tool” on page 263.

## 3.3 I/O and cryptographic feature cards

I/O cards have ports to connect the z9 EC to external devices, networks, or to other servers. I/O cards are plugged into I/O slots in an I/O cage, and their specific locations are based on z9 EC configuration rules. There are different types of I/O cards, one for each channel or link type. I/O cards can be installed or replaced concurrently.

Optional Crypto Express2 features are also plugged into an I/O slot in an I/O cage. They carry coprocessors and accelerator cryptographic functions. OSA Express2 features can be installed or replaced concurrently.

### 3.3.1 I/O feature cards

Table 3-3 gives a summary of all I/O feature cards that can be ordered on new build z9 EC servers.

Table 3-3 z9 EC I/O feature cards

I/O card types	Feature codes (FC)
ESCON	2323 for the 16 port card and 2324 for each 4 channels purchased
FICON Express4 10KM LX	3321
FICON Express4 4KM LX	3324
FICON Express4 SX	3322
OSA-Express2 GbE LX	3364
OSA-Express2 GbE SX	3365
OSA-Express2 1000BASE-T Ethernet	3366

I/O card types	Feature codes (FC)
OSA-Express 2 10 GbE LR	3368
ISC-3	0218 (ISC-D), 0217 (ISC-M)
ISC-3 up to 20 km	RPQ 8P2197 (ISC-D)
ETR	6155

### I/O cards carried forward with an upgrade

The following features can be carried forward to the z9 EC when upgrading from a z900 or z990.

- ▶ FICON Express LX (FC 2319) and SX (FC 2320).
- ▶ FICON Express2 LX (FC 3319) and SX (FC 3320).
- ▶ OSA-Express GbE LX (FC 2364) and SX (FC 2365).
- ▶ OSA-Express GbE LX (FC 1364) and SX (FC 1365).
- ▶ OSA-Express 1000BASE-T Ethernet (FC 1366).
- ▶ OSA-Express Fast Ethernet (FC 2366).

Cryptographic functions on the z9 EC are provided by CPACF and the Crypto Express2 feature. The Crypto Express2 (FC 0863) feature combines both the coprocessor and the accelerator functions that were previously offered separately. Detailed information about the Crypto Express2 feature is found in Chapter 5, "Cryptography" on page 149.

### 3.3.2 Physical Channel IDs (PCHIDs)

A Physical Channel ID is the number assigned to a port of an I/O card or PCI-X cryptographic adapter on a Crypto Express2 feature. Each enabled port has its own PCHID number, which is based on its I/O slot location in the I/O cage (except for ESCON sparing).

In the case of an ICB-4 link, the PCHID number is based on its CEC cage location. Figure 3-4 shows an example of a PCHID Report.

```

Machine: 2094-S18   NEW1
-----
Book/Fanout/Jack   Cage    Slot   F/C    PCHID/Ports
1/D8/J00           A01B   D101   0218   100/J00  101/J01
0/D8/J00           A01B   D102   0218   110/J00  111/J01
1/D8/J00           A01B   03     3324   120/J00  121/J01  122/J02  123/J03
0/D8/J00           A01B   04     3364   130/J00  131/J01
1/D8/J00           A01B   06     0863   140/P00  141/P00
0/D8/J00           A01B   07     0863   150/P00  151/P01

Legend:
A19B   Top of A frame
A01B   Bottom of A frame
D1xx   Half high card in top of slot xx
0218   ISC D <10KM
3324   FICON Express4 LX (4 port)
3364   OSA Express GbE LX (2 port)
0863   Crypto Express2

```

Figure 3-4 PCHID Report example



I/O slot 01 has an ISC-3 Daughter (ISC-D) half-high card (FC 0218) in the top, connected to STI 0 (Jack J00) from MBA fanout D8 of book 1. Its two enabled ports have PCHID numbers 100 and 101.

I/O slot 02 has an ISC-3 Daughter (ISC-D) half-high card (FC 0218) in the top, connected to STI 0 (Jack J00) from MBA fanout D8 of book 0. Its two enabled ports have PCHID numbers 110 and 111.

I/O slot 03 has a FICON Express4LX card (FC 3324), connected to STI 0 (Jack J00) from MBA fanout D8 of book 1. Its four enabled ports have PCHID numbers 120 through 123.

I/O slot 04 has an OSA Express2 GbE LX card (FC 3364), connected to STI 0 (Jack J00) from MBA fanout D8 of book 0, and its two ports have PCHID numbers 130 and 131.

I/O slot 06 has a Crypto Expresss2 feature (FC 0863), connected to STI 0 (Jack J00) from MBA fanout D8 of book 1. PCHID numbers 140 and 141 are assigned. Since no ports are present, no jack numbers are assigned, but its two PCI-X cryptographic adapters are identified by P0 and P1.

I/O slot 07 has a Crypto Expresss2 feature (FC 0863), connected to STI 0 (Jack J00) from MBA fanout D8 of book 0. PCHID numbers 140 and 141 are assigned. Since no ports are present, no jack numbers are assigned, but its two PCI-X cryptographic adapters are identified by P0 and P1.

The pre-assigned PCHID number of each I/O port relates directly to its physical location (jack location in a specific slot) except for ESCON sparing; refer to Figure 3-6 on page 107 for an ESCON sparing example.

Table 3-4 shows the PCHID number range for each I/O slot of each I/O cage.

Table 3-4 PCHID numbers and locations

I/O cage slot	PCHID numbers		
	1st I/O cage	2nd I/O cage	3rd I/O cage
01 (front)	100 - 10F	300 - 30F	500 - 50F
02 (front)	110 - 11F	310 - 31F	510 - 51F
03 (front)	120 - 12F	320 - 32F	520 - 52F
04 (front)	130 - 13F	330 - 33F	530 - 53F
06 (front)	140 - 14F	340 - 34F	540 - 54F
07 (front)	150 - 15F	350 - 35F	550 - 55F
08 (front)	160 - 16F	360 - 36F	560 - 56F
09 (front)	170 - 17F	370 - 37F	570 - 57F
10 (front)	180 - 18F	380 - 38F	580 - 58F
11 (front)	190 - 19F	390 - 39F	590 - 59F
12 (front)	1A0 - 1AF	3A0 - 3AF	5A0 - 5AF
13 (front)	1B0 - 1BF	3B0 - 3BF	5B0 - 5BF
15 (front)	1C0 - 1CF	3C0 - 3CF	5C0 - 5CF
16 (front)	1D0 - 1DF	3D0 - 3DF	5D0 - 5DF

I/O cage slot	PCHID numbers		
	1st I/O cage	2nd I/O cage	3rd I/O cage
17 (front)	1E0 - 1EF	3E0 - 3EF	5E0 - 5EF
18 (front)	1F0 - 1FF	3F0 - 3FF	5F0 - 5FF
19 (rear)	200 - 20F	400 - 40F	600 - 60F
20 (rear)	210 - 21F	410 - 41F	610 - 61F
21 (rear)	220 - 22F	420 - 42F	620 - 62F
22 (rear)	230 - 23F	430 - 43F	630 - 63F
24 (rear)	240 - 24F	440 - 44F	640 - 64F
25 (rear)	250 - 25F	450 - 45F	650 - 65F
26 (rear)	260 - 26F	460 - 46F	660 - 66F
27 (rear)	270 - 27F	470 - 47F	670 - 67F
29 (rear)	280 - 28F	480 - 48F	680 - 68F
30 (rear)	290 - 29F	490 - 49F	690 - 69F
31 (rear)	2A0 - 2AF	4A0 - 4AF	6A0 - 6AF
32 (rear)	2B0 - 2BF	4B0 - 4BF	6B0 - 6BF

Note that:

- ▶ I/O cage slot numbers 05, 14, 23, and 28 are reserved for STI-MP cards.
- ▶ The PCHID number range from 000 to 0FF is reserved for ICB-4 links. ICB-4 links are directly connected to a STI port in a MBA fanout card in a book.

Table 3-5 shows the ICB-4 PCHID numbers range for each book.

Table 3-5 PCHID numbers for ICB-4 links

CEC cage book	PCHID numbers
0	010 - 01F
1	020 - 02F
2	030 - 03F
3	000 - 00F

**Important:** If the STI Rebalance feature (feature code 2400) is selected on a z9 EC server upgrade, the current ICB-4 PCHID numbers will change. This requires the corresponding update of the ICB-4 link definition in the z9 EC server I/O configuration.

The PCHID Report has all the installed PCHID numbers. When the configuration is defined, PCHIDs are assigned to Channel Path IDs (CHPIDs) using the CHPID Mapping Tool, or HCD/HCM, or IOCP. The CHPID assignment associates the CHPID number to a physical channel port location (PCHID).

HiperSockets (IQD) and IC links (ICP) do not have PCHIDs, as they are virtual and not physical links, but they do require CHPID numbers.

The Crypto Express2 feature does not require CHPID numbers, but its PCI-X cryptographic adapters are assigned PCHIDs.

### ***z9 EC CHPID Mapping Tool (CMT)***

The z9 EC CHPID Mapping Tool is highly recommended for PCHID-to-CHPID assignments. For complex configurations and configurations using multiple Channel Subsystems, using the CMT is recommended.

CMT has a manual mapping and an availability mapping function; the latter does the PCHID-to-CHPID assignments for the best availability. For more about the z9 EC CMT, see Appendix B, “CHPID mapping tool” on page 263.

## **3.4 Connectivity**

Input/output (I/O) channels are components of the z9 EC server Channel Subsystem (CSS). They provide a pipeline through which data is exchanged between processors, or between a processor and external devices or networks. The most common type of device attached to a channel is a control unit (CU). The CU controls I/O devices such as disk and tape drives.

Server-to-server communications are most commonly implemented using Inter-System Channels (ISC-3), Integrated Cluster Bus (ICB4, or ICB-3) links, or channel-to-channel (CTC) connections.

Network connectivity (LAN and WAN) is provided with OSA Express2 features. There are specific OSA Express2 cards to support Gigabit Ethernet (GbE), 1000BASE-T Ethernet, 100BASE-T Ethernet, and 10BASE-T Ethernet. For additional, detailed information about z9 EC connectivity, see the IBM Redbooks publication *IBM System z Connectivity Handbook*, SG24-5444.

### **3.4.1 I/O and cryptographic features support and configuration rules**

Table 3-6 summarizes all available I/O and cryptographic features on z9 EC servers, the maximum number of ports for each type, the number of I/O slots required to achieve this number, and the port increments for new build systems. Cards that are carried forward from z900 or z990 upgrades are not shown in the table.

*Table 3-6 I/O and cryptographic features support*

I/O feature	Feature codes	Number of		Max. number of		PCHID	CHPID definition	Notes
		Ports per card	Ports increments	Ports	I/O slots			
ESCON	2323 2324 (ports)	16 (one spare)	4 (LIC-CC)	1024	69	Yes	CNC, CVC, CTC, CBY	1, 2
FICON Express LX/SX	2319/2320	2	2	120	60	Yes	FC, FCP FCV	9, 10
FICON Express2 LX/SX	3319/3320	4	4	336	84	Yes	FC, FCP	3
FICON Express4 LX/SX	3321/3324/ 3322	4	4	336	84	Yes	FC, FCP	3
OSA-Express2 Gb Ethernet LX/SX	3364/3365	2	2	48	24	Yes	OSD, OSN (z9 only)	

I/O feature	Feature codes	Number of		Max. number of		PCHID	CHPID definition	Notes
		Ports per card	Ports increments	Ports	I/O slots			
OSA-Express2 10Gb LR	3368	1	1	24	24	Yes	OSD	
OSA-Express2 1000BASE-T Ethernet	3366	2	2	48	24	Yes	OSE, OSD, OSC, OSN	
ICB-3 (1 Gbps)	0993	2	1	16	8	Yes	CBP	4
ICB-4 (2.0 Gbps)	3393	-	1	16	0	Yes	CBP	4
ISC-3 at 10km (2 Gbps)	0217 (ISC-M) 0218 (ISC-D) 0219 (ports)	4/ISC-M 2/ISC-D	1 (LIC-CC)	48	12	Yes	CFP	4, 5
ISC-3 20km support (1 Gbps)	RPQ 8P2197 (ISC-D)	4/ISC-M 2/ISC-D	2	48	12	Yes	CFP	4, 5
HiperSockets	-	-	1	16	0	No	IQD	6
IC	-	-	2	32	0	No	ICP	4, 6
ETR	6155	1	-	2	-	No	-	7
Crypto Express2	0863	2	2	16	8	Yes	-	8

- The ESCON 16-port card feature code is 2323, while individual ESCON ports are ordered in increments of four using feature code 2324. The ESCON card has one spare port and up to 15 usable ports.
- The maximum number of ESCON ports on a z9 EC Model S08 is 960.
- The maximum number of FICON Express2 (carry forward only) or FICON Express4 ports on a z9 EC Model S08 model is 240.
- The sum of IC, ICB-3, ICB-4, active ISC-3, and RPQ 8P2197 links supported on a z9 EC is limited to 64.
- There are three feature codes for the ISC-3 card:
  - Feature code 0217 is for the ISC mother card (ISC-M).
  - Feature code 0218 is for the ISC daughter card (ISC-D).
  - Feature code 0219 is an ISC-3 link on the ISC daughter card feature code 0218.

One ISC mother card supports up to two ISC daughter cards, and each ISC daughter card contains two ports. Port activation must be ordered using feature code 0219. RPQ 8P2197 is available to extend the distance of ISC-3 links to 20 km at 1 Gbps. When RPQ 8P2197 is ordered, both ports (links) in the card are activated.
- There are two types of internal links that can be defined and that require CHPID numbers, but do not have PCHID numbers:
  - Internal Coupling (IC) links. Each IC link pair requires two CHPID numbers.
  - HiperSockets, also called Internal Queued Direct I/O (iQDIO). Up to 16 virtual LANs can be defined, each one requiring a CHPID number.

7. Two ETR cards are automatically included in a server configuration if any Coupling Link I/O feature (ISC-3, ICB-3, or ICB-4) is selected.
8. The initial order is two features, then the additional increment is one feature. The Crypto Express2 feature does not require CHPIDs, but use PCHIDs; one PCHID per PCI-X cryptographic adapter.
9. Only when carried forward from a z990 or z900 upgrade. the I/O slot limit is from z990, as this is the maximum a z990 can have.
10. FCV is supported by FICON Express LX only (FC 2319).

At least one channel I/O feature (FICON Express or ESCON) or one Coupling link I/O feature (ISC-3, ICB-3, or ICB-4) must be present in a configuration.

The maximum number of configurable CHPIDs is 256 per Channel Subsystem (CSS) and per operating system image.

### Spanned and shared channels

The Multiple Image Facility (MIF) allows channels to be shared among multiple logical partitions in a server.

- ▶ Shared channels can be shared by logical partitions within a Channel Subsystem.
- ▶ Spanned channels can be shared by logical partitions within and across CSSs.

The following ESCON channels *cannot* be shared or spanned:

- ▶ ESCON channels defined with CHPID type CVC or CBY

The following channels can be shared but *not* spanned:

- ▶ ESCON channels defined with CHPID type CNC or CTC
- ▶ FICON channels defined with CHPID type FCV

All other channels can be shared and spanned:

- ▶ FICON Express when defined as CHPID type FC or FCP
- ▶ FICON Express2 and FICON Express4
- ▶ OSA-Express and OSA-Express2
- ▶ Coupling links channels in peer mode: ICB-4, ICB-3, and ISC-3
- ▶ Internal channels: IC and HiperSockets

The Crypto Express2 feature does not have a CHPID type, but logical partitions in all CSSs have access to the feature; each PCI-X Cryptographic Adapter can be defined to up to 16 active logical partitions.

### I/O features cables and connectors

**Attention:** All fiber optic cables, cable planning, labeling, and installation are customer responsibilities for new z9 EC installations and upgrades. Fiber optic conversion kits and Mode Conditioning Patch (MCP) cables are not orderable as features on z9 EC servers; only ICB (copper) cables are orderable. All other cables have to be sourced separately.

IBM Fiber Cabling Services offer a total cable solution service to help with cable ordering needs, and is highly recommended. These services take into consideration the requirements for all of the protocols/media types supported (for example, ESCON, FICON, Coupling Links, and OSA), whether the focus is the data center, the Storage Area Network (SAN), Local Area Network (LAN), or the end-to-end enterprise.

The Enterprise Fiber Cabling Services employ the use of a proven modular cabling system, the Fiber Transport System (FTS), which includes trunk cables, zone cabinets, and panels for servers, directors, and storage devices.

FTS supports Fiber Quick Connect (FQC), a fiber harness integrated in the frame of a z9 EC for “quick” connect, which is offered as a feature on z9 EC for connection to ESCON channels.

Whether a packaged service or a custom service is chosen, high quality components are used to facilitate moves, adds, and changes in the enterprise to prevent extending the maintenance window.

Table 3-7 lists the required connectors and cable types for each I/O feature on z9 EC servers.

Table 3-7 I/O features connectors and cables types

Feature code	Feature name	Connector type	Cable type
0219	ISC-3 link	LC Duplex	9 micron SM <sup>1</sup>
6155	ETR	MT-RJ	62.5 micron MM <sup>2</sup>
2324	ESCON channel	MT-RJ	62.5 micron MM
2319 <sup>4</sup>	FICON Express LX	LC Duplex	9 micron SM
3319 <sup>4</sup>	FICON Express2 LX	LC Duplex	9 micron SM
3324	FICON Express4 LX 4km	LC Duplex	9 micron SM
3321	FICON Express4 LX 10km	LC Duplex	9 micron SM
2320 <sup>4</sup>	FICON Express SX	LC Duplex	50, 62.5 micron MM
3320 <sup>4</sup>	FICON Express2 SX	LC Duplex	50, 62.5 micron MM
3322	FICON Express4 SX	LC Duplex	50, 62.5 micron MM
1364 <sup>4</sup>	OSA-E GbE LX <sup>3</sup>	LC Duplex	9 micron SM
2364 <sup>4</sup>	OSA-E GbE LX	SC Duplex	9 micron SM
3364	OSA-E2 GbE LX	LC Duplex	9 micron SM
1365 <sup>4</sup>	OSA-E GbE SX	LC Duplex	50, 62.5 micron MM
2365 <sup>4</sup>	OSA-E GbE SX	SC Duplex	50, 62.5 micron MM
3365	OSA-E2 GbE SX	LC Duplex	50, 62.5 micron MM
3368	OSA-E2 GbE LR	SC Duplex	9 micron SM
1366 <sup>4</sup>	OSA-E 1000BASE-T	RJ-45	Category 5 UTP <sup>5</sup>
3366	OSA-E2 1000BASE-T	RJ-45	Category 5 UTP <sup>5</sup>
2366 <sup>4</sup>	OSA-E Fast Ethernet	RJ-45	Category 5 UTP <sup>5</sup>

1. SM is single mode fiber.
2. MM is multimode fiber.
3. OSA-E refers to OSA-Express.
4. Brought forward to z9 EC on an upgrade only.
5. UTP is Unshielded Twisted Pair.

## 3.4.2 ESCON channels

ESCON channels support the ESCON architecture and directly attach to ESCON-supported I/O devices.

### 16-port ESCON feature

The 16-port ESCON feature (FC 2323) occupies one I/O slot in an I/O cage. Each port on the feature uses a 1300 nanometer (nm) light-emitting diode (LED) transceiver, designed to be connected to 62.5 micron multimode fiber optic cables only.

The feature has 16 ports with one PCHID associated with each port, up to a maximum of 15 active ESCON channels per feature. There is a minimum of one spare port per feature, to allow for channel sparing in the event of a failure of one of the other ports.

The 16-port ESCON feature port utilizes a small form factor optical transceiver that supports a fiber optic connector called MT-RJ. The MT-RJ is an industry standard connector that has a much smaller profile compared with the original ESCON Duplex connector. The MT-RJ connector, combined with technology consolidation, allows for the much higher density packaging implemented with the 16-port ESCON feature.

**Note:** The z9 EC 16-port ESCON feature does *not* support a multimode fiber optic cable terminated with an ESCON Duplex connector. However, 62.5 micron multimode ESCON Duplex jumper cables *can* be reused to connect to the 16-port ESCON feature. This is done by installing an MT-RJ/ESCON Conversion kit between the 16-port ESCON feature MT-RJ port and the ESCON Duplex jumper cable. This protects the investment in the existing ESCON Duplex cabling infrastructure.

Fiber optic conversion kits and Mode Conditioning Patch (MCP) cables are not orderable as features on z9 EC. Fiber optic cables, cable planning, labeling, and installation are all customer responsibilities for new z9 EC installations and upgrades.

IBM Fiber Cabling Services offer a total cable solution service to help with cable ordering needs, and is highly recommended.

### ESCON channel port enablement feature

The 15 active ports on each 16-port ESCON feature are activated in groups of four ports through Licensed Internal Code - Control Code (LIC-CC), by using the ESCON channel port feature (FC 2324).

The first group of four ESCON ports requires two 16-port ESCON features. After the first pair of ESCON cards is fully allocated (by seven ESCON ports groups, using 28 ports), single cards are used for additional ESCON ports groups.

Ports are activated equally across all installed 16-port ESCON features for high availability. In most cases, the number of physically installed channels is greater than the number of active channels that are LIC-CC enabled. This is not only because the last ESCON port (J15) of every 16-port ESCON channel card is a spare, but also because several physically installed channels are typically inactive (LIC-CC protected). These inactive channel ports are available to satisfy future channel adds.

If there is a requirement to increase the number of ESCON channel ports (minimum increment is four), and there are sufficient unused ports already available to fulfill this requirement, an LIC-CC diskette is sent to concurrently enable the number of additional ESCON ports ordered. This is illustrated in Figure 3-5. In this case, no additional hardware is installed.

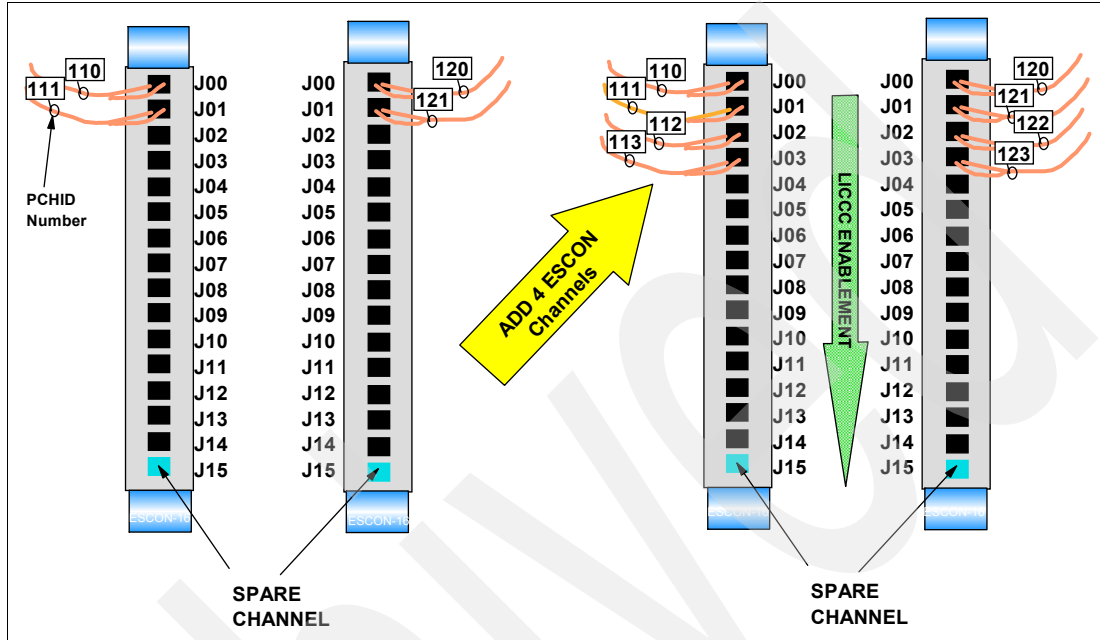


Figure 3-5 16-port ESCON - LIC-CC

An ESCON channel add will never activate the spare channel port. However, if the spare port on a card was previously used, then the add may activate all the remaining ports on that card.

If there are not enough inactive ports on existing 16-port ESCON cards installed to satisfy the additional channel order, an additional 16-port ESCON channel cards is shipped with an LIC-CC diskette.

If there has been multiple sparing on a 16-port ESCON card and, by replacing that card the additional channel add can be satisfied, the card will be replaced.

A maximum of 1024 ESCON ports can be activated on a z9 EC server. This maximum requires 69 16-port ESCON channel cards to be installed. The z9 EC Model S08 can have up to 960 ESCON ports, on 64 channel cards. This number is limited by the number of available STIs on the S08 model.



### 16-port ESCON channel sparing

The last ESCON port on a 16-port ESCON channel card (normally J15) is assigned as a spare port. Should an LIC-CC-enabled ESCON port on the card fail, the spare port is used to replace it, as shown in Figure 3-6.

If the initial first spare port (J15) is already in use and a second LIC-CC-enabled port fails, then the highest LIC-CC-protected port (for example, J14) is used to spare the failing port.

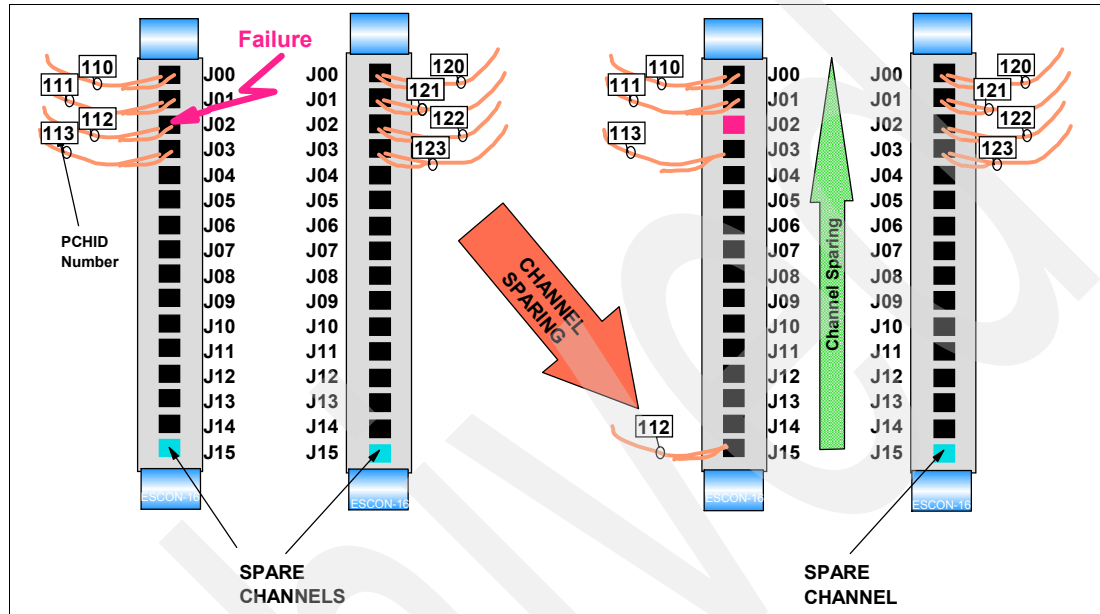


Figure 3-6 16-port ESCON channel sparing

Channel port sparing is only done between ports on the same 16-port ESCON card. A failing ESCON channel port cannot be spared with a port on another 16-port ESCON card.

Channel sparing is a service repair action performed by a service representative using the z9 EC server maintenance package Repair and Verify procedure. If sparing can take place, the IBM SR moves the external fiber optic cable from the failing port to the spare port. When sparing occurs, the PCHID moves to the spare port (PCHID 112 in Figure 3-6). If sparing cannot be performed, the 16-port ESCON card is replaced.

### Fiber Quick Connect (FQC) for ESCON Quick Connect

The Fiber Quick Connect (FQC) features are optional features for factory installation of the IBM Fiber Transport System (FTS) fiber harnesses for connection to ESCON channels in the I/O cage. Each direct-attach fiber harness connects to six ESCON channels at one end and one coupler in a Multi-Terminated Push-On Connector (MTP) coupler bracket at the opposite end. When ordered, the features support all of the installed ESCON features in all of the installed I/O cages. FQC cannot be ordered on a partial or one cage basis.

FQC supports all of the ESCON channels in the I/O cage. FQC cannot be ordered for selected channels.

### 3.4.3 FICON channels

The FICON Express4, FICON Express2, and FICON Express LX and SX features conform to the Fiber Connection (FICON) architecture and directly attach to FICON-supported I/O devices.

FICON channels can be shared among logical partitions and be defined as spanned. All ports on a FICON feature must be of the same type, either LX or SX.

#### **FICON Express4**

Three types of FICON channel transceivers are supported on new build z9 EC servers, two long wavelength (LX) laser versions, and one short wavelength (SX) LED version:

- ▶ FICON Express4 10km LX feature FC 3321, with four ports per feature, supporting LC Duplex connectors.
- ▶ FICON Express4 4km LX feature FC 3324, with four ports per feature, supporting LC Duplex connectors.
- ▶ FICON Express4 SX feature FC 3322, with four ports per feature, supporting LC Duplex connectors.

All channels on a feature are of the same type, either 10 km LX, 4 km LX, or SX. The features are connected to a FICON-capable control unit, either point-to-point or switched point-to-point, through a Fibre Channel switch.

Up to 336 FICON Express4 channels (up to 84 features) can be installed in the z9 EC. The Model S08 can have up to 256 FICON channels (64 features). The number is limited by the number of available STIs on the S08 model.

All FICON Express4 features use Small Form Factor Pluggable (SFP) optics that allow for concurrent repair or replacement for each SFP. The data flow on the unaffected channels on the same feature can continue. A problem with one FICON port may no longer require replacement of a complete feature.

There are two FICON Express4 LX features: one supports an unrepeated distance of 10 kilometers, and the other an unrepeated distance of 4 kilometers, using 9 micron single mode fiber. The FICON Express4 SX feature supports varying distances depending on the fiber used (50, or 62.5 micron multimode fiber) and the link speed (1 Gbps, 2 Gbps, or 4 Gbps).

FICON Express4 features are listed below with a short description of their respective support levels and attachment capabilities.

#### ***FICON Express4 10km LX feature (FC 3321)***

The FICON Express4 10 km LX feature occupies one I/O slot in the I/O cage. The feature occupies one I/O slot in the I/O cage; it has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated distance of 10 km (6.2 miles). Interoperability of 10 km transceivers with 4 km transceivers is supported, provided the unrepeated distance between the two transceivers does not exceed 4 km (2.5 miles).

Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps, 2 Gbps, 4 Gbps, and FICON LX Fibre Channels.
- ▶ Control units that support 1 Gbps, 2 Gbps, and 4 Gbps FICON LX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the z9 EC FICON Express4 10 km LX feature uses a 1300 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector. Use of MCP cables limits the link speed to 1 Gbps and the unrepeated distance to 550 meters (1804 feet).

#### ***FICON Express4 4km LX feature (FC 3324)***

The FICON Express4 4km LX feature occupies one I/O slot in the I/O cage. The feature occupies one I/O slot in the I/O cage; it has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated distance of 4 km (2.5 miles). Interoperability of 10 km transceivers with 4 km transceivers is supported, provided the unrepeated distance between the two transceivers does not exceed 4 km.

Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps, 2 Gbps, 4 Gbps, and FICON LX Fibre Channels.
- ▶ Control units that support 1 Gbps, 2 Gbps, and 4 Gbps FICON LX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the z9 EC FICON Express4 4km LX feature uses a 1300 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector. Use of MCP cables limits the link speed to 1 Gbps and the unrepeated distance to 550 meters (1804 feet).

#### ***FICON Express4 SX feature (FC 3322)***

The FICON Express4 SX feature occupies one I/O slot in the I/O cage. The feature occupies one I/O slot in the I/O cage. It has two Peripheral Component Interconnect (PCI) cards. The PCI cards have a higher performing infrastructure, which can improve performance compared to the FICON Express2 LX feature. Each PCI card has two ports supporting an LC duplex connector, with one CHPID associated with each port, and supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps. Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps, 2 Gbps, 4 Gbps, and FICON SX Fibre Channels.
- ▶ Control units that support 1 Gbps, 2 Gbps, and 4 Gbps FICON SX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the FICON Express SX feature uses an 850 nanometer (nm) fiber bandwidth SX transceiver. The port supports connection to a 62.5 micron or 50 micron multimode fiber optic cable terminated with an LC Duplex connector. Unrepeated distances vary with the use of 50 micron or 62.5 micron fiber optic cable and the data rate.

**Note:** IBM has qualified the 50 micron multimode 2000 MHz-km ISO/IEC OM3, TIA 850 nanometer laser-optimized 50/125 micrometer fiber optic cable for use when attaching System z to servers, switches/directors, disks, tapes, and printers. Support of the OM3 cable is designed to help facilitate use of the industry-standard fiber optic cabling when the unrepeated distances offered by 50 micron 500 MHz-km multimode fiber optic cabling are insufficient for data center requirements.

## FICON Express2

Two types of FICON Express2 channel transceivers are supported on z9 EC servers when carried forward on an upgrade: A long wavelength (LX) laser version, and a short wavelength (SX) LED version.

- ▶ z9 EC FICON Express2 LX feature FC 3319, with four ports per feature, supporting LC Duplex connectors.
- ▶ z9 EC FICON Express2 SX feature FC 3320, with four ports per feature, supporting LC Duplex connectors.

The features are connected to a FICON-capable control unit, either point-to-point or switched point-to-point, through a Fibre Channel switch.

Up to 336 FICON Express2 channels (84 features) can be installed in the z9 EC. The Model S08 can have up to 256 FICON channels (64 features). The number is limited by the number of available STIs on the S08 model.

All supported FICON Express2 features on the z9 EC are listed below with a short description of their respective support levels and attachment capabilities.

### ***FICON Express2 LX feature (FC 3319)***

The FICON Express2 LX feature occupies one I/O slot in the I/O cage. The feature occupies one I/O slot in the I/O cage; it has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps and 2 Gbps FICON LX Fibre Channels.
- ▶ Control units that support 1 Gbps and 2 Gbps FICON LX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the z9 EC FICON Express2 LX feature uses a 1300 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector.

### ***FICON Express2 SX feature (FC 3320)***

The FICON Express2 SX feature occupies one I/O slot in the I/O cage. The feature occupies one I/O slot in the I/O cage; it has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps and 2 Gbps FICON SX Fibre Channel.
- ▶ Control units that support 1 Gbps and 2 Gbps FICON SX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the FICON Express SX feature uses an 850 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 62.5 micron or 50 micron multimode fiber optic cable terminated with an LC Duplex connector.

## FICON Express

FICON Express features (FC 2319 and FC 2320) are carried forward to the z9 EC when upgrading from a z900 or z990 server.

### ***FICON Express LX feature (FC 2319)***

The FICON Express LX feature occupies one I/O slot in the z9 EC I/O cage. The feature occupies one I/O slot in the I/O cage; it has two ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- ▶ FICON LX Bridge one port feature of an IBM 9032 ESCON Director at 1 Gbps *only*.
- ▶ Fibre Channel Switches that support 1 Gbps and 2 Gbps FICON LX Fibre Channels.
- ▶ Control units that support 1 Gbps and 2 Gbps FICON LX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the z9 EC FICON Express LX feature uses a 1300 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector.

### ***FICON Express SX feature (2320)***

The FICON Express SX feature occupies one I/O slot in the z9 EC I/O cage. The feature occupies one I/O slot in the I/O cage; it has two ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- ▶ Fibre Channel Switches that support 1 Gbps and 2 Gbps FICON SX Fibre Channels.
- ▶ Control units that support 1 Gbps and 2 Gbps FICON SX Fibre Channels.
- ▶ FICON channels in Fibre Channel Protocol (FCP) mode.

Each port of the FICON Express SX feature uses an 850 nanometer (nm) fiber bandwidth transceiver. The port supports connection to a 62.5 micron or 50 micron multimode fiber optic cable terminated with an LC Duplex connector.

**Note:** A multimode (62.5 or 50 micron) fiber optic cable may be used with the z9 EC FICON Express LX, FICON Express2 LX, and FICON Express4 LX features for 1 Gbps *only*. The use of this multimode cable type requires a Mode Conditioning Patch (MCP) cable to be used at each end of the fiber optic link, or at each optical port in the link. Use of the single mode to multimode MCP cables reduces the supported distance of the 1 Gbps link to an end-to-end maximum of 550 meters.

Fiber optic conversion kits and Mode Conditioning Patch (MCP) cables are not orderable as features. Fiber optic cables, cable planning, labeling, and installation are all customer responsibilities for new installations and upgrades.

IBM Fiber Cabling Services offer total a cable solution service to help with cable ordering needs, and is highly recommended.

### **FICON performance**

FICON Express4 features exploit the up to 2 GB per second bandwidth of the STIs and show improved performance over its predecessors. It should however be noted that as the link data rate increase, the unrepeated distance decreases. Care must be taken when migrating from a 2 Gbps to an 4 Gbps infrastructure.

The Modified Indirect Data Address Word facility improves FICON performance. The MIDAW facility is unique for System z9 and is described in detail in 4.2, “The MIDAW facility” on page 144.

### ***FICON Express4 performance improvement for native FICON***

A FICON Express4 channel, when operating at 4 Gbps, is designed to achieve a maximum throughput of up to 330 MBps when processing all read or all write (half-duplex data transfers) large sequential data transfer I/O operations. This represents approximately a 65% increase compared to a FICON Express2 channel operating at 2 Gbps.

A FICON Express4 channel, when operating at 4 Gbps, is designed to achieve up to 350 MBps when processing a mix of read and write large sequential data transfer I/O operations. This represents approximately a 25% increase compared to a FICON Express2 channel operating at 2 Gbps.

These large sequential data transfer measurements for native FICON (CHPID type FC) are examples of the maximum throughput that can be achieved in a laboratory environment using one FICON Express4 channel on a z9 EC with z/OS V1.7 with no other processing occurring and do not represent actual field measurements.

### ***FICON Express4 performance improvement for FCP***

A FICON Express4 FCP channel, when operating at 4 Gbps, is designed to achieve a maximum throughput of up to 400 MBps when processing all read and up to 392 MBps when processing all write (half-duplex data transfers) large sequential data transfer I/O operations. This represents greater than a 100% increase compared to a FICON Express2 channel operating at 2 Gbps.

A FICON Express4 FCP channel, when operating at 4 Gbps, is designed to achieve up to 525 MBps when processing a mix of read and write large sequential data transfer I/O operations. This represents approximately a 50% increase compared to a FICON Express2 channel operating at 2 Gbps.

These large sequential data transfer measurements for FCP (CHPID type FCP to communicate with SCSI devices) are examples of the maximum throughput that can be achieved in a laboratory environment using one FICON Express4 FCP channel on a z9 EC with Linux on System z, SUSE Linux SLES 9 SP3, with no other processing occurring, and do not represent actual field measurements.

### ***FCP performance metrics for Linux on System z***

When a FICON channel is configured as CHPID type FCP, I/O information is made available and can be extracted using Linux on System z. These performance metrics measure the portions of response time spent in the I/O subchannel and the FCP fabric. It helps with analysis of FCP channels. Performance metrics information applies the FICON Express4 and FICON Express2 features (CHPID type FCP) on System z9 EC and System z9 BC in the Linux on System z environment.

### **FICON availability**

All FICON Express4 features use Small Form Factor Pluggable (SFP) optics that allow for concurrent repair or replacement for each SFP. The data flow on the unaffected channels on the same feature can continue. A problem with one FICON port may not require replacement of a complete feature.

In a fiber optic infrastructure with long distances, cabling problems can be challenging. To address this issue, FICON supports Request Node Identification Data (RNID). FICON Link incident reports may be captured and analyzed. FICON purge path support takes care of reporting error-related data on a FICON path.

### ***Request Node Identification Data (RNID)***

RNID assists with isolation of cabling detected errors. Operating systems can request RNID data for each device or control unit attached to FICON channels and have it displayed by the operator command (Display M=DEV).

### ***FICON Link incident reporting***

Operating systems like z/OS can capture data regarding link incidents for link problem determination. The data is displayed on the console and recorded in LOGREC.

Link Incident reporting is integral to the FICON (and ESCON) architecture. When a problem on a link occurs, this mechanism identifies the two connected nodes between which the problem occurred, potentially leading to faster problem determination and service. For FCP, Link Incident reporting is not a requirement for the architecture (though it may be offered as an optional switch function). Consequently, important problem determination information may not be available should a problem occur on an FCP link.

### ***FICON purge path extended***

The purge path extended function provides the capability for FICON problem determination. The FICON purge path error-recovery function transfers error-related data and statistics between the channel and entry switch and the control unit and its entry switch to the host operating system.

### ***FICON error recovery***

The z9 EC in combination with z/OS Version 1.7 (and later) recovery processing are designed to allow for the system to detect switch, director, or fabric problems that may cause FICON links to fail. If problems persist, this may mean that recovery repeatedly takes place, substantially reducing FICON throughput. FICON error recovery will keep an affected path offline, until operator action is taken, limiting the performance impact of recovery actions.

### ***FICON channel in Fibre Channel Protocol (FCP) mode***

When FICON channels are configured for FCP mode, the FICON Express, FICON Express2, and FICON Express4 features can access FCP devices either:

- ▶ Through a FICON channel in FCP mode through a single Fibre Channel switch or multiple switches to an FCP device.
- ▶ Through a FICON channel in FCP mode through a single Fibre Channel switch or multiple switches to a Fibre Channel-to-SCSI bridge.
- ▶ As FICON point-to-point connections (not routed through a switch or director) to access devices, or to IPL an operating system.

### ***FCP Adapter interruptions***

Hardware, Linux on System z, and z/VM work together to improve performance by exploiting extensions to the Queued Direct Input/Output (QDIO) architecture. Adapter Interruptions, first added to z/Architecture with HiperSockets, provide an efficient, high-performance technique for I/O interruptions to reduce path lengths and overhead in both the host operating system and the FICON Express adapter when using the FCP CHPID type.

In extending the use of adapter interruptions to FCP channels, the programming overhead to process a traditional I/O interruption is reduced. This benefits FCP support in Linux on System z.

Adapter interruptions apply to a FICON Express, FICON Express2, and FICON Express4 channels when in FCP mode (FCP CHPID type), in support of SCSI devices in a Linux on System z environment.

### ***FCP SCSI IPL feature (FC 9904)***

This feature allows z/VM and Linux on System z to do IPL from a SCSI or FCP disk. It means z/VM and Linux can be started and run completely from SCSI or FCP disks. Both IPLs of logical partition images and z/VM guests are supported.

Further, a stand-alone dump program can be loaded from such a SCSI or FCP disk in order to dump the contents of a logical partition, and the dump data can be written to this same disk.

### ***FCP concurrent patch***

FICON channels, when configured as CHPID type FCP, support concurrent patches allowing the application of a new Licensed Internal Code (LIC-CC) without requiring a configuration of off/on. This is a FCP availability feature, available with FICON Express feature codes 2319 and 2320 and FICON Express2 feature codes 3319 and 3320, and FICON Express4 feature codes 3321, 3322, and 3324.

### ***FCP N\_Port ID Virtualization (NPIV)***

NPIV is the industry-standard solution that allows sharing of a single FCP channel among operating systems in logical partitions, or in z/VM guests. The operating systems use the virtual FCPs as though they were dedicated FCP channels. The Fibre Channel switch to which the FCP channel attaches must have NPIV support.

NPIV allows a single FCP port to register multiple Worldwide Port Names (WWPN) with fabric name server (NS). A unique WWPN is assigned to each operating system image sharing the port. Each registered WWPN is assigned a unique N\_Port ID. With NPIV, a single FCP port can appear as multiple WWPNs in the FCP fabric.

### ***FCP point-to-point attachments***

FCP point-to-point connections (connections not routed through a switch or director) to access devices or to IPL an operating system from a device (for example, using the SCSI IPL feature) are allowed. NPIV is not supported for point-to-point attachments.

### ***z/VM Linux guest performance assists***

z/VM Linux guests benefit from QDIO Enhanced Buffer-State Management (QEBSM) on the System z9. It uses instructions that help to reduce the time for hypervisor interception. Host Page-Management Assist assists z/VM with hardware assign, lock, and unlock of page frames without z/VM intervention.

The performance assists not only apply to FCP, but also to HiperSockets (CHPID type IQD), and all OSA features with CHPID type OSD specified.



### 3.4.4 OSA-Express2 and OSA-Express adapters

What follows is a discussion of the connectivity options by both the OSA-Express2 and OSA-Express environments.

The following OSA-Express2 features can be installed on new build z9 EC systems:

- ▶ OSA-Express2 Gigabit Ethernet (GbE) Long Wavelength (LX), feature code 3364
- ▶ OSA-Express2 Gigabit Ethernet (GbE) Short Wavelength (SX), feature code 3365
- ▶ OSA-Express2 Gigabit Ethernet 10 GbE LR, feature code 3368
- ▶ OSA-Express2 1000BASE-T Ethernet, feature code 3366

The following OSA-Express features are brought forward to the z9 EC on an upgrade only:

- ▶ OSA-Express Gigabit Ethernet (GbE) Long Wavelength (LX), feature code 1364
- ▶ OSA-Express Gigabit Ethernet (GbE) Short Wavelength (SX), feature code 1365
- ▶ OSA-Express Gigabit Ethernet (GbE) Long Wavelength (LX), feature code 2364
- ▶ OSA-Express Gigabit Ethernet (GbE) Short Wavelength (SX), feature code 2365
- ▶ OSA-Express 1000BASE-T Ethernet, feature code 1366
- ▶ OSA-Express Fast Ethernet, feature code 2366.

**Note:** If FDDI or ATM connectivity is desired, a multiprotocol switch or router with the appropriate network interface (for example, 1000BASE-T Ethernet, Gigabit Ethernet) can be used to provide connectivity between the z9 EC server and an FDDI or ATM network.

A z9 EC server supports up to 24 OSA-Express and OSA-Express2 features (48 ports).

Table 3-8 OSA-Express2 and OSA Express features support

I/O feature	Feature codes	Number of		Maximum number		PCHID	CHPID definition
		Ports per card	Ports increments	Ports	I/O slots		
OSA Express2 GbE LX/SX	3364/3365	2	2	48	24	Yes	OSD, OSN
OSA Express2 10 GbE LR	3368	1	1	24	24	Yes	OSD
OSA-Express2 1000BASE-T Ethernet	3366	2	2	48	24	Yes	OSE, OSD, OSC, OSN
OSA-Express GbE Ethernet LX/SX	1364/1365 2364/2365	2	2	48 (24)	24 (12)	Yes	OSD
OSA-Express 1000BASE-T Ethernet	1366	2	2	48	24	Yes	OSE, OSD, OSC,
OSA-E Fast Ethernet	2366	2	2	24	12	Yes	OSE, OSD

All supported OSA-Express2 and OSA-Express features on the z9 EC are listed below with a short description of their respective support levels and attachment capabilities.

#### **OSA-Express2 GbE LX (FC 3364)**

The OSA-Express2 Gigabit (GbE) Long Wavelength (LX) feature occupies one slot in an I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports a connection to a 1 Gbps Ethernet LAN through a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector. This feature utilizes a long wavelength (LX) laser as the optical transceiver.

A multimode (62.5 or 50 micron) fiber cable may be used with the OSA-Express2 GbE LX feature. The use of these multimode cable types requires a mode conditioning patch (MCP) cable to be used at each end of the fiber link. Use of the single-mode to multimode MCP cables reduces the supported optical distance of the link to a maximum end-to-end distance of 550 meters.

The OSA-Express2 GbE LX feature supports Queued Direct Input/Output (QDIO) and OSN mode only, full-duplex operation, jumbo frames, and checksum offload. It is defined with CHPID type OSD or OSN. For OSN mode, see “OSA-Express2 OSN - Open System Adapter for NCP” on page 121.

### ***OSA-Express2 GbE SX (FC 3365)***

The OSA-Express2 Gigabit (GbE) Short Wavelength (SX) feature occupies one slot in an I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports a connection to a 1 Gbps Ethernet LAN through a 62.5 micron or 50 micron multimode fiber optic cable terminated with an LC Duplex connector. The feature utilizes a short wavelength (SX) laser as the optical transceiver.

The OSA-Express2 GbE SX feature supports Queued Direct Input/Output (QDIO) and OSN mode only, full-duplex operation, jumbo frames, and checksum offload. It is defined with CHPID type OSD or OSN. For OSN mode, see “OSA-Express2 OSN - Open System Adapter for NCP” on page 121.

### ***OSA-Express2 10 GbE LR (FC 3368)***

The OSA-Express 10 GbE LR feature occupies one slot in an I/O cage and has one port that connects to a 10 Gbps Ethernet LAN through a 9 micron single mode fiber optic cable terminated with an SC Duplex connector. The feature supports an un-repeated maximum distance of 10 km.

The OSA-Express2 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. The OSA-Express 10 GbE LR feature is defined as CHPID type OSD.

### ***OSA-Express2 1000BASE-T Ethernet (FC 3366)***

The OSA-Express2 1000BASE-T Ethernet occupies one slot in the I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports connection to either a 1000BASE-T (1000 Mbps), 100BASE-TX (100 Mbps), or 10BASE-T (10 Mbps) Ethernet LAN. The LAN must conform either to the IEEE 802.3 (ISO/IEC 8802.3) standard or the DIX V2 specifications.

Each port has an RJ-45 receptacle for cabling to an Ethernet switch that is appropriate for the LAN speed. The RJ-45 receptacle is required to be attached using EIA/TIA category 5 unshielded twisted pair (UTP) cable with a maximum length of 100 m (328 ft).

The OSA-Express2 1000BASE-T Ethernet feature supports auto-negotiation and automatically adjusts to 10 Mbps, 100 Mbps, or 1000 Mbps, depending upon the LAN.

LAN speed or the duplex mode can be set explicitly using OSA/SF or the OSA Advanced Facilities function of the z9 EC server Hardware Management Console (HMC). The explicit settings will override the OSA-Express2 feature port ability to auto-negotiate with its attached Ethernet switch. The OSA-Express2 1000BASE-T Ethernet feature supports CHPID types, OSC, OSD, OSE, and OSN. For OSN mode, see “OSA-Express2 OSN - Open System Adapter for NCP” on page 121.

Any one of the following settings for the OSA-Express2 1000BASE-T Ethernet and OSA-Express2 1000BASE-T Ethernet features can be chosen:

- ▶ Auto-negotiate
- ▶ 10 Mbps half-duplex or full-duplex
- ▶ 100 Mbps half-duplex or full-duplex
- ▶ 1000 Mbps / 1 Gbps full-duplex

LAN speed and duplexing mode default to auto negotiation. The feature port and the attached switch automatically negotiate these settings. If the attached switch does not support auto-negotiation, the port enters the LAN at the default speed of 1000 Mbps and full duplex mode.

The 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, OSE, or OSN.

Non-QDIO operation mode requires CHPID type OSE. When configured at 1 Gbps, the 1000BASE-T Ethernet feature has the same attributes as the fiber Gigabit Ethernet features:

- ▶ Operates in QDIO mode only (CHPID type OSD).
- ▶ Carries TCP/IP packets only.
- ▶ Operates in full-duplex mode only.
- ▶ Supports jumbo frames.
- ▶ Supports checksum offload.

#### ***OSA-Express GbE LX (FC 1364, upgrade only)***

The OSA-Express Gigabit (GbE) Long Wavelength (LX) feature occupies one slot in an I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports a connection to a 1 Gbps Ethernet LAN through a 9 micron single-mode fiber optic cable terminated with an LC Duplex connector. This feature utilizes a long wavelength (LX) laser as the optical transceiver.

A multimode (62.5 or 50 micron) fiber cable may be used with the OSA-Express GbE LX feature. The use of these multimode cable types requires a mode conditioning patch (MCP) cable to be used at each end of the fiber link. Use of the single mode to multimode MCP cables reduces the supported optical distance of the link to a maximum end-to-end distance of 550 meters.

The OSA-Express GbE LX feature supports Queued Direct Input/Output (QDIO) mode only, full-duplex operation, jumbo frames, and checksum offload. It is defined with CHPID type OSD.

#### ***OSA-Express GbE SX (FC 1365, upgrade only)***

The OSA-Express Gigabit (GbE) Short Wavelength (SX) feature occupies one slot in an I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports a connection to a 1 Gbps Ethernet LAN through a 62.5 micron or 50 micron multimode fiber optic cable terminated with an LC Duplex connector. The feature utilizes a short wavelength laser as the optical transceiver.

The OSA-Express GbE SX feature supports Queued Direct Input/Output (QDIO) mode only, full-duplex operation, jumbo frames, and checksum offload. It is defined with CHPID type OSD.

### ***OSA-Express GbE LX (FC 2364, upgrade only)***

The OSA-Express GbE LX feature occupies one slot in an I/O cage and has two independent ports with one PCHID associated with each port.

Each port supports a connection to a 1 Gbps Ethernet LAN through a 9 micron single-mode fiber optic cable terminated with an SC Duplex connector.

A multimode (62.5 or 50 micron) fiber cable may be used with the OSA-Express GbE LX feature. The use of these multimode cable types requires a mode conditioning patch (MCP) cable to be used at each end of the fiber link. Use of the single mode to multimode MCP cables reduces the supported optical distance of the link to a maximum end-to-end distance of 550 meters.

The OSA-Express GbE LX feature only supports QDIO mode and TCP/IP. It is defined with CHPID type OSD. The Enterprise Extender (EE) function of Communications Server for z/OS allows you to run SNA applications and data on IP networks and IP-attached clients.

### ***OSA-Express GbE SX (FC 2365, upgrade only)***

The OSA-Express GbE SX feature occupies one slot in an I/O cage and has two independent ports with one PCHID associated with each port.

Each port supports connection to a 1 Gbps Ethernet LAN through a 62.5 micron or 50 micron multimode fiber optic cable terminated with an SC Duplex connector.

The OSA-Express GbE SX feature only supports QDIO mode and TCP/IP. It is defined with CHPID type OSD. The Enterprise Extender (EE) function of Communications Server for z/OS allows you to run SNA applications and data on IP networks and IP-attached clients.

### ***OSA-Express 1000BASE-T Ethernet (FC 1366, upgrade only)***

The OSA-Express 1000BASE-T Ethernet occupies one I/O slot in the I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports a connection to either a 1000BASE-T (1000 Mbps), 100BASE-TX (100 Mbps), or 10BASE-T (10 Mbps) Ethernet LAN. The LAN must conform either to the IEEE 802.3 (ISO/IEC 8802.3) standard or the DIX V2 specifications.

Each port has an RJ-45 receptacle for cabling to an Ethernet switch that is appropriate for the LAN speed. The RJ-45 receptacle is required to be attached using an EIA/TIA category 5 unshielded twisted pair (UTP) cable with a maximum length of 100 m (328 ft).

The OSA-Express 1000BASE-T Ethernet feature supports auto-negotiation and automatically adjusts to 10 Mbps, 100 Mbps, or 1000 Mbps, depending upon the LAN.

LAN speed or the duplex mode can be set explicitly, using OSA/SF or the OSA Advanced Facilities function of the z9 EC server Hardware Management Console (HMC). The explicit settings will override the OSA-Express feature port ability to auto-negotiate with its attached Ethernet switch. The OSA-Express 1000BASE-T Ethernet feature supports CHPID types, OSC, OSD, and OSE.

### ***OSA-Express Fast Ethernet (FC code 2366, upgrade only)***

The OSA-Express FENET feature occupies one I/O slot in an I/O cage and has two independent ports, with one PCHID associated with each port.

Each port supports connection to either a 100 Mbps or 10 Mbps Ethernet LAN. The LAN must conform either to the IEEE 802.3 (ISO/IEC 8802.3) standard or the Ethernet V2.0 specifications, and the 10BASE-T or 100BASE-TX standard transmission schemes.

Each port has an RJ-45 receptacle for cabling to an Ethernet switch that is appropriate for the LAN speed. The RJ-45 receptacle is required to be attached using an EIA/TIA category 5 unshielded twisted pair (UTP) cable with a maximum length of 100 m (328 ft).

It is possible to choose any one of the following settings for the OSA-Express FENET feature:

- ▶ Auto negotiate
- ▶ 10 Mbps half-duplex or full-duplex
- ▶ 100 Mbps half-duplex or full-duplex

LAN speed and the duplex mode can be set explicitly using OSA/SF or the OSA Advanced Facilities function of the z9 EC server hardware management console (HMC). The explicit settings will override the OSA-Express feature port ability to auto-negotiate with its attached Ethernet switch.

The OSA-Express FENET feature supports auto-negotiation with its attached Ethernet hub, router, or switch. If the LAN speed is set to auto-negotiation, the FENET OSA-Express and the attached hub, router, or switch auto-negotiates the LAN speed setting between them. If the attached Ethernet hub, router, or switch does not support auto-negotiation, the OSA enters the LAN at the default speed of 100 Mbps in half-duplex mode.

If the auto-negotiate is not used, the OSA will attempt to join the LAN at the specified speed/mode; however, the speed/mode settings are only used when the OSA is first in the LAN. If this fails, the OSA will attempt to join the LAN as though auto-negotiate were specified.

The OSA-Express FENET feature can be defined with CHPID type OSD or OSE. The HPDT MPC mode is no longer available on the FENET.

### ***OSA-Express and OSA Express2 Integrated Console Controller (OSA-ICC)***

The 1000BASE-T Ethernet features also provide the Integrated Console Controller (OSA-ICC) function, which supports TN3270E (RFC 2355) and non-SNA DFT 3270 emulation. The OSA-ICC function uses a definition as OSC CHIPD and console controller, and has multiple logical partitions support, both as shared or spanned channels.

With the OSA-ICC function, 3270 emulation for console session connections is integrated in the z9 EC through a port on the OSA-Express 1000BASE-T Ethernet or OSA-Express2 1000BASE-T features. This eliminates the requirement for external console controllers, like 2074 or 3174, helping to reduce cost and complexity. Each port can support up to 120 console session connections.

OSA-ICC can be configured on a port-by-port basis, and is supported at any of the feature settings (10, 100, or 1000 Mbps, half or full-duplex).

### **Checksum Offload for IPv4 packets when in QDIO mode**

A function called Checksum Offload, offered on the OSA-Express GbE, OSA-Express2 GbE, OSA-Express 1000BASE-T, and OSA-Express2 1000BASE-T Ethernet features, is in support of Linux on System z and z/OS environments. Checksum Offload provides the capability of calculating the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Internet Protocol (IP) header checksum. Checksum verifies the correctness of files. By moving the checksum calculations to a Gigabit or 1000BASE-T Ethernet feature, host CPU cycles are reduced and performance is improved.

When checksum is off-loaded, the OSA-Express feature performs the checksum calculations for Internet Protocol Version 4 (IPv4) packets. This function applies to packets that actually go onto the Local Area Network (LAN) or come in from the LAN. When multiple IP stacks share an OSA-Express, and an IP stack sends a packet to a next hop address owned by

another IP stack sharing the OSA-Express, OSA-Express sends the IP packet directly to the other IP stack without placing it out on the LAN. Checksum Offload does not apply to such IP packets.

This function does not apply to IPv6 packets. TCP/IP will continue to perform all checksum processing for IPv6 packets. This function also does not apply to ICMP checksum processing. TCP/IP will continue to perform processing for ICMP checksum.

Checksum Offload is supported by the GbE features (FC 1364, FC 1365, FC 3364, and FC 3365) and the 1000BASE-T Ethernet features (FC 1366 and FC 3366) when operating at 1000 Mbps (1 Gbps). This is applicable to the QDIO mode only (channel type OSD).

z/OS support for Checksum Offload is available in all in-service z/OS releases. For Linux on System z support, refer to the following Web site for further information:

<http://www.ibm.com/developerworks>

### **Adapter interruptions for QDIO**

Hardware, Linux on System z, and z/VM work together to provide performance improvements by exploiting extensions to the Queued Direct Input/Output (QDIO) architecture. Adapter interruptions, first added to z/Architecture with HiperSockets, provide an efficient, high-performance technique for I/O interruptions to reduce path lengths and overhead in both the host operating system and the adapter (OSA-Express when using the OSD CHPID type).

In extending the use of adapter interruptions to OSD (QDIO) channels, the programming overhead to process a traditional I/O interruption is reduced. This benefits OSA-Express TCP/IP support in both Linux on System z and z/VM.

Adapter interruptions apply to all of the OSA-Express features available on z9 EC, whether offered as a new build or on an upgrade when in QDIO mode (CHPID type OSD).

## OSA-Express2 OSN - Open System Adapter for NCP

OSA-Express2 GbE and OSA-Express2 1000BASE-T Ethernet features can provide channel connectivity from an operating system in a z9 EC to the IBM Communication Controller for Linux on System z (CCL) with the Open Systems Adapter for NCP, in support of the Channel Data Link Control (CDLC) protocol (see Figure 3-7). OSA-Express2 OSN eliminates the need for an external communication medium for communications between the operating system and the CCL image.

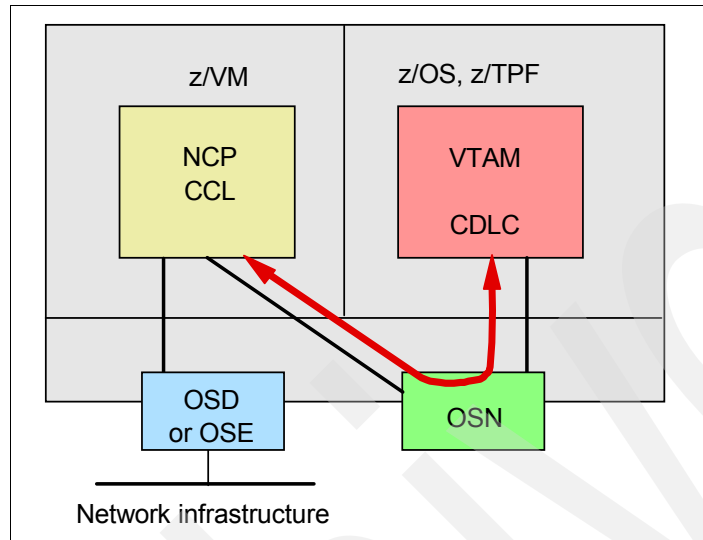


Figure 3-7 OSA for NCP

With OSN, you no longer need to use an external LAN or ESCON channel. The logical partition-to-logical partition data flow is accomplished by the OSA-Express2 feature without ever exiting the card. OSN support allows multiple connections between the same CCL image and the same operating system (such z/OS or TPF). The operating system must reside in the same physical server as the CCL image.

### OSA-Express2 OSN:

- ▶ Is designed to appear to the operating system as a channel connected 374x device type exploiting CDLC protocols.
- ▶ Has the capability to configure, manage, and operate the CCL NCPs as though running in a channel attached 374x Communications Controller.
- ▶ Supports NCP channel-related functions such as loading and dumping.
- ▶ Does not need external switches and cables.
- ▶ Provides support for up to 180 connections per CHPID per OSN.
- ▶ May span CSSs.

For more planning information about CCL, refer to the Redbooks publication *IBM Communication Controller for Linux on System z V1.2.1*, SG24-7223.

## VLAN management

To simplify the network administration and management of VLANs, the GARP VLAN Registration Protocol (GVRP) is used (see Figure 3-8). All OSA-Express2 features support VLAN prioritization, a component of the IEEE 802.1 standard. With this support, it is no longer necessary to manually enter VLAN IDs at the switch. The OSA-Express2 features when in QDIO mode (CHPID type OSD) can have GVRP dynamically register VLAN IDs.

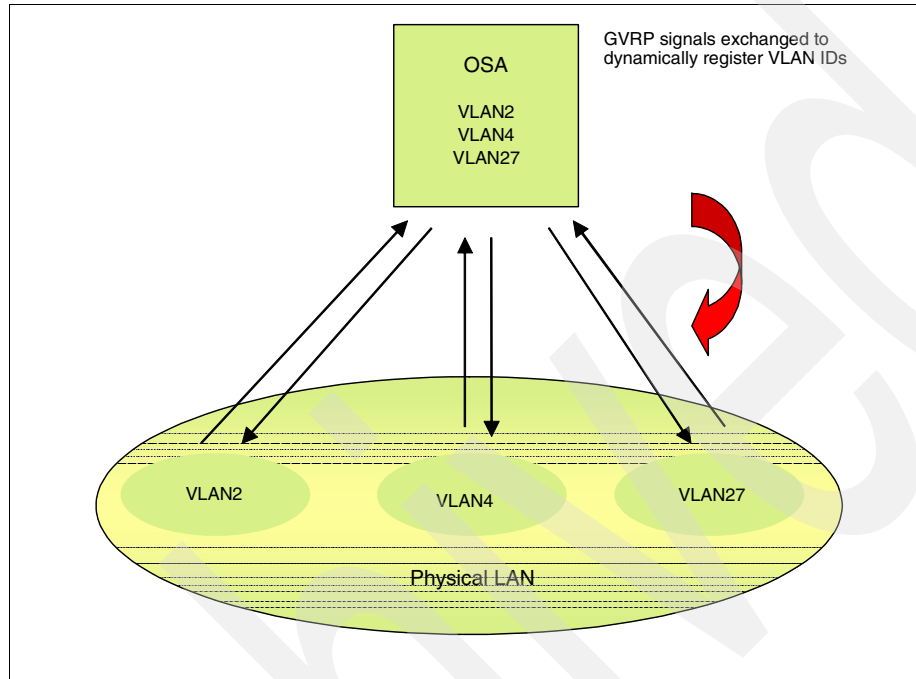


Figure 3-8 VLAN management

### OSA Layer 3 Virtual MAC for z/OS and z/OS.e environments

To help simplify the infrastructure and to facilitate load balancing when an logical partition is sharing the same OSA Media Access Control (MAC) address with another logical partition, each operating system instance can have its own unique "logical" or "virtual" MAC (VMAC) address. All IP addresses associated with a TCP/IP stack are accessible using their own VMAC address, instead of sharing the MAC address of an OSA port. This applies to Layer 3 mode and to an OSA port spanned among Channel Subsystems.

This support is designed to:

- ▶ Improve IP workload balancing.
- ▶ Dedicate a Layer 3 VMAC to a single TCP/IP stack.
- ▶ Remove the dependency on Generic Routing Encapsulation (GRE) tunnels.
- ▶ Improve outbound routing.
- ▶ Simplify configuration setup.
- ▶ Allow WebSphere Application Server content-based routing to work with z/OS in an IPv6 network.
- ▶ Allow z/OS to use a "standard" interface ID for IPv6 addresses.
- ▶ Remove the need for PRIROUTER/SECROUTER function in z/OS.



OSA Layer 3 VMAC is exclusive to System z9, and is applicable to the OSA-Express2 and OSA-Express features when configured as CHPID type OSD (QDIO), and is supported by z/OS V1.8.

### ***OSA-Express2 QDIO Diagnostic Synchronization***

QDIO Diagnostic Synchronization is designed to provide system programmers and network administrators the ability to coordinate and simultaneously capture both software and hardware traces. It allows z/OS to signal an OSA-Express2 feature (using a Diagnostic Assist function) to stop traces and capture the current trace records.

This function is exclusive to System z9 and OSA-Express2 features when configured as CHPID type OSD. z/OS V1.8 implements software support for this feature.

### ***OSA-Express2 Network Traffic Analyzer***

With the large volume and complexity of today's network traffic, the System z9 offers systems programmers and network administrators the ability to more easily solve network problems. With the availability of the OSA-Express Network Traffic Analyzer and QDIO Diagnostic Synchronization on the System z9, customers have the ability to capture trace/trap data and forward it to z/OS tools for easier problem determination and resolution.

This function is exclusive to System z9 and OSA-Express2 features when configured as CHPID type OSD. Support is available in z/OS V1.8.

### ***OSA Dynamic LAN idle***

OSA Dynamic LAN idle parameter change is designed to help reduce latency and improve performance by dynamically adjusting the inbound blocking algorithm. System administrators can authorize the TCP/IP stack to enable a dynamic setting, which was previously a static setting.

For latency sensitive applications, the blocking algorithm is modified to be "latency sensitive." For streaming (throughput sensitive) applications, the blocking algorithm is adjusted to maximize throughput. In all cases, the TCP/IP stack determines the best setting based on the current system and environmental conditions (inbound work load volume, CPU utilization, traffic patterns, and so on) and can dynamically update the settings. An OSA-Express2 feature will "adapt" to the changes, avoiding thrashing and frequent updates to the OSA address table (OAT). OSA will hold packets before "presenting" the packets to the host based upon the TCP/IP settings. A dynamic setting is designed to avoid or minimize host interrupts.

OSA Dynamic LAN idle is exclusive to System z9, is supported by the OSA-Express2 features (CHPID type OSD), and is exploited by z/OS V1.8 with PTFs. PTFs are planned for availability in 3Q 2007.

### ***Link aggregation support for z/VM***

z/VM Virtual Switch-controlled (VSWITCH-controlled) link aggregation (IEEE 802.3ad) allows the dedication of an OSA-Express2 port to the z/VM operating system when the port is participating in an aggregated group configured in Layer 2 mode. Link aggregation (trunking) is designed to allow combining multiple physical OSA-Express2 ports into a single logical link for increased throughput and for nondisruptive failover in the event that a port becomes unavailable. Aggregation allows:

- ▶ Aggregated link to be viewed as one logical trunk and containing all of the Virtual LANs (VLANs) required by the LAN segment
- ▶ Load balance communications across several links in a trunk to prevent a single link from being overrun
- ▶ Link aggregation between a VSWITCH and the physical network switch

- ▶ Point-to-point connections
- ▶ Up to eight OSA-Express2 ports in one aggregated link
- ▶ Ability to dynamically add/remove OSA ports for "on demand" bandwidth
- ▶ Full-duplex mode (send and receive)

Target links for aggregation must be of the same type.

Link aggregation is exclusive to System z9, is applicable to the OSA-Express2 features when configured as CHPID type OSD (QDIO), and is supported by z/VM V5.3.

### **HiperSockets function**

The HiperSockets function, also known as internal Queued Direct Input/Output (iQDIO) or internal QDIO, is an integrated function of the z9 EC server that provides users with attachments to up to sixteen high-speed "virtual" Local Area Networks (LANs) with minimal system and network overhead.

HiperSockets eliminates the need to utilize I/O subsystem operations and the need to traverse an external network connection to communicate between logical partitions in the same z9 EC server. HiperSockets offers significant value in server consolidation connecting many virtual servers, and can be used instead of certain coupling link configurations in a Parallel Sysplex.

HiperSockets can be customized to accommodate varying traffic sizes. Since HiperSockets does not use an external network, it can free up system and network resources, eliminating attachment costs while improving availability and performance.

## **3.4.5 Coupling links**

Coupling links provide connectivity options in the Parallel Sysplex environment. For more information about Parallel Sysplex, see 7.2.4, "Coupling link connectivity" on page 190.

Besides providing connectivity for the Parallel Sysplex environment, CF coupling links are also used to transmit timekeeping messages when Server Time protocol (STP) is enabled. Refer to *Server Time Protocol Planning Guide*, SG24-7280, and *Server Time Protocol Implementation Guide*, SG24-7281 for detailed information about STP.

### **Coupling link features**

The z9 EC supports the following Coupling link features:

- ▶ Inter-System Channel-3, ISC-3 in Peer mode only, feature codes 0217, 0218, and 0219
- ▶ Integrated Cluster Bus-4, ICB-4 in Peer mode, feature code 3393
- ▶ Integrated Cluster Bus-3, ICB-3 in Peer mode, feature code 0993
- ▶ Internal Channel, IC in Peer mode; Licensed Internal Code (LIC-CC) function defined using HCD/IOCP

#### ***ISC-3 link***

The ISC-3 feature is made up of the following feature codes:

- ▶ ISC-3 mother card, feature code 0217
- ▶ ISC-3 daughter card, feature code 0218
- ▶ ISC-3 Port, feature code 0219

The ISC-3 mother card occupies one slot in the I/O cage. The ISC-3 mother card supports up to two ISC-3 daughter cards. Each ISC-3 daughter card has two independent ports with one PCHID associated with each active port. The ISC-3 ports are activated through Licensed Internal Code (LIC-CC).

When the quantity of ISC links (FC 0219) is selected, the quantity of ISC-3 Port features selected determines the appropriate number of ISC-3 mother and daughter cards to be included in the configuration, up to a maximum of 12 ISC-M cards. Additional ISC-M cards can be ordered, up to the number of ISC-D features or twelve, whichever is smaller.

Each active ISC-3 port in peer mode supports a 2-Gbps connection through 9 micron single mode fiber optic cables terminated with an LC-Duplex connector. ISC3 links can be defined as *timing-only links* when STP is enabled. Timing-only links are coupling links that allow two servers to be synchronized using STP messages when a CF does not exist at either end of the link.

**RPQ 8P2197: Extended distance option**

RPQ 8P2197 ISC-3 daughter card has two links per card. Both links are active when installed and do not need to be activated through LIC-CC.

The RPQ allows an ISC-3 link in Peer mode to operate at a 1Gbps rate instead of 2 Gbps. Since the link operates at half the data rate, the maximum distance of the ISC-3 link may be extended to 20 km with this RPQ. For Peer mode, one RPQ daughter card is required at each end of the link between the z9 EC, z990, z900, z890, or z800 servers. STP supports this RPQ when defined either as coupling link or timing-only link.

Table 3-9 lists the various ISC-3 link characteristics.

Table 3-9 ISC-3 link characteristics

Mode of operation	IOCP definition	Bandwidth	Open Fiber Control (OFC)	Intended attachment	Maximum distance
Peer	CFP	2 Gbps	No	z9 EC, z9 BC, z990, z900, z890, or z800	10 km
Peer with RPQ 8P2197	CFP	1 Gbps	No	z9 EC, z9 BC, z990, z900, z890, or z800	20 km

**ICB-4 link (FC 3393)**

The Integrated Cluster Bus-4 (ICB-4) link is a member of the family of Coupling link options available on z9 EC servers. ICB-4 is a “native” connection used to connect z9 EC, z990, and z890 servers to other z9 EC, z990, and z890 servers. An ICB-4 connection consists of one link that attaches directly to an MBA fanout port (STI port) in the CEC cage, does not require connectivity to a card in the I/O cage, and operates at 2 GBps.

One ICB-4 feature is required for each end of the link and each end of the ICB-4 link has a PCHID number.

The ICB-4 cable, feature code 0228, is a unique 10 meter (33 feet) copper cable to be used with ICB-4 links only. ICB-4 cables are automatically ordered to match the quantity of the ICB-4 feature on order. Order one cable per connection, not per feature. The quantity of ICB cables can be reduced, but cannot exceed the quantity of ICB features on order.

When STP is enabled, ICB-4 links can be defined as timing-only links to other z9 EC, z990, and z890 servers.

### **ICB-3 link (FC 0993)**

The Integrated Cluster Bus-3 (ICB-3) link is a member of the family of coupling link options available on z9 EC servers. ICB-3 links are used to connect z900 and z800 servers to z9 EC, z990, and z890 servers. In the I/O cage is an STI-3 card that provides two ports to support the ICB-3 connections. The STI-3 card converts the 2 GBps input into two 1 GBps ICB-3s. One ICB-3 feature is required for each end of the link. Each ICB-3 link at the z9 EC end has a PCHID number.

The ICB-3 cable (feature code 0227) is a unique 10 meter (33 feet) 1.0 GB copper cable to be used with ICB-3 links. Existing 10 meter 1.0 GB ICB-3 cables can be reused. ICB-3 cables will be automatically ordered to match the quantity of ICB-3s on order. Order one cable per connection, not per feature. The quantity of ICB cables can be reduced, but cannot exceed the quantity of ICB features on order.

When STP is enabled, ICB-3 links can be defined as timing-only links to other z9 EC, z990, and z890 servers.

### **ICB links summary**

Table 3-10 shows a summary of the z9 EC Integrated Cluster Bus (ICB) link types.

Table 3-10 z9 EC ICB links summary

ICB link type	Feature code	IOCP definition	Bandwidth	Intended attachment	Maximum cable length
ICB-4	3393	CBP	2 GBps	z9 EC, z9 BC, z990, and z890	10 meters
ICB-3	0993	CBP	1 GBps	z900 and z800	10 meters

For more information about timing-only links, see “Coupling links and Server Time Protocol” on page 192.

### **IC links**

IC links are used when an ICF logical partition is on the same server as other system images participating in the sysplex. An IC link is the fastest Coupling link, using just memory-to-memory data transfers. IC links do not have PCHID numbers, but do require CHPIDs.

IC links require ICP channel path definition at the z/OS and the CF end of a channel connection to operate in peer mode. They are always defined and connected in pairs.

## **3.4.6 External Time Reference feature (FC 6155)**

The External Time Reference (ETR) is an optional feature.

Each ETR feature consists of one ETR card and each card has one port. When a quantity of one is ordered, two features are shipped. The ETR features must be explicitly ordered since with the availability of Server Time Protocol they are no longer automatically added if any coupling link feature (ISC-3, ICB-3, or ICB-4) is ordered.

The ETR cards provide attachment to the Sysplex Timer. Each ETR card should connect to a different 9037 Sysplex Timer in an Expanded Availability configuration. Each feature has a single port supporting an MT-RJ fiber optic connector to provide the capability to attach to a Sysplex Timer Unit. The two ETR cards are supported in one CEC cage card slot in the rear and provide attachment to a 9037 Sysplex Timer. The 9037 Sysplex Timer provides the synchronization for the Time-of-Day (TOD) clocks of multiple servers, and thereby allows events started by different servers to be properly sequenced in time. When multiple servers

update the same database and database reconstruction is necessary, all updates are required to be time stamped in proper sequence.

**Important:** A Sysplex Timer unit is assigned a unique two-digit ID at installation, called a unit ID. This unit ID is referenced as an External Time Reference ID (ETR ID) in the output of the z/OS command Display ETR and Support Element panels.

This function requires the ETR Network ID of the attached Sysplex Timer Network to be manually set in the Support Element at installation. This function checks that the ETR Network ID being received in the timing signals through each of the server's two ETR ports matches the ETR Network ID manually set in the server's Support Element (SE); refer to 7.2.1, "Sysplex configurations and Time Synchronization" on page 185 for more information.

The port cards support concurrent maintenance. The ETR card port has a small form factor optical transceiver that supports an MT-RJ connector only.

**Note:** The ETR card does not support a multimode fiber optic cable terminated with an ESCON Duplex connector as on the Sysplex Timer.

Current 62.5 micron multimode ESCON Duplex jumper cables can be reused to connect to the ETR card. This is done by installing an MT-RJ/ESCON Conversion kit between the ETR card MT-RJ port and the ESCON Duplex jumper cable.

Fiber optic conversion kits and Mode Conditioning Patch (MCP) cables are not orderable as features on z9 EC. Fiber optic cables, cable planning, labeling, and installation are all customer responsibilities for new z9 EC installations and upgrades.

IBM Fiber Cabling Services offer a total cable solution service to help with cable ordering needs, and is highly recommended.

### 3.4.7 Cryptographic features

Cryptographic functions on the z9 EC are provided by CPACF and the Crypto Express2 feature.

#### **Crypto Express2 feature (FC 0863)**

No CryptoExpress2 feature needs to be installed; the minimum installed number of CryptoExpress2 features is two. After the initial configuration, the number of features increase one feature at a time up to a maximum of eight.

Each Crypto Express2 feature holds two PCI-X cryptographic adapters that can be configured as coprocessors or accelerators. Either of the adapters can be configured by the installation as a coprocessor or an accelerator.

Each Crypto Express2 feature occupies one I/O slot in an I/O cage, has no CHPIDs assigned, but uses two PCHIDs.

Cryptographic functions are described in Chapter 5, "Cryptography" on page 149.

Archived

# Channel Subsystem

This chapter describes the concept of multiple Channel Subsystems (CSSs) first implemented on the z990 and familiarizes you with the technology, terminology, and implementation aspects of the Channel Subsystem. The Channel Subsystem has been enhanced on the z9 EC with the support of Multiple Subchannel Sets (MSS) and the addition of MIDAW to improve FICON, and in some cases ESCON, performance by reducing channel, director, and Control unit overhead for data sets that utilize striping and compression (VSAM, PDSE, HFS, and zFS).

The following topics are discussed:

- ▶ 4.1, “Channel Subsystem (CSS)” on page 130
- ▶ 4.1.9, “Configuration management” on page 143
- ▶ 4.2, “The MIDAW facility” on page 144

## 4.1 Channel Subsystem (CSS)

The role of the Channel Subsystem is to control communication of internal and external channels to control units and devices. The configuration definitions of the CSS define the operating environment for the correct execution of all system Input/Output (I/O) operations. The CSS provides the server communications to external devices through channel connections. The channels permit transfer of data between main storage and I/O devices or other servers under the control of a channel program. The CSS allows channel I/O operations to continue independently of other operations within the central processors.

The building blocks that make up a Channel Subsystem are listed in Figure 4-1.

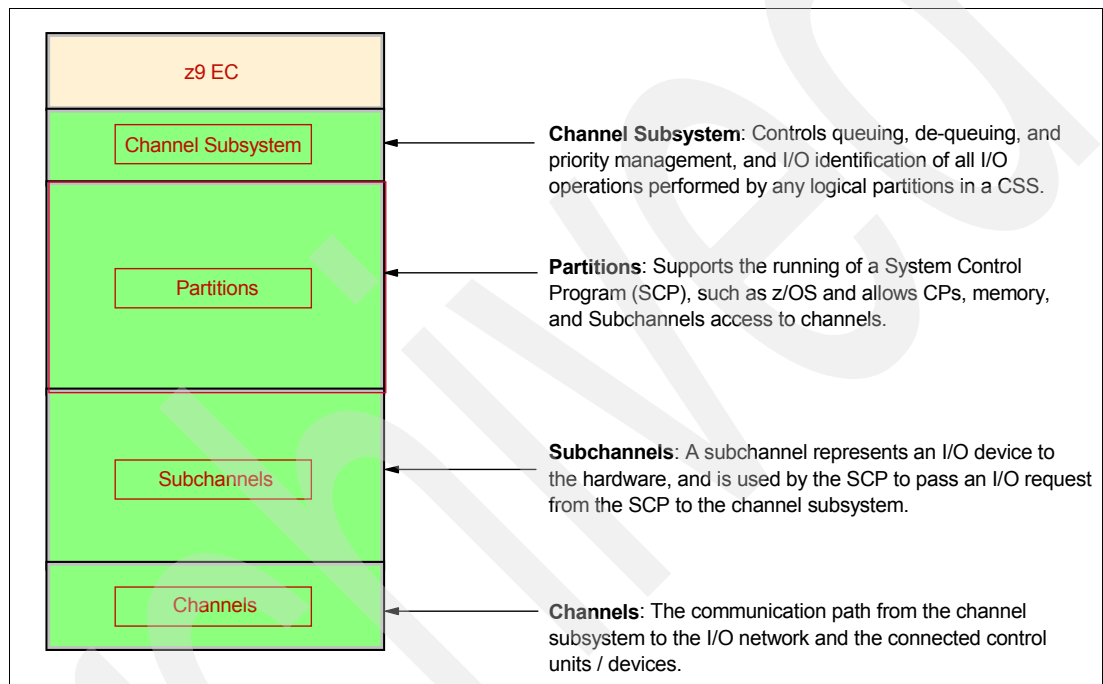


Figure 4-1 Channel Subsystem overview



The structure provides up to four Channel Subsystems (see Figure 4-2). Each CSS has from one to 256 CHPIDs, and may be configured with up to 15 logical partitions that relate to that particular Channel Subsystem. CSSs are numbered from 0 to 3, and are sometimes referred to as the CSS Image ID (CSSID 0, 1, 2, and 3).

**Note:** The z9 EC provides for up to four Channel Subsystems, 1024 CHPIDs, and up to 60 logical partitions for the total system.

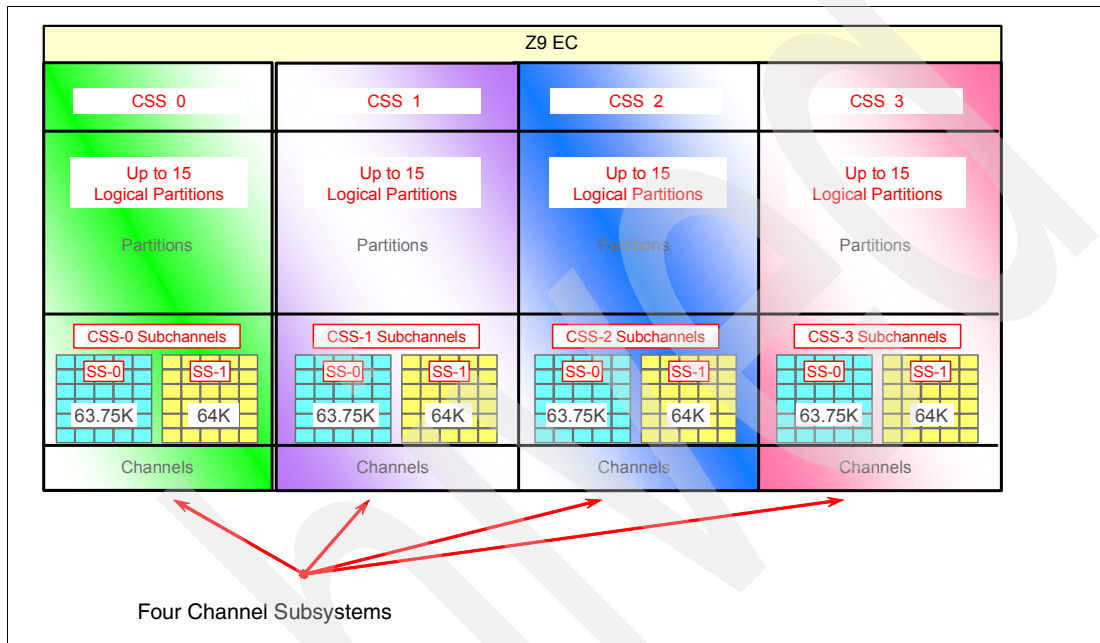


Figure 4-2 Four Channel Subsystems

### 4.1.1 Multiple CSSs concept

The multiple Channel Subsystems concept provides the ability to define more than 256 CHPIDs in System z servers. The z9 EC supports up to four CSSs. The design of the System z servers offers considerable processing power, memory sizes, and I/O connectivity. In support of the larger I/O capability, the CSS concept has been scaled up correspondingly. This provides relief for the number of supported logical partitions, channels, and devices available to the server.

Each CSS may have from 1 to 256 channels and may in turn be configured with 1 to 15 logical partitions, with a maximum of 60 logical partitions on z9 EC servers. CSSs are numbered from 0 to 3 and are sometimes referred to as the CSS Image ID (CSSID 0, 1, 2 or 3).

### 4.1.2 Multiple CSSs structure

The structure of the multiple CSSs provides channel connectivity to the defined logical partitions in a manner that is transparent to subsystems and application programs.

The System z servers provide the ability to define more than 256 CHPIDs in the system through the multiple CSSs. CSS defines CHPIDs, control units, subchannels, and so on. This enables the definition of a balanced configuration for the processor and I/O capabilities.

For ease of management, we strongly recommend that the Hardware Configuration Definitions (HCD) be used to build and control the z9 EC, z9 BC, z890, or z990 input/output configuration definitions. HCD support for multiple Channel Subsystems is available with z/VM and z/OS. HCD provides the capability to make both dynamic hardware and software I/O configuration changes.

A z9 EC must have at least one CSS defined. No logical partitions can exist without at least one defined CSS. Logical partitions are defined to an CSS, not to a server. A logical partition is associated with one CSS only. CHPID numbers are unique within a CSS and range from 00 to FF. However, the same CHPID number can be reused within any other CSS.

All Channel Subsystem Images (CSS Image) are defined within a single I/O configuration data set (IOCDS). The IOCDS is loaded and initialized into the Hardware System Area during Power-on Reset.

There is no HSA expansion support for dynamic I/O on the z9 EC. The HSA allocation is controlled by the “Maximum number of devices” field on the HCD Channel Subsystem List panel. This value can only be changed by a Power-on Reset.

Figure 4-3 shows a logical view of the relationships. It must be noted that each CSS supports up to 15 logical partitions; system wide, a total of up to 60 logical partitions are supported.

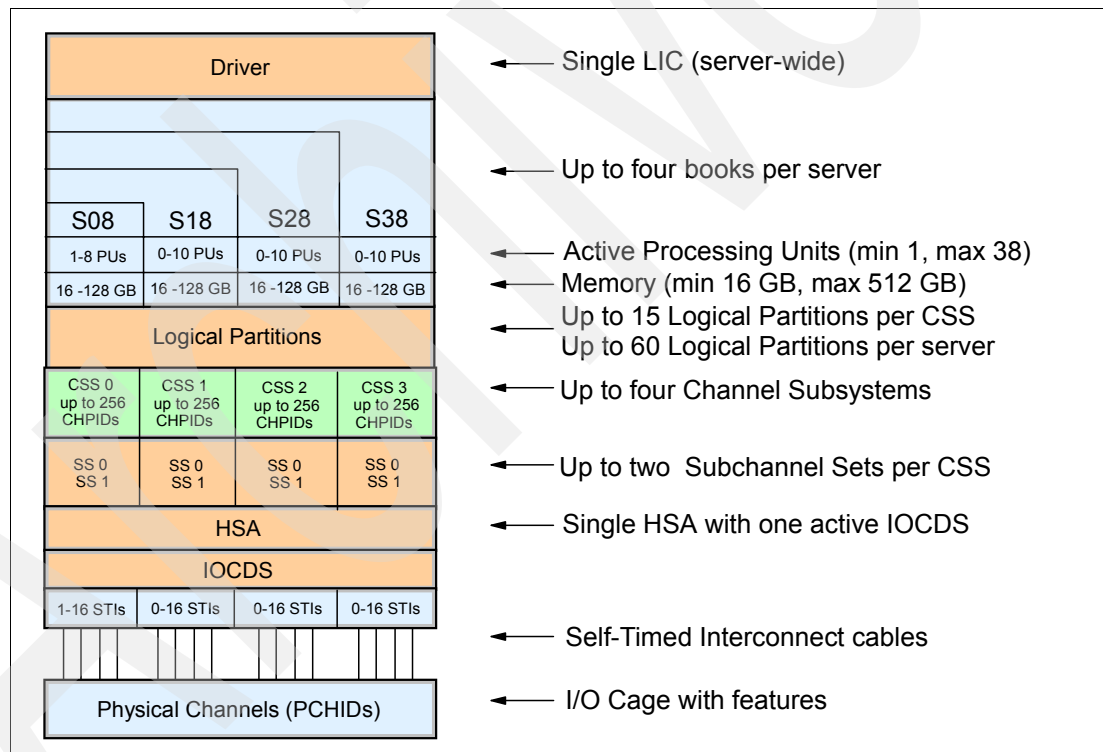


Figure 4-3 Logical view of z9 EC models,<sup>1</sup> CSSs, IOCDS, and HSA

**Note:** The HSA can be moved from one book to a different book in an enhanced availability configuration as part of a concurrent book repair action.

The channel definitions of a CSS are not bound to a single book. A CSS may define resources that are physically connected to any STIs of any book in a multi-book z9 EC.

<sup>1</sup> To keep the diagram simple, Model S54 is not represented.

### 4.1.3 Multiple Subchannel Sets (MSS)

The Multiple Subchannel Sets functionality should not be confused with multiple Channel Subsystems.

In most cases, a subchannel represents an addressable device. A disk control unit with 30 drives uses 30 subchannels (for base addresses), and so forth. An addressable device is associated with a device number and the device number is commonly (but incorrectly) known as the device address.

Subchannel numbers (including their implied path information to a device) are limited to four hexadecimal digits by architecture. Four hexadecimal digits provides 64 K addresses, known as a *set*. IBM reserved 1024 subchannels, leaving 63 K subchannels for general use.<sup>2</sup>

Again, addresses, device numbers, and subchannels are often used as synonyms, although this is not technically correct. We may hear that there is *a maximum of 63 K addresses* or *a maximum of 63 K device numbers*.

The advent of technology such as metro mirror, GDPS®-Hyperswap, global mirror, FlashCopy®, and Parallel Access to Volumes (PAV) has made this limitation (63 K subchannels) a problem for larger installations. For example, a single disk drive (with PAV) often consumes at least four subchannels<sup>3</sup>.

The solution allows *sets* of subchannels (addresses), with a current implementation of two sets. Each set provides 64 K addresses. (Subchannel set 0, the first set, still reserves subchannels for IBM use although the number of reserved subchannels is being reduced from 1024 to 256.) Subchannel set 1 provides a full range of 64 K subchannels.

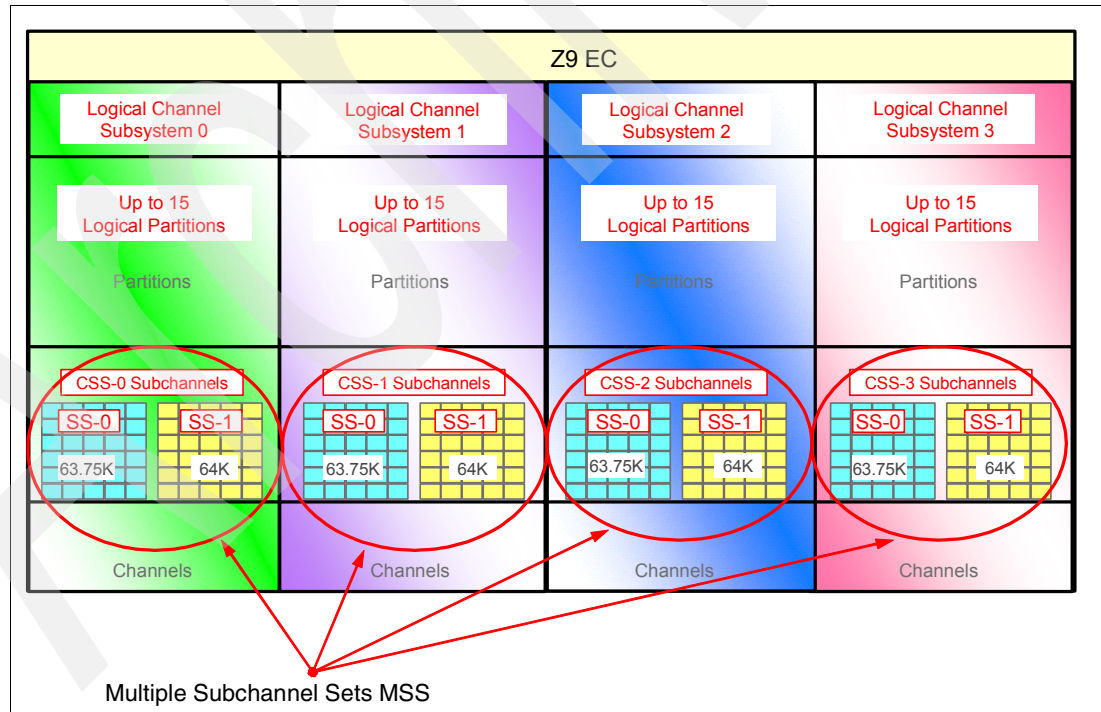


Figure 4-4 Multiple Subchannel Sets (MSS)

<sup>2</sup> The number of reserved subchannels is changed from 1024 to 256. We abbreviate this to 63K in this discussion to easily differentiate it from the 64K subchannels available in subchannel set 1.

<sup>3</sup> Four appears to be a popular number for PAV. It represents the base address and three alias addresses.

In principle, subchannels in either set could be used for any device addressing purpose. However, the current implementation (in z/OS) restricts subchannel set 1 to disk *alias* subchannels. Subchannel set 0 may be used for base addresses and for alias addresses.

There is no required correspondence between addresses in the two sets. For example, we might have device number 8000 in subchannel set 0 and device number 8000 in subchannel set 1 and they might refer to completely separate devices. (We know that the device number in subchannel set 1 must be an alias for z/OS, but that is all we can know from the device number.) Likewise, device number 1234 (subchannel set 0) and device number 4321 (subchannel set 1) might be the base and an alias for the same device. There is no *required* correspondence between the device numbers used in the two subchannel sets.

The additional subchannel set, in effect, adds an extra high-order digit (either 0 or 1) to existing device numbers. For example, we might think of an address as 08000 (subchannel set 0) or 18000 (subchannel set 1). This is not done in system code or in messages because of the architectural requirement for four-digit addresses (device numbers or subchannels). However, some messages will contain the subchannel set number and one can mentally use it as a high-order digit for device numbers. There should be few requirements for this since subchannel set 1 is used only for alias addresses and users, through JCL, messages, or programs rarely refer directly to an alias address.

Moving the alias devices into the second subchannel set has created additional space for device number growth. Although all the aliases have been moved to a subchannel set (Figure 4-5), this is not a requirement. There can still be a mixture of base and alias devices in subchannel set 0.

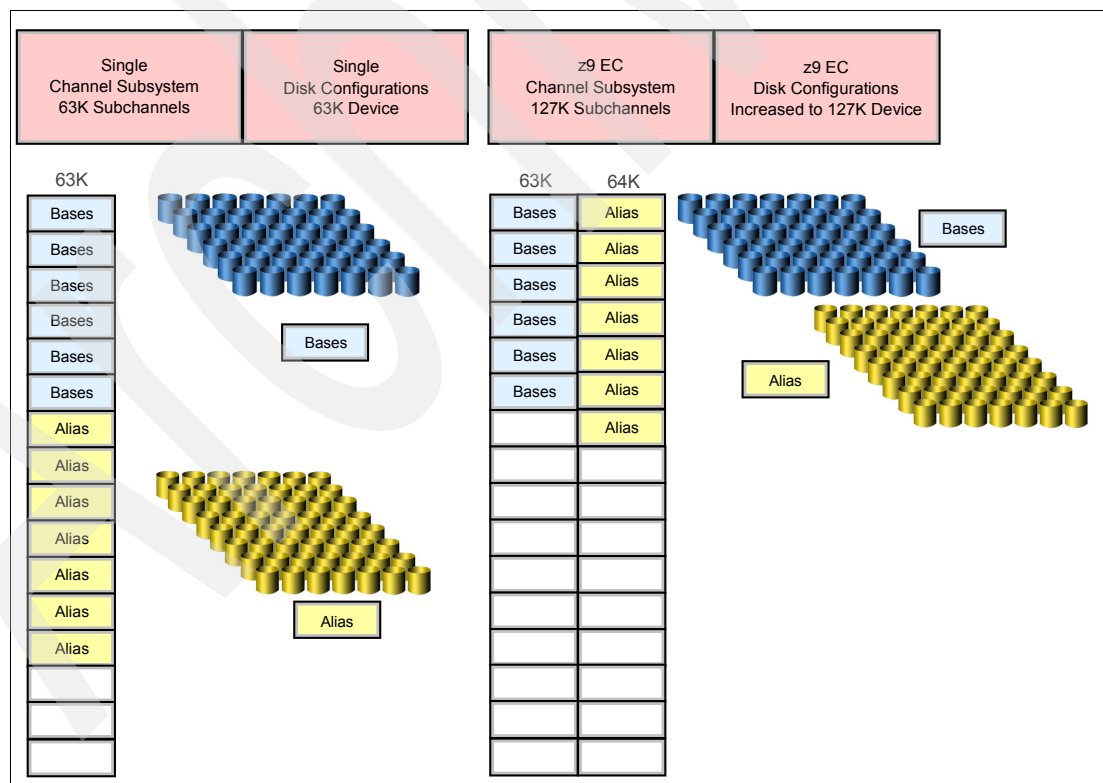


Figure 4-5 Multiple subchannel sets

The appropriate subchannel set number must be included in IOCP definitions (or in the HCD definitions that produce the IOCDs). The subchannel set number defaults to zero and IOCP changes are needed only when using subchannel set 1.

### 63.75 K subchannels

Systems prior to the z9 EC reserved 1024 subchannels out of the potential maximum of 64 K subchannels. The z9 EC has reduced this reserved number to 256 subchannels, thus increasing the number of subchannels available. The reserved subchannels only exist in subchannel set 0; no subchannels are reserved in subchannel set 1.

The informal name, 63.75 K, represents  $(63 \times 1024 + 0.75 \times 1024)$  or 65280.

## 4.1.4 CSS-related numbers

Table 4-1 shows CSS-related information in terms of maximum values for devices, subchannels, logical partitions, and CHPIDs.

Table 4-1 z9 EC CSS at a glance

Setting	z9 EC
Maximum number of CSSs	Four
Maximum number of CHPIDs	1024
Maximum number of LPs supported per CSS	15
Maximum number of LPs supported per system	60
Maximum number of HSA subchannels	7665 K (127.75 K per partition *60 partitions)
Maximum number of devices	511 K (Four CSSs * 127.75 K devices)
Maximum number of CHPIDs per CSS	256
Maximum number of CHPIDs per logical partition	256
Maximum number of devices/subchannels per logical partition	127.75 K

### Multiple Image Facility (MIF)

Multiple Image Facility enables resource sharing across logical partitions within a single CSS or across the multiple CSSs. When a channel resource is shared across logical partitions in multiple CSSs, this is known as *spanning*; refer to 4.1.6, “Channel spanning” on page 138 for more information.

Because of multiple CSSs, the IOCDS logical partition MIF Image ID is not unique within the z9 EC. Therefore, the logical partition identifier value has to provide a unique value for each logical partition within the same z9 EC. The following terminology applies.

► Logical partition number

The logical partition number cannot be specified by the installation; actually, it is not even visible to the user. On the z9 EC, it is assigned at Power-on Reset by PR/SM and is based on the total number of partitions defined in the RESOURCE statements in the IOCDS. It is unique for each logical partition.

► Logical partition identifier

The logical partition identifier is a number in the range from '00' to '3F'. It is assigned by the user on the image profile through the Support Element (SE) or the Hardware Management Console. It is unique across the z9 EC and may also be referred to as the User Logical Partition ID (UPID).

► MIF Image ID (MIFID)

The Multiple Image facility enables channel sharing among logical partitions pertaining to the same Channel Subsystem.

The MIF Image ID is a number that is defined through the Hardware Configuration Dialog (HCD) or directly through the IOCP. It is a number that is specified in the RESOURCE statement in the configuration definitions. It is in the range '1' to 'F' and is unique within an CSS, but it is not unique within the z9 EC. Multiple CSSs may specify the same MIF Image ID.

The combination of CSSID.MIFID is unique across the CPC.

► Logical partition name

This name is user defined through HCD or the IOCP and is the partition name in the RESOURCE statement in the configuration definitions. Each name must be unique across all CSSs defined for the z9 EC.

Figure 4-6 summarizes the identifiers and how they are defined, using a four CSSs configuration example.

Logical Partition Name PROD1 PROD2			Logical Partition Name TST2 PROD3 PROD4			Log Part Name TST3	Logical Partition Name TST4 PROD5		Specified in HCD / IOCP
Logical Partition ID 02 04 0A			Logical Partition ID 14 16 1D			Log Part ID 22	Logical Partition ID 35 3A		Specified in HMC Image Profile
Logical Partition Number			Logical Partition Number			Log Part Number	Logical Partition Number		Assigned by PR/SM at POR
MIF ID 2 4 A			MIF ID 4 6 D			MIF ID 2	MIF ID 5 A		Specified in HCD / IOCP
CSS0 SS 0 SS 1		CSS1 SS 0 SS 1		CSS2 SS 0		CSS3 SS 0 SS 1		Specified in HCD / IOCP	

Figure 4-6 CSSs, logical partition, and identifiers example

We suggest that you establish a naming convention for the logical partition identifiers. As shown in Figure 4-6, you could use the CSS number concatenated to the MIF Image ID, which means logical partition ID 3A is in CSS 3 with MIF ID A. This fits within the allowed range of logical partition IDs and conveys useful information to the user.

**Dynamic addition or deletion of a logical partition name**

In order to have a partition defined for future use, it must be reserved with a CSS beforehand in the IOCDs that is used for Power-On Reset. A reserved partition is defined with a partition name placeholder, a MIF ID, a usage type, and a description. The reserved partition can be assigned a logical partition name to be later used in I/O commands of HCD.

**Important:** Some HCD and HCM panels may still refer the user to the definition of a *logical partition number*. For a z9 EC configuration, this is incorrect, and the user should understand that the panel refers to the definition of a MIF ID.

As previously mentioned, on a z9 EC, the logical partition number is assigned by PR/SM during Power-on Reset and cannot be visualized or modified by the user.

### 4.1.5 Physical Channel ID (PCHID)

A Physical Channel ID, or PCHID, reflects the physical identifier of a channel-type interface. A PCHID number is based on the I/O cage location, the channel feature slot number, and the port number of the channel feature. A CHPID does not directly correspond to a hardware channel port on a z9 EC, and may be arbitrarily assigned. A hardware channel is identified by a PCHID.

You can address 256 CHPIDs within a single Channel Subsystem. That gives a maximum of 1024 CHPIDs when four CSSs are defined. Each CHPID number is associated with a single channel. The physical channel, which uniquely identifies a connector jack on a channel feature, is known by its PCHID number.

PCHIDs identify the physical ports on cards located in I/O cages and follows the numbering scheme shown in Table 4-2.

Table 4-2 PCHIDs numbering scheme

Cage	Front PCHID ##	Rear PCHID ##
I/O Cage 1	100 - 11F	200 - 21F
I/O Cage 2	300 - 31F	400 - 41F
I/O Cage 3	500 - 51F	600 - 61F
CEC cage	000 - 03F reserved for ICB-4s	

CHPIDs are not pre-assigned. It is the responsibility of the installation to assign the CHPID numbers through the use of the CHPID Mapping Tool (CMT) or HCD/IOCP. Assigning CHPIDs means that a CHPID number is associated with a physical channel port location (PCHID), and a CSS. The CHPID number range is still from '00' to 'FF' and must be unique within a CSS. Any non-internal CHPID not defined with a PCHIDs will fail validation when an attempt is made to build a production IODF or an IOCDs.

A pictorial view of a z9 EC with multiple CSSs is shown in Figure 4-7. In this example, two Channel Subsystems are defined (CSS0 and CSS1). Each CSS has three logical partitions with their associated MIF Image Identifiers.

In each CSS, the CHPIDs are shared across all logical partitions. The CHPIDs in each CSS can be mapped to their designated PCHIDs using the CHPID Mapping Tool (CMT), or manually using HCD or IOCP. The output of the CMT is used as input to HCD or the IOCP to establish the CHPID to PCHID assignments. See Appendix B, “CHPID mapping tool” on page 263 for details on the CMT.

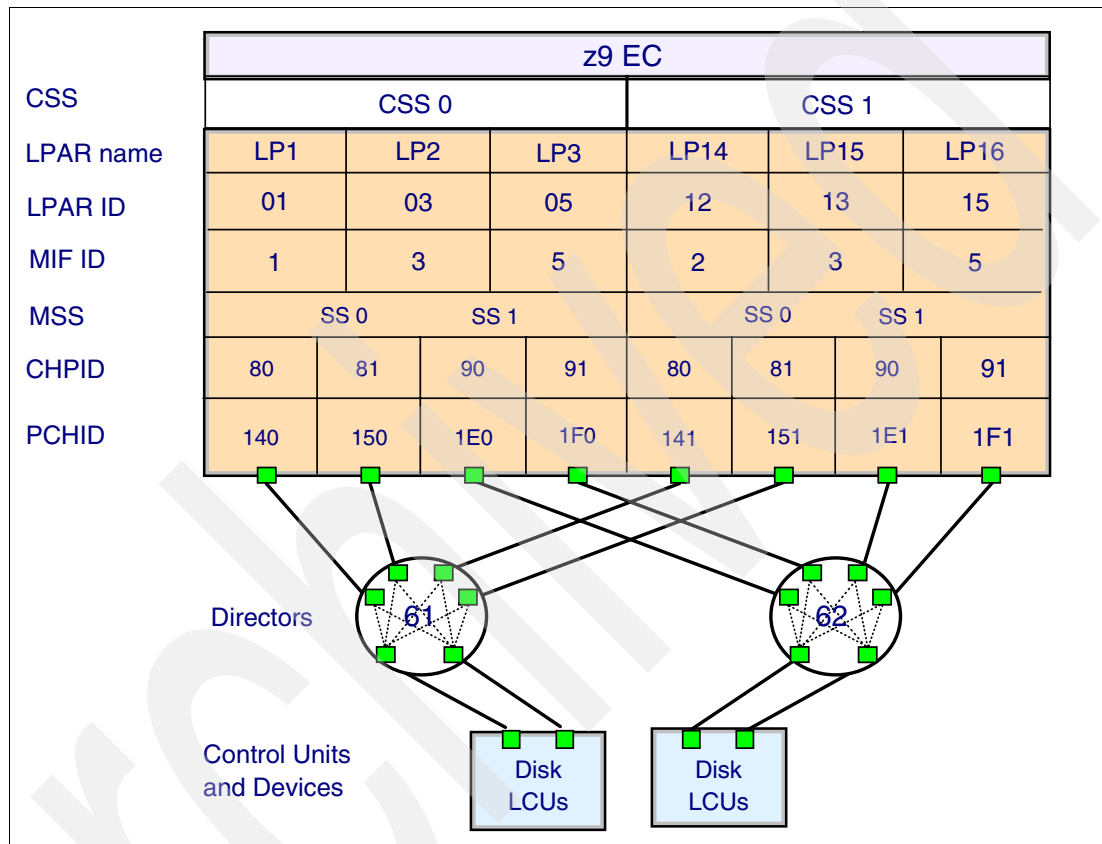


Figure 4-7 z9 EC CSS connectivity

### 4.1.6 Channel spanning

Channel spanning extends the MIF concept of sharing channels across logical partitions to sharing channels across logical partitions *and* Channel Subsystems.

Spanning is the ability for a physical channel (PCHID) to be mapped to CHPIDs defined in multiple Channel Subsystems. When defined that way, the channels can be transparently shared by any or all of the configured logical partitions, regardless of the Channel Subsystem to which the logical partition is configured.

A channel is considered a spanned channel if the same CHPID number in different CSSs is assigned to the same PCHID in IOCP, or is defined as *spanned* in HCD.

In the case of internal channels (for example, IC links and HiperSockets), the same applies, but there is no PCHID association. They are defined with the same CHPID number in multiple CSSs.



CHPIDs that span CSSs reduce the total number of channels available on the z9 EC. The total is reduced, since no CSS can have more than 256 CHPIDs. For a z9 EC with two CSSs, a total of 512 CHPIDs are supported. If all CHPIDs are spanned across the two CSSs, then only 256 channels are supported. For a z9 EC with four CSSs, a total of 1024 CHPIDs are supported. If all CHPIDs are spanned across the four CSSs, then only 256 channels can be supported.

Channel spanning is supported for internal links (HiperSockets and Internal Coupling (IC) links) and for some external links (FICON Express2 channels, OSA-Express2, and Coupling Links).

**Note:** Spanning of ESCON channels and FICON converter (FCV) channels is not supported.

In Figure 4-8, CHPID 04 is spanned to CSS0 and CSS1. Since it is not an external channel link, there is no PCHID assigned. CHPID 06 is an external spanned channel and has a PCHID assigned.

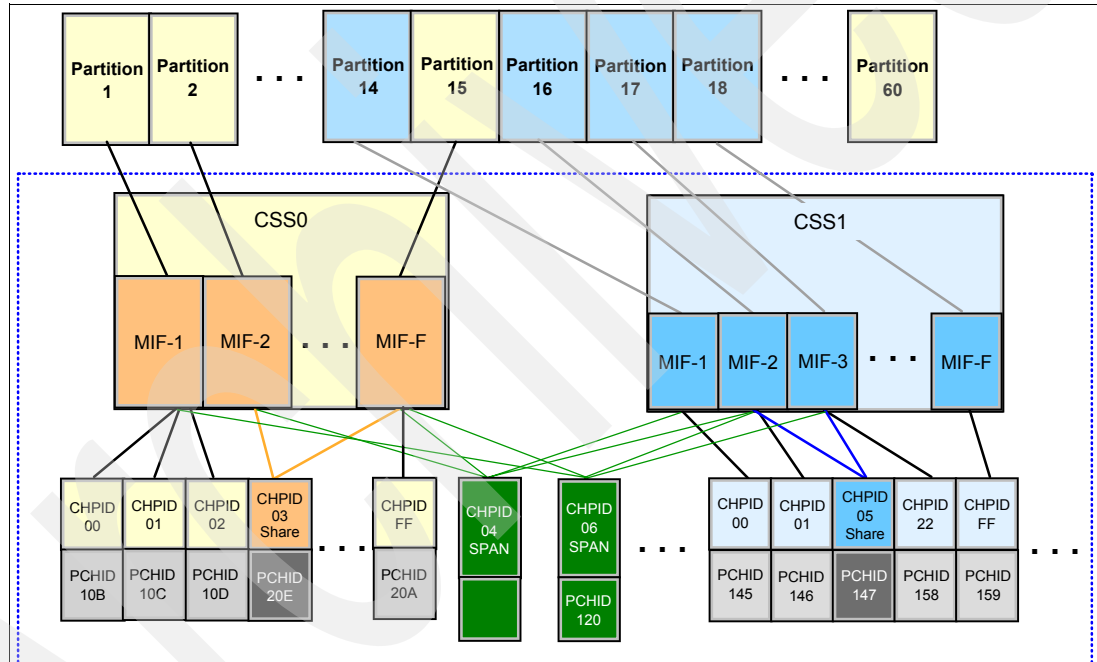


Figure 4-8 z9 EC CSS - Two Channel Subsystems with channel spanning

### 4.1.7 IOCP example

CSS, MSS, and PCHID information is included in an IOCDs. An IOCDs is created through the HCD program or by use of a native IOCP program.

For illustration purposes, we examine an IOCP file and assume the reader is generally familiar with such files.

The following IOCP definition describes the server shown in Figure 4-9. This is not intended to represent a practical server—it has no consoles, for example—but it illustrates the elements involved.

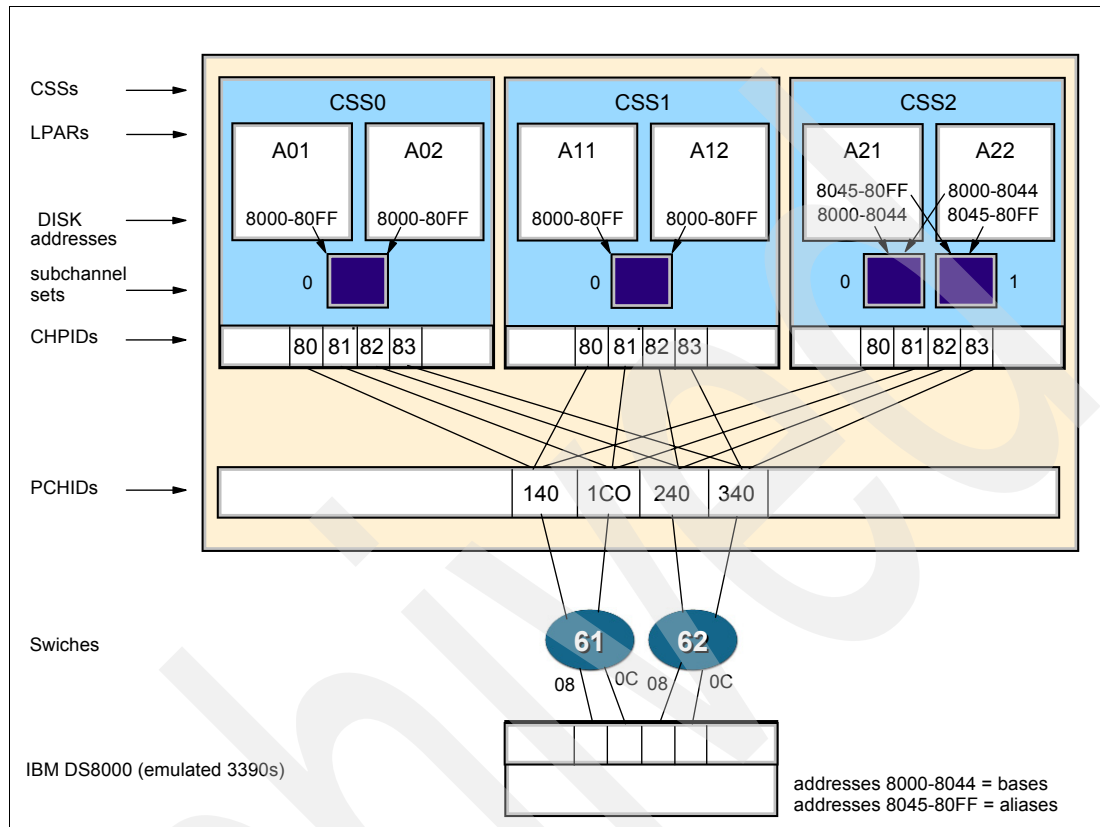


Figure 4-9 IOCP Example with CSSs, MSS, and PCHIDs

```

ID   MSG1='IODF01',
      MSG2='MY.IODF01.WORK - 2006-03-23 10:44',
      SYSTEM=(2094,1),
      TOK=('SCZP101',008001136A3A2084135941210105143F00000000,
          00000000,'06-03-23','10:44',' ',' ')
      RESOURCE PARTITION=((CSS(0),(A01,1),(A02,2)),(CSS(1),(A11,1),
          (*A12,2)),(CSS(2),(A21,1),(A22,2))),
      MAXDEV=((CSS(0),65280,0),(CSS(1),65280,0),(CSS(2),65280,
          65535))
      CHPID PATH=(CSS(0,1,2),80),SHARED,SWITCH=61,PCHID=140,TYPE=FC
      CHPID PATH=(CSS(0,1,2),81),SHARED,SWITCH=61,PCHID=1C0,TYPE=FC
      CHPID PATH=(CSS(0,1,2),82),SHARED,SWITCH=62,PCHID=240,TYPE=FC
      CHPID PATH=(CSS(0,1,2),83),SHARED,SWITCH=62,PCHID=340,TYPE=FC
      CNTLUNIT CUNUMBR=8000,
      PATH=((CSS(0),80,81,82,83),(CSS(1),80,81,82,83),(CSS(2),
          80,81,82,83)),UNITADD=((00,256)),
      LINK=((CSS(0),08,0C,08,0C),(CSS(1),08,0C,08,0C),(CSS(2),
          08,0C,08,0C)),CUADD=0,UNIT=2105
      IODEVICE ADDRESS=(8000,069),CUNUMBR=(8000),STADET=Y,UNIT=3390B
      IODEVICE ADDRESS=(8045,187),CUNUMBR=(8000),STADET=Y,
      SCHSET=((CSS(2),1)),UNIT=3390A
  
```

Key elements in this IOCP include the following:

- ▶ Three Channel Subsystems (CSS0, CSS1, and CSS2) are defined. The CSSs used in the IOCDs created from this IOCP are stated in the RESOURCE statement. This *defines* the CSSs.
- ▶ Two logical partitions are defined in each Channel Subsystem (A01, A02, and so forth). These are also defined in the RESOURCE statement.
- ▶ The number of subchannels (for each Channel Subsystem) is defined in the RESOURCE statement. CSS0 and CSS1 have the single, default subchannel set (with a maximum of 65280 devices that may be defined). CSS2 has two subchannel sets (with 65280 for the first and 65535 for the second as the maximum number of devices).
- ▶ Four spanned channels are used and PCHIDs 140, 1C0, 240, and 340. Since these are spanned (across CSSs), they must have the same CHPIDs in each CSS. (They are assigned 80, 81, 82, and 83 in this example.) Each CHPID statement specifies its CSS number as part of the PATH parameter.
- ▶ Each CHPID statement *must* include a PCHID parameter to associate the CHPID with a physical channel, except for internal CHPID types, such as ICP and IQD. The PCHID parameters can be added to the IOCP definitions by the CHPID Mapping Tool, through HCD definitions, or defined in an IOCP input file, but the PCHID parameters must be present in all CHPID statements in order to create an IOCDs.
- ▶ PATH and LINK parameters in CNTLUNIT statements must indicate the CSS number associated with each path and link.
- ▶ The target disk subsystem has four channel interfaces and 256 unit addresses. It has 69 base devices defined at addresses 8000-8044, and 187 aliases defined at addresses 8045-80FF.
- ▶ CSS0 and CSS1 use subchannels from subchannel set 0 to access both the base and alias addresses. CSS2 uses subchannels from subchannel set 0 base devices and subchannels from subchannel set 1 to access aliases. The first IODEVICE statement has no SCHSET keyword, so all devices default to SS-0. The second IODEVICE statement defaults to SCHSET 0 for CSS0 and CSS1. For CSS2, use SCHSET 1.

This example does not clearly illustrate the device address relief provided by Multiple Subchannel Sets. However, it does exist in CSS2 in the example. Addresses 8045 through 80FF are free in subchannel set 0 and can be used for additional devices. Addresses 8045-80FF are used for alias addresses in CSS0 and CSS1. The alias addresses have been moved to subchannel set 1 in CSS2.

As can be seen from this example, the basic concepts and definitions for multiple channel subsystems are straightforward and consistent with existing IOCP parameters. Equivalent fields have been added to HCD panels for entering CSS, MSS, and PCHID definitions.

## The display ios,config command

The `display ios,config(all)` command, shown in Figure 4-10, includes information about the MSS for the z9 EC.

Another thing to note about this is that we have defined SS-0 and SS-1 but have no devices defined to SS-1 as yet. The maxdev definition in the IOCDs is set to the largest number for MSS-1 at 65335.

```
D IOS,CONFIG(ALL)
IOS506I 13.40.02 I/O CONFIG DATA 908
ACTIVE IODF DATA SET = SYS6.IODF19
CONFIGURATION ID = TEST2094      EDT ID = 01
TOKEN: PROCESSOR DATE      TIME      DESCRIPTION
SOURCE: SCZP101 05-07-26 15:51:03 SYS6      IODF19
ACTIVE CSS: 2      SUBCHANNEL SETS IN USE: 0, 1
HARDWARE SYSTEM AREA AVAILABLE FOR CONFIGURATION CHANGES
PHYSICAL CONTROL UNITS          8133
CSS 0 - LOGICAL CONTROL UNITS    4037
  SS 0  SUBCHANNELS              57888
  SS 1  SUBCHANNELS              65335
CSS 1 - LOGICAL CONTROL UNITS    4041
  SS 0  SUBCHANNELS              58144
  SS 1  SUBCHANNELS              65335
CSS 2 - LOGICAL CONTROL UNITS    4041
  SS 0  SUBCHANNELS              58144
ELIGIBLE DEVICE TABLE LATCH COUNTS
      0 OUTSTANDING BINDS ON PRIMARY EDT
IEA631I OPERATOR HEWITT  NOW INACTIVE, SYSTEM=SC76 ,
```

Figure 4-10 `D ios,config(all)` display with MSS

### 4.1.8 IODF Version 5

The HCD version shipped in z/OS V1R7 generates an IODF with a Version 5 format. Planning is important when migrating to a Version 5 IODF. You must make sure that coexistence code is installed on z/OS V1R6 systems that access the IODF.

Compatibility support (SPE) is provided so that activation functions can be processed using a pre-V1.7 release of z/OS when using the IODF V5 format. For z/OS V1R6 HCD, SPE OA07875 is provided that allows read access to an IODF V5.

In z/OS V1.7, HCD provides a conversion function to upgrade from a V4 to a V5 IODF.

When accessing an IODF in V4 format from a z/OS 1.7 system, the IODF format is converted to an in-storage IODF V5, and message CBDG549 is issued to inform the user that a back-level IODF is being accessed. However, as long as no migration is requested, the V5 IODF format will not be saved and the copy on disk will remain in the V4 IODF format.

After migration to the Version 5 IODF, only a z/OS V1R7 system can make updates to this IODF version.

Table 4-3 summarizes the combinations to be considered for a migration from IODF V4 to V5.

Table 4-3 IODF migration to V5

Version of IODF	Function	z/OS V1R7	z/OS V1R6	
			Compatibility SPE installed	Compatibility SPE <i>not</i> installed
IODF V4 created by HCD prior to z/OS V1R7	HCD Read IODF	Yes	Yes	Yes
	HCD Write to IODF	NO Requires migration to IODF V5	Yes	Yes
	IPL	Yes	Yes	Yes
	Dynamic Activation	Yes	Yes	Yes
IODF V5 created by HCD z/OS V1R7	HCD Read IODF	Yes	Yes	NO
	HCD Write to IODF	Yes	NO	NO
	IPL	Yes	Yes	NO
	Dynamic Activation	Yes	Yes	NO

### 4.1.9 Configuration management

Tools are provided to help maintain and optimize the I/O configuration of a z9 EC.

► IBM Configurator for e-business (e-Config)

The e-Config tool is available to your IBM representative. It is used to configure new configurations or upgrades of an existing configuration, and maintains installed features of those configurations. Reports produced by e-Config will be helpful in understanding the changes being made for a system upgrade and what the final configuration will look like.

► Hardware Configuration Dialog (HCD)

HCD supplies an interactive dialog to generate your I/O definition file (IODF) and subsequently your Input/Output Configuration Data Set (IOCDS). It is strongly recommended that HCD or HCM be used to generate your I/O configuration, as opposed to writing your own IOCP. The validation checking that HCD performs as you enter data helps minimize the risk of errors before you implement your I/O configuration.

► z9 EC CHPID Mapping Tool (CMT)

The CHPID Mapping Tool provides a mechanism to map CHPIDs onto PCHIDs as required on a z9 EC. Additional enhancements have been built into the CMT to cater for the requirements of the z9 EC. It provides the best availability recommendations for the installed z9 EC features and defined configuration.

For further details on the CMT, see Appendix B, “CHPID mapping tool” on page 263.

### 4.1.10 System-initiated CHPID reconfiguration

System-initiated CHPID reconfiguration function is designed to reduce the duration of a repair action and minimize operator interaction when an ESCON or FICON channel, an OSA port, or an ISC-3 link is shared across logical partitions on IBM System z9 server. When an I/O card is replaced for a repair, it usually has some failed channels and some still functioning channels. To remove the card, all channels need to be configured offline from all logical partitions sharing those channels. Without system-initiated CHPID reconfiguration, this means the CE must contact each logical partition operators and have them set the channels offline, and then after the repair, contact them again to configure the channels back online.

With system-initiated CHPID reconfiguration support, the System z9 Support Element sends a signal to the IOP that a channel needs to be configured offline. The IOP determines all the logical partitions sharing that channel and sends an alert to the operating systems in those logical partitions. The operating system then configures the channel offline without any operator intervention. This is repeated for each channel on the card. When the card is replaced, Support Element sends another signal to the IOP for each channel. This time the IOP alerts the operating system that the channel should be configured back online. This is designed to minimize operator interaction to configure channels offline and online.

System-initiated CHPID reconfiguration is supported by z/OS V1R6 and later.

### 4.1.11 Multipath Initial Program Load (IPL)

Multipath IPL is designed to help increase availability and to help eliminate manual problem determination when executing an IPL. It does so by allowing IPL to complete if possible using alternate paths when executing an IPL from a device connected through ESCON and FICON channels. If an error occurs, an alternate path is selected.

Multipath IPL is applicable to ESCON channels (CHPID type CNC) and to FICON channels (CHPID type FC).

z/OS V1R6 and later supports multipath IPL.

## 4.2 The MIDAW facility

The Modified Indirect Data Address Word (MIDAW) facility is a system architecture and software exploitation designed to improve FICON performance. This facility is only available on System z9 servers and is exploited by media manager in z/OS V1.6 (with PTFs) and later releases. The MIDAW facility provides a more efficient CCW/IDAW structure for certain categories of data chaining I/O operations.

- ▶ MIDAW can significantly improve FICON performance for extended format data sets.
  - Non-extended data sets can also benefit from MIDAW.
- ▶ MIDAW can improve channel utilization and can significantly improve I/O response time.
  - Reduces FICON channel connect time, director ports, and control unit overhead.

MIDAW is for Modified IDAW. An IDAW is an Indirect Address Word that is used to specify data addresses for I/O operations in a virtual environment.<sup>4</sup> The existing IDAW design allows the first IDAW in a list to point to any address within a page. Subsequent IDAWs in the same list must point to the first byte in a page; also, all but the first and last IDAW in a list must deal with complete 2K or 4K units of data. This limits the usability of IDAWs to straightforward

Figure 4-11 shows a single CCW to control the transfer of data that spans non-contiguous 4 K frames in main storage. When the IDA flag is set, the data address in the CCW points to a list of words (IDAWs), each of which contains an address designating a data area within real storage.

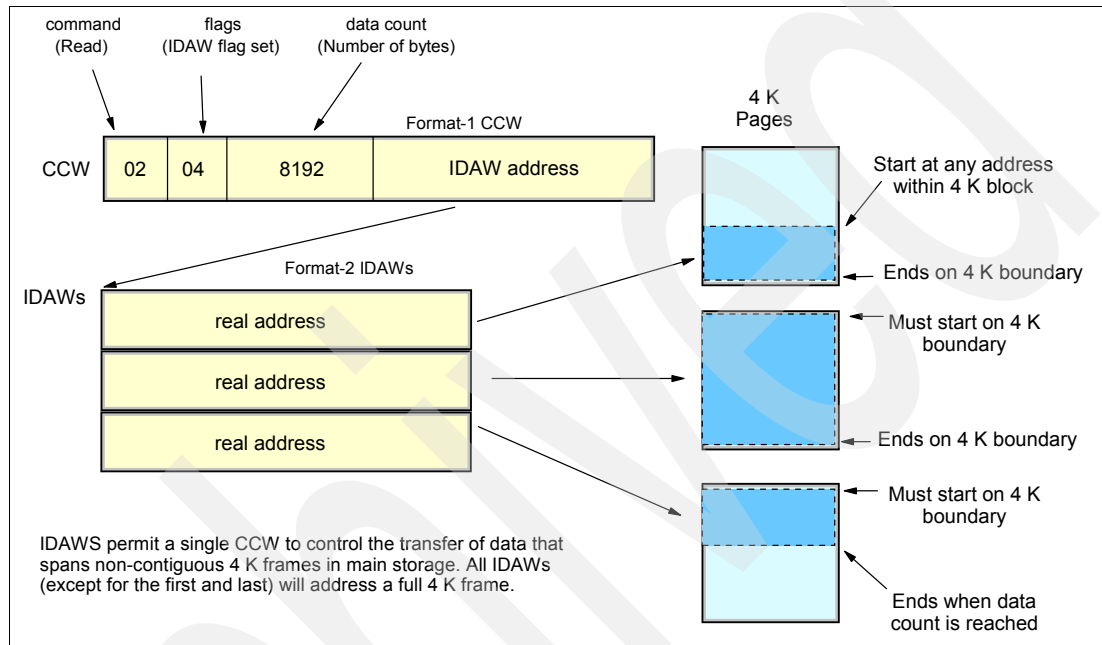


Figure 4-11 IDAW usage

The number of IDAWs required for a CCW is determined by the IDAW format as specified in the ORB, by the count field of the CCW, and by the data address in the initial IDAW. When, for example, (1) the ORB specifies format-2 IDAWs with 4K-byte blocks, (2) the CCW count field specifies 8K bytes, and (3) the first IDAW designates a location in the middle of a 4K-byte block, then three IDAWs are required.

CCWs with *data chaining* may be used to process I/O data blocks that have a more complex internal structure in which portions of the data block are directed into separate buffer areas (this is sometimes known as scatter-read or scatter-write). However, as technology evolves and link speed increases, data chaining techniques are becoming less efficient in modern I/O environments for reasons involving switch fabrics, control unit processing and exchanges, and so on.

<sup>4</sup> There are exceptions to this statement and we skip a number of details in the following description. We assume the reader will merge this brief description with an existing understanding of I/O operations in a virtual memory environment.

The MIDAW facility is a method of gathering/scattering data from/into discontinuous storage locations during an I/O operation. The modified IDAW (MIDAW) format is shown in Figure 4-12. It is 16 bytes long and is aligned on a quadword.

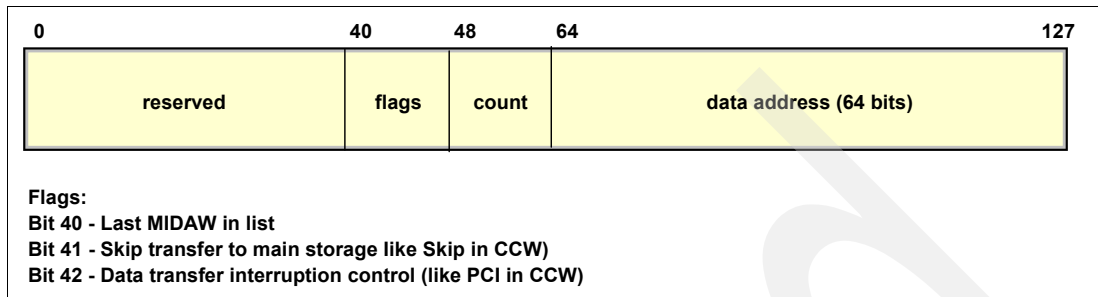


Figure 4-12 MIDAW format

An example of MIDAW usage is shown in Figure 4-13.

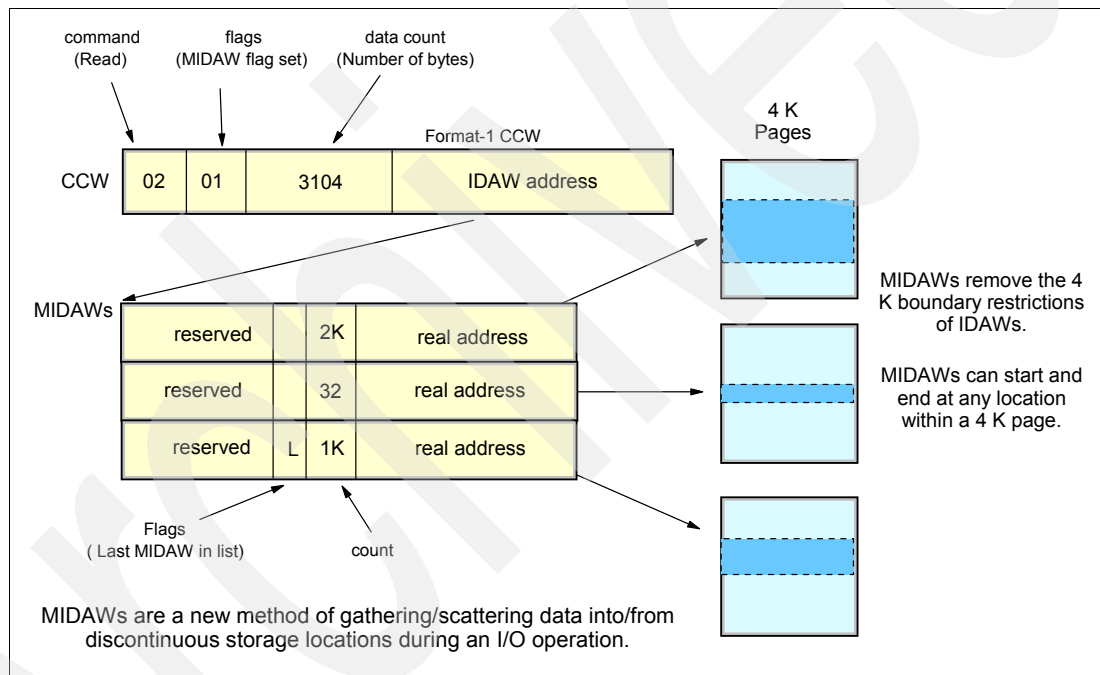


Figure 4-13 MIDAW usage

The use of MIDAWs is indicated by the MIDAW bit in the CCW. If this bit is set, then the *skip flag* may not be set in the CCW; the skip flag in the MIDAW may be used instead. The data count in the CCW should equal the sum of the data counts in the MIDAWs. The CCW operation ends when the CCW count goes to zero or the last MIDAW (with the *last* flag) ends. The combination of the address and count in a MIDAW cannot cross a page boundary; this means the largest possible count is 4 K. The maximum data count of all the MIDAWs in a list cannot exceed 64 K. (This is because the associated CCW count cannot exceed 64 K.)

The scatter-read or scatter-write effect of the MIDAWs makes it possible to efficiently send small control blocks embedded in a disk record to separate buffers than those used for larger data areas within the record. MIDAW operations are on a single I/O block, in the manner of data chaining. Do not confuse this operation with CCW *command* chaining.



### **Extended format data sets**

z/OS extended format data sets use internal structures (usually not visible to the application program) that require scatter-read (or scatter-write) operation. This means that CCW data chaining is required and this produces less than optimal I/O performance. Since the most significant performance benefit of MIDAWs is achieved with Extended Format (EF) data sets, a brief review of the EF data sets are included here.

Both VSAM and non-VSAM (DSORG=PS) can be defined as extended format data sets. In the case of non-VSAM data sets, a 32-byte suffix is appended to the end of every physical record (that is, block) on disk. VSAM appends the suffix to the end of every Control Interval, which normally corresponds to a physical record (a 32 K CI is split into two records in order to span tracks.) This suffix is used to improve data reliability and facilitates other functions described below. Thus, for example, if the DCB BLKSIZE or VSAM CI Size is equal to 8192, the actual block on DASD consists of 8224 bytes. The control unit itself does not distinguish between suffixes and user data, The suffix is transparent to the access method or database.

Besides reliability, EF data sets enable three other functions: DFSMS™ striping, access method compression, and Extended Addressability (EA). EA is especially useful for creating large DB2 partitions (larger than 4 GB). Striping can be used to increase sequential throughput, or to spread random I/Os across multiple logical volumes. DFSMS striping is especially useful for utilizing multiple channels in parallel for one data set. The DB2 logs are often striped to optimize the performance of DB2 sequential inserts.

To process an I/O operation to an EF data set would normally require at least two CCWs with data chaining. One CCW would be used for the 32 byte suffix of the EF data set. With MIDAW, the additional CCW for the EF data set suffix can be eliminated.

MIDAWs benefit both EF and non-EF data sets. For example, to read twelve 4 K records from a non-EF data set on a 3390 track, media manager would chain 12 CCWs together using data chaining. To read twelve 4 K records from an EF data set, 24 CCWs would be chained (two CCWs per 4 K record). Using media manager track-level command operations and MIDAWs, a whole track can be transferred using a single CCW.

### **Performance benefits**

Media Manager has the I/O channel programs support for implementing Extended Format data sets, it automatically exploits MIDAWs when appropriate. Today, most disk I/Os in the system are generated using media manager.

Users of the EXCPVR may construct channel programs containing MIDAWs provided they construct an IOBE with the IOBEMIDA bit set. Users of EXCP *may not* construct channel programs containing MIDAWs.

The MIDAW facility removes the 4 K boundary restrictions of IDAWs and in the case of EF data sets, reduce the number of CCWs. Decreasing the number of CCWs helps to reduce the FICON channel processor utilization. Media manager and MIDAWs will not cause the bits to move any faster across the FICON link, but they reduce the number of frames and sequences flowing across the link, thus utilizing the channel resources more efficiently.

Use of the MIDAW facility with FICON Express4, operating at 4 Gbps, compared to use of Indirect Data Address Words (IDAWs) with FICON Express2, operating at 2 Gbps, showed an improvement in throughput of greater than 220% for all reads (270 MBps versus 84 MBps) on DB2 table scan tests with Extended Format datasets.

These measurements are examples of what has been achieved in a laboratory environment using one FICON Express4 channel operating at 4 Gbps (CHPID type FC) on a z9 EC with z/OS V1.7 and DB2 UDB for z/OS V8.

The performance of a specific workload may vary according to the conditions and hardware configuration of the environment. IBM laboratory tests found that DB2 gains significant performance benefits using the MIDAW facility in the following areas:

- ▶ Table scans
- ▶ Logging
- ▶ Utilities
- ▶ Using DFSMS striping for DB2 data sets

Media manager with the MIDAW facility can provide significant performance benefits when used in combination applications that use EF data sets (such as DB2) or long chains of small blocks.

For additional information relating to FICON and MIDAW, check the following Web site for the material on FICON channel performance:

<http://www.ibm.com/systems/z/connectivity/>

See also the book *IBM TotalStorage® DS8000™ Series: Performance Monitoring and Tuning*, SG24-7146.



# Cryptography

This chapter describes the hardware cryptographic functions available on the z9 EC. As for the z990 and z890, the Cryptographic Assist Architecture (CAA), along with the CP Assist for Cryptographic Function, offer a balanced use of resources and unmatched scalability.

Included in this chapter are the following topics:

- ▶ 5.1, “Cryptographic functions” on page 150
- ▶ 5.2, “CP Assist for Cryptographic Function (CPACF)” on page 154
- ▶ 5.3, “Crypto Express2” on page 154
- ▶ 5.4, “TKE workstation feature” on page 159
- ▶ 5.5, “Cryptographic functions comparison” on page 160
- ▶ 5.6, “Software support” on page 161

## 5.1 Cryptographic functions

The z9 EC includes both standard cryptographic hardware and optional cryptographic features for flexibility and growth capability. IBM has a long history of providing hardware cryptographic solutions, from the development of Data Encryption Standard (DES) in the 1970s to delivering integrated cryptographic hardware in a server to achieve the US Government's highest FIPS 140-2 Level 4 rating for secure cryptographic hardware.

The z9 EC cryptographic functions include the full range of cryptographic operations needed for e-business, e-commerce, and financial institution applications. In addition, custom cryptographic functions can be added to the set of functions that the z9 EC offers.

Today, e-business applications are increasingly relying on cryptographic techniques to provide the confidentiality and authentication required in this environment. Secure Sockets Layer/Transport Layer Security (SSL/TLS) technology is a key technology for conducting secure e-commerce using Web servers, and it is in use by a rapidly increasing number of e-business applications, demanding new levels of security and performance.

### 5.1.1 Cryptographic synchronous functions

Synchronous cryptographic functions are provided by the CP Assist for Cryptographic Function (CPACF).

The z9 EC hardware includes the implementation of algorithms as hardware synchronous operations, that is, holding the PU processing of the instruction flow until the operation has completed. Note that keys, when needed, are to be provided in clear form only.

- ▶ Data encryption/decryption algorithms
  - Data Encryption Standard (DES)
    - Double length-key DES
    - Triple length-key DES (TDES)
  - Advanced Encryption Standard (AES) for 128-bit keys
- ▶ Hashing algorithms: SHA-1 and SHA-256
- ▶ Message authentication code (MAC):
  - Single-key MAC
  - Double-key MAC
- ▶ Pseudo Random Number Generation (PRNG)

### 5.1.2 Cryptographic asynchronous functions

Asynchronous cryptographic functions are provided by the PCI-X cryptographic adapters.

The following secure key functions are provided as cryptographic asynchronous functions. System internal messages are passed to the cryptographic coprocessors to initiate the operation and messages are passed back from the coprocessors to signal completion of the operation.

- ▶ Data encryption/decryption algorithms
  - Data Encryption Standard (DES)
  - Double length-key DES
  - Triple length- key DES

- ▶ DES key generation and distribution
- ▶ PIN generation, verification, and translation functions
- ▶ Pseudo Random Number Generator (PRNG)
- ▶ Public Key Algorithm (PKA) Facility

These commands are intended for application programs using public key algorithms, including:

- Importing RSA public-private key pairs in clear and encrypted forms
- Rivest-Shamir-Adelman (RSA)
  - Key generation, up to 2048-bit
  - Signature verification, up to 2048-bit
  - Import and export of DES keys under an RSA key, up to 2048-bit
- Public Key Encrypt (PKE)

Public Key Encrypt service is provided for assisting the SSL/TLS handshake. When used with the Mod\_Raised\_to Power (MRP) function, it is also used to offload compute-intensive portions of the Diffie-Hellman protocol onto the PCI-X cryptographic adapter.

- Public Key Decrypt (PKD)

Public Key Decrypt supports a Zero-Pad option for clear RSA private keys. PKD is used as an accelerator for raw RSA private operations, such as those required by the SSL/TLS handshake and digital signature generation. The Zero-Pad option is exploited by Linux to allow use of PCI-X cryptographic adapter for improved performance of digital signature generation.

- Derived Unique Key Per Transaction (DUKPT)

The service is provided to write applications that implement the DUKPT algorithms as defined by the ANSI X9.24 standard. DUKPT provides additional security for point-of-sale transactions that are standard in the retail industry. DUKPT algorithms are supported on the Crypto Express2 feature coprocessor for triple-DES with double-length keys.

- Europay Mastercard VISA (EMV) 2000 standard

Applications may be written to comply with the EMV 2000 standard for financial transactions between heterogeneous hardware and software. Support for EMV 2000 applies only to the Crypto Express2 feature coprocessor of the z9 EC.

Other key functionalities of the Crypto Express2 feature serve to enhance the security of public/private key encryption processing:

- ▶ Remote Loading of Initial ATM Keys

Provides the ability to remotely load the initial ATM keys. Remote Key Loading refers to the process of loading DES keys to ATM from a central administrative site without the need for personnel to visit each machine to manually load the DES keys. The process uses ICSF callable services along with the Crypto Express2 feature to perform the remote load.

ICSF has added two callable services, Trusted Block Create (CSNDTBC) and Remote Key Export (CSNDRKX). CSNDTBC is a callable service that is used to create a trusted block containing a public key and some processing rules. This callable service is used to create a trusted block containing a public key and some processing rules. The rules define the ways and formats in which keys are generated and exported. CSNDRKX is a callable service that uses the trusted block to generate or export DES keys for local use and for distribution to an ATM or other remote device. The PKA Key Import (CSNDPKI), PKA Key Token Change (CSNDKTC), and Digital Signature Verify (CSFNDFV) callable Services were changed and are used to support remote key loading.

▶ Key Exchange with Non-CCA Cryptographic Systems

Allows for the changing of the operational keys between the remote site and the non-CCA system like the ATM. IBM Common Cryptographic Architecture (CCA) employs Control Vectors to control usage of cryptographic keys. Non-CCA systems use other mechanisms, or may use keys that have no associated control information. The key exchange functions added to CCA enhance the ability to exchange keys between CCA systems, and systems that do not use Control Vectors, by allowing the CCA system owner to define permitted types of key import and export while preventing uncontrolled key exchange that can open the system to an increased threat of attack.

▶ Support for ISO 16609 CBC Mode T-DES MAC

In support of ISO 16609:2004, the cryptographic facilities in the System z9 support the requirements for Message Authentication, using symmetric techniques. The Crypto Express2 provides the ISO 16609 CBC Mode T-DES MAC support. This support is accessible through ICSF callable services. ICSF callable services used to invoke the support are MAC Generate (CSNBMGN) and MAC Verify (CSNVMVR).

▶ Retained key support (RSA private keys generated and kept stored within the secure hardware boundary)

▶ Support for 4753 Network Security Processor migration

▶ User-Defined Extensions (UDX) support, including:

- For Activate UDX requests:
  - Establish Owner
  - Relinquish Owner
  - Emergency Burn of Segment
  - Remote Burn of Segment
- Import UDX File function
- Reset UDX to IBM default function
- Query UDX Level function

UDX allows the user to add customized operations to a cryptographic processor. User-Defined Extensions to the Common Cryptographic Architecture (CCA) support customized operations that execute within the Crypto Express2 feature. UDX is supported through an IBM, or approved third-party, service offering.

More information can be found on the IBM cryptocards Web site:

<http://www.ibm.com/security/cryptocards>

The Web site will direct the customer's request to an IBM Global Services location appropriate for the customer's geographic location. A special contract will be negotiated between IBM Global Services and the customer, covering development of the UDX by IBM Global Services per the customer's specifications, as well as an agreed-upon level of the UDX.

Under a special contract with IBM, Crypto Express2 feature customers will gain the flexibility to define and load custom cryptographic functions themselves. This service offering can be requested through the IBM Cryptocards Web site indicated above by selecting the **Custom programming** option.

### 5.1.3 Cryptographic feature codes

What follows is a list of the cryptographic features available with the z9 EC.

Feature code	Description
3863	CPACF DES/TDES enablement The enablement feature is a prerequisite to use CPACF (except for SHA-1 and SHA-256) and Crypto Express2 features.
0863	Crypto Express2 feature A maximum of eight features may be ordered. Each feature contains two PCI-X cryptographic adapters.
0839	Trusted Key Entry 5.0 (TKE 5.0) workstation The TKE 5.0 workstation is optional. It offers local and remote key management and supports connectivity to an Ethernet LAN at operating speeds of 10, 100, and 1000 Mbps. This workstation may also be used to control z9 BC, z990, z890, and z900 servers. It is not designed to control a z800. Up to three features per z9 EC may be installed.
0859	Trusted Key Entry 5.0 (TKE 5.0) workstation The TKE 5.0 workstation is optional. It offers local and remote key management and supports connectivity to an Ethernet LAN at operating speeds of 10, 100, and 1000 Mbps. It may also be used to control z9 BC, z990, z890, z900, and z800 servers. Up to three features per z9 EC may be installed.
0855	TKE 5.0 Licensed Internal Code
0856	TKE 5.1 Licensed Internal Code
0887	TKE Smart Card Reader
0888	TKE additional smart cards

If the TKE option is chosen for key management of the PCI-X cryptographic adapters, a TKE workstation with the TKE 5.0 LIC or later is required.

**Note:** The PCI CC (#0861), PCI CA (#0862), and PCI-X CC (#0868) features are *not* available on the z9 EC.

**Important:** Products that include any of the cryptographic feature codes contain cryptographic functions that are subject to special export licensing requirements by the United States Department of Commerce. It is the customer's responsibility to understand and adhere to these regulations whenever moving, selling, or transferring these products.

## 5.2 CP Assist for Cryptographic Function (CPACF)

The CP Assist for Cryptographic Function offers a set of symmetric cryptographic functions that enhance the encryption and decryption performance of clear key operations for SSL, VPN, and data storing applications that do not require FIPS 140-2 level 4 security.

Cryptographic keys must be protected by the application system, as these keys have to be provided in clear form to the CPACF.

The CP Assist for Cryptographic Function feature provides hardware acceleration for DES, TDES, AES (128 bit), MAC, SHA-1, and SHA-256 cryptographic services. It provides high-performance hardware encryption, decryption, and hashing support.

The following five instructions support the cryptographic assist function:

<b>KMAC</b>	Compute Message Authentic Code
<b>KM</b>	Cipher Message
<b>KMC</b>	Cipher message with chaining
<b>KIMD</b>	Compute Intermediate Message Digest
<b>KLMD</b>	Compute Last Message Digest

The functions are provided as problem state z/Architecture instructions, directly available to application programs. When enabled, the CP Assist for Cryptographic Function runs at z9 EC processor speed and since the facility is available on every CP and IFL in the system, there are no affinity issues.

The cryptographic architecture includes DES, T-DES, AES data encryption and decryption, MAC message authentication, and SHA-1 and SHA-256 hashing.

The functions of the CP Assist for Cryptographic Function must be explicitly enabled using FC 3863, by the manufacturing process or at the customer site as an MES installation, except for SHA-1 and SHA-256, which are always enabled.

## 5.3 Crypto Express2

The Crypto Express2 feature has two PCI-X cryptographic adapters. Each of the PCI-X cryptographic adapters can be configured as a cryptographic coprocessor or a cryptographic accelerator.

Reconfiguration of the PCI-X cryptographic adapter between coprocessor and accelerator mode is an exclusive of the z9 EC system and is also supported for Crypto Express2 features brought forward from z990 and z890 systems to the z9 EC.

- ▶ When the PCI-X cryptographic adapter is configured as a coprocessor, the adapter provides equivalent functions (plus some additional) as the PCICC card on previous systems with a higher level of performance. When the PCI-X adapter is configured as a coprocessor, the adapter also provides equivalent functions (plus some additional) as the PCI CA card on previous systems with the same level of performance.



- ▶ When the PCI-X cryptographic adapter is configured as an accelerator, it provides PCICA-equivalent functions with an expected throughput of approximately three times the PCICA throughput on previous systems.

The z9 EC supports up to eight Crypto Express2 features (up to sixteen PCI-X cryptographic adapters) to be installed. Each PCI-X adapter either act as cryptographic coprocessor or as cryptographic accelerator.

A physical layout of the Crypto Express2 feature is shown in Figure 5-1.

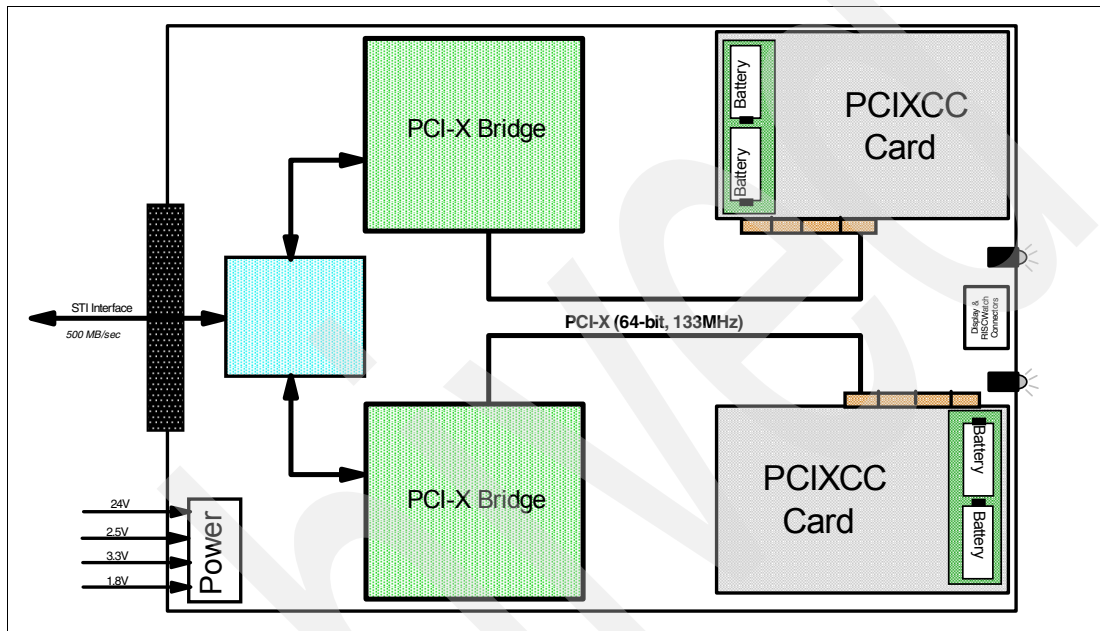


Figure 5-1 Crypto Express2 feature

The Crypto Express2 feature does not have ports and does not use fiber optic or other cables. It does not use CHPIDs, but requires one slot in the I/O cage and one PCHID for each PCI-X cryptographic adapter. The feature is attached to an STI and has no other external interfaces. Removal of the feature or card *zeroizes* the content.

The z9 EC supports a maximum of eight Crypto Express2 features, offering a combination of up to 16 coprocessor and accelerators. Access to the PCI-X cryptographic adapter is controlled through the setup in the image profiles on the SE.

**Note:** While PCI-X cryptographic adapters have no CHPID type and are not identified as external channels, all logical partitions in all CSSs can have access to the adapter (up to 16 logical partitions per adapter). To have access to the adapter requires setup in the image profile for the partition, the adapter must be in the candidate list. For details on setting up the image profile, refer to *IBM Systems z9-109 Configuration Setup SG24-7203*.

### 5.3.1 Crypto Express2 coprocessor

The Crypto Express2 coprocessor is a Peripheral Component Interconnect eXtended (PCI-X) cryptographic adapter configured as a coprocessor and provides a high-performance cryptographic environment with added functions.

The Crypto Express2 coprocessor provides asynchronous functions only.

The Crypto Express2 feature contains two PCI-X cryptographic adapters. The two adapters are actually two PCI-X CC cryptographic processors, and as such these processors provide equivalent plus additional functions as the PCIXCC feature on the z990 with doubled throughput.

PCI-X cryptographic adapters, when configured as coprocessors, are designed for FIPS 140-2 Level 4 compliance rating for secure cryptographic hardware modules. Unauthorized removal of the adapter or feature *zeroizes* its content.

The Crypto Express2 coprocessor enables the user to:

- ▶ Encrypt and decrypt data utilizing secret-key algorithms. Triple-length key DES and double-length key DES algorithms are supported.
- ▶ Generate, install, and distribute cryptographic keys securely using both public and secret key cryptographic methods.
- ▶ Generate, verify, and translate personal identification numbers (PINs).
- ▶ Ensure the integrity of data by using message authentication codes (MACs), hashing algorithms, and Rivest-Shamir-Adelman (RSA) public key algorithm (PKA) digital signatures.

The Crypto Express2 coprocessor also provides (natively) the functions described below for the Crypto Express2 accelerator, however, with a lower performance than the Crypto Express2 accelerator can provide.

Three methods of master key entry are provided by ICSF for the Crypto Express2 feature coprocessor:

1. A pass phrase initialization method that generates and enters all master keys that are necessary to fully enable the cryptographic system in a minimal number of steps.
2. A simplified master key entry procedure provided through a series of Clear Master Key Entry panels from a TSO terminal.
3. In enterprises that require enhanced key-entry security, a Trusted Key Entry (TKE) workstation is available as an optional feature.

The security-relevant portion of the cryptographic functions is performed inside the secure physical boundary of a tamper-resistant card. Master keys and other security-relevant information is also maintained inside this secure boundary.

A Crypto Express2 coprocessor operates with the Integrated Cryptographic Service Facility (ICSF) and IBM Resource Access Control Facility (RACF®), or equivalent software products, in a z/OS operating environment to provide data privacy, data integrity, cryptographic key installation and generation, electronic cryptographic key distribution, and personal identification number (PIN) processing.

PR/SM fully supports the Crypto Express2 feature coprocessor to establish a logically partitioned environment on which multiple logical partitions can use the cryptographic functions. A 128-bit data-protection master key, and one 192-bit Public Key Algorithm (PKA) master key, are provided for each of 16 cryptographic domains that a coprocessor can serve.

Using the dynamic add/delete of a logical partition name, a logical partition can be renamed. Its name can be changed from 'NAME1' to '\*' and then changed again from '\*' to 'NAME2'. The logical partition number and MIF ID are retained across the logical partition name change. The master keys in the Crypto Express2 feature coprocessor that were associated with the old logical partition 'NAME1' are retained. There is no explicit action taken against a cryptographic component for this dynamic change.

**Note:** Cryptographic coprocessors are not tied to logical partition numbers or MIF IDs. They are set up with PCI-X adapter numbers and domain indices that are defined in the partition image profile. The customer can assign them to the partition and change or clear them when needed.

### 5.3.2 Crypto Express2 accelerator

The Crypto Express2 accelerator is actually a coprocessor that is re-configured by the installation to only use a subset of the coprocessor functions at a higher speed. The re-configuration is done through the z9 EC Support Element.

Note that:

- ▶ Re-configuration is done at the PCI-X cryptographic adapter level, that is, a Crypto Express2 feature can host a coprocessor and an accelerator, two coprocessors, or two accelerators.
- ▶ Re-configuration works both ways, from coprocessor to accelerator and from accelerator to coprocessor. Master keys in the coprocessor domain can be optionally preserved when reconfigured to an accelerator.
- ▶ Re-configuration is disruptive to coprocessor and accelerator operations. The coprocessor or accelerator must be deactivated before engaging the re-configuration.
- ▶ FIPS 140-2 certification is not relevant to the accelerator since it operates with clear keys only.
- ▶ The function extension capability through UDX is not available to the accelerator.

The functions that remain available when configured as an accelerator are used for the acceleration of modular arithmetic operations, that is, the RSA cryptographic operations used with the SSL/TLS protocol:

- ▶ PKA Decrypt (CSNDPKD), with PKCS-1.2 formatting
- ▶ PKA Encrypt (CSNDPKE), with Zero-Pad formatting
- ▶ Digital Signature Verify

The RSA encryption and decryption functions support key lengths of 512-bit to 2048-bit, in the Modulus Exponent (ME) and Chinese Remainder Theorem (CRT) formats.

The maximum number of SSL transactions per second that can be supported on a z9 EC by any combination of CPACF, and accelerators is limited by the amount of cycles available to perform the software portion of the SSL/TLS transactions.

When both PCI-X cryptographic adapters are configured as accelerators on a z9 EC, the Crypto Express2 feature is designed to perform up to 6000 SSL handshakes per second. This represents approximately a 3x performance improvement compared to a z990 using a PCI Cryptographic Accelerator (PCI CA) feature with two accelerators per feature.

### 5.3.3 Configuration rules

All Crypto Express2 features may reside in one I/O cage. The maximum number of Crypto Express2 features per I/O cage is eight; the maximum number of PCI-X cryptographic adapters is 16.

Table 5-1 summarizes configuration information for Crypto Express2 on the z9 EC.

Table 5-1 Crypto Express2 feature

Minimum number of orderable features per server	2
Order Increment above two features	1
Maximum number of features per server	8
Number of PCI-X cryptographic adapters per feature (coprocessor or accelerator)	2
Maximum number of PCI-X adapters per z9 EC	16
Number of cryptographic domains per PCI-X adapter <sup>a</sup>	16

a. More than one partition, defined to the same CSS or to different CSSs, can use the same domain number when assigned to different PCI-X cryptographic adapters.

The minimum initial order of Crypto Express2 features on a z9 EC is two. After the initial order, additional Crypto Express2 can be ordered one feature at a time up to a maximum of eight.

The concept of *dedicated processor* does not apply to the PCI-X cryptographic adapter. Whether configured as coprocessor or accelerator, the PCI-X cryptographic adapter is made available to a logical partition as directed by the domain assignment and the candidate list in the logical partition image profile, regardless of the shared or dedicated status given to the CPs in the partition.

When installed in a non-concurrent way, Crypto Express2 features are assigned PCI-X cryptographic adapter numbers sequentially during the Power-on Reset following the installation. When a Crypto Express2 feature is installed concurrently, it is possible for the installation to select an out-of-sequence number from the unused range. When a Crypto Express2 feature is removed concurrently, the PCI-X adapter numbers are automatically freed.

The definition of domain indexes and PCI-X cryptographic adapter numbers in the Candidate list for each logical partition should be planned ahead to allow for nondisruptive changes.

- ▶ A change to a logical partition image profile to modify its domain indexes or Candidate list is disruptive to the partition. It requires a partition deactivation-activation to take effect.
- ▶ The same usage domain index may be defined more than once across multiple logical partitions. However, the PCI-X cryptographic adapter number coupled with the usage domain index specified must be unique across all active logical partitions.

The same PCI-X cryptographic adapter number and usage domain index combination may be defined for more than one logical partition. This may be used, for example, to define a configuration for backup situations. Note that only one of the logical partitions can be active at any one time.

The z9 EC allows for up to 60 logical partitions to be active concurrently. Each PCI-X adapter supports 16 domains, whether it is configured as a Crypto Express2 accelerator or a Crypto Express2 coprocessor. The server configuration must include at least two Crypto Express2 (four PCI-X adapters and 16 domains per PCI-X adapter) when all 60 logical partitions require concurrent access to cryptographic functions. More Crypto Express2 features may be needed to satisfy application performance or availability requirements.

For availability, assignment of multiple PCI-X adapters of the same type (Crypto Express2 accelerator or coprocessor) to one logical partition should be spread across multiple features.

The Crypto Express2 feature is supported on the z990 (as a PCI-X cryptographic coprocessor only) and is carried forward when upgraded to a z9 EC.

## 5.4 TKE workstation feature

The TKE workstation is an optional feature that offers key management functions. The TKE 5.0 workstation (with TKE 5.0 or later Licensed Internal Code) is required to support cryptographic key management on the z9 EC.

TKE support across current and older TKE versions is as follows:

- ▶ TKE 5.0 workstation can control cryptographic features on z9 EC, z9 BC, z990, and z890 servers.
- ▶ TKE 4.x workstations can control cryptographic features on z990 and z890 servers.

**Note:** The TKE 5.0 workstation only supports Ethernet adapters to connect to a Local Area Network.

A TKE workstation is part of a customized solution for using the Integrated Cryptographic Service Facility for z/OS program product (ICSF for z/OS) to manage cryptographic keys of a z9 EC that has Crypto Express2 features installed and is configured for using Data Encryption Standard (DES) and Public Key Algorithm (PKA) cryptographic keys.

The TKE workstation provides secure control for the Crypto Express2 coprocessors, including loading of master keys.

If one or more logical partitions are customized for using Crypto Express2 coprocessors, the TKE workstation can be used to manage DES master keys and PKA master keys for all cryptographic domains of each Crypto Express2 coprocessor feature assigned to logical partitions defined by the TKE workstation.

Each logical partition in the same system using a domain managed through a TKE workstation connection is either a TKE host or a TKE target. A logical partition with a TCP/IP connection to the TKE is referred to as TKE host; all other partitions are TKE targets.

The cryptographic controls as set for a logical partition through the z9 EC Support Element determine whether the workstation is a TKE host or TKE target.

### 5.4.1 Optional TKE Feature

Along with the TKE workstation and the TKE LIC, you can add an optional Smart Card Reader (Feature Code 0887). The reader supports the use of smart cards that contain an embedded microprocessor and associated memory for data storage that can contain the keys to be loaded into the Crypto Express2 feature. Access to and use of confidential data on the smart card is protected by a user-defined Personal Identification Number (PIN). Additional smart cards can be ordered for backup; the smart card feature code is 0888.

## 5.5 Cryptographic functions comparison

Table 5-2 summarizes the functions and attributes of the cryptographic hardware features.

Table 5-2 *Cryptographic functions on z9 EC*

Functions or attributes	CPACF	Crypto Express2 Coprocessor	Crypto Express2 Accelerator
Supports z/OS applications using ICSF	X	X	X
Encryption and decryption using secret-key algorithm		X	
Provides highest SSL handshake performance			X <sup>(a)</sup>
Provides highest symmetric (clear key) encryption performance	X		
Provides highest asymmetric (clear key) encryption performance			X
Provides highest asymmetric (encrypted key) encryption performance		X	
Disruptive process to enable		(b)	(b)
Requires IOCDs definition			
Uses CHPID numbers			
Uses PCHIDs		X <sup>(c)</sup>	X <sup>(c)</sup>
Physically embedded on each CP and IFL	X		
Requires CPACF DES/TDES enablement FC 3863	X <sup>(d)</sup>	X <sup>(d)</sup>	X <sup>(d)</sup>
Requires ICSF to be active		X	X
Offers user programming function (UDX)		X	
Usable for data privacy - encryption and decryption processing	X	X	
Usable for data integrity - hashing and message authentication	X	X	
Usable for financial processes and key management operations		X	
Crypto performance RMF™ monitoring		X	X
Requires system master keys to be loaded		X	
System (master) key storage		X	
Retained key storage		X	
Tamper-resistant hardware packaging		X	X <sup>e</sup>
Designed for FIPS 140-2 Level 4 certification		X	
Supports SSL functions	X	X	X

Functions or attributes	CPACF	Crypto Express2 Coprocessor	Crypto Express2 Accelerator
Supports Linux applications doing SSL handshakes			X
RSA functions		X	X
High performance SHA-1, and SHA-256	X		
Clear key DES/T-DES	X		
Advanced Encryption Standard (AES) for 128-bit keys	X		
Pseudo Random Number Generation (PRNG)	X	X	
Clear key RSA			X
Double length DUKPT support		X	
Europay Mastercard VISA (EMV) support		X	
Public Key Decrypt (PKD) support for Zero-Pad option for clear RSA private keys		X	X
Public Key Encrypt (PKE) support for MRP function		X	X
Remote Loading of Initial Keys in ATM		X	
Improved Key Exchange with non CCA System		X	
ISO 16609 CBC Mode T-DES MAC Support		X	

a. Requires CPACF DES/TDES enablement feature code 3863.

b. In order to make the addition of Crypto Express2 feature nondisruptive, the logical partition must be pre-defined with the appropriate PCI-X cryptographic adapter number selected in its candidate list in the partition image profile.

c. One PCHID is required per PCI-X cryptographic adapter.

d. Not required for Linux if only RSA clear key operations are used; DES or TDES encryption requires CPACF to be enabled.

e. Physically present but not used when configured as an accelerator (clear key only).

## 5.6 Software support

The minimum software support levels are listed in the following sections. The PSP buckets should be pulled and reviewed to ensure the latest support levels are known and included as part of the implementation plan.

### 5.6.1 CPACF

The minimum software requirement to support the CPACF on System z9 EC is:

- ▶ z/OS V1.6 or V1.7 with Enhancements for Cryptographic Support for z/OS and z/OS.e V1R6/R7
- ▶ z/VM Version 5.1
- ▶ Linux on System z, in a future distribution release or service update for Linux on System z

## 5.6.2 Crypto Express2

Table 5-3 lists the minimum software requirements for the Crypto Express2 feature when configured as a coprocessor or an accelerator and support for the base or enhanced functions.

Table 5-3 *Crypto Express2 support on z9 EC*

Operating system	Support
z/OS V1 R9	Included in base product.
z/OS V1 R8	Included in base product.
z/OS V1 R7	Web deliverable.
z/OS V1 R6	Web deliverable.
z/VM V5 R3	Any guest that can exploit the feature.
z/VM V5 R2	Any guest that can exploit the feature; PTFs.
z/VM V5 R1	Any guest that can exploit the feature; PTFs.
z/VSE V4 R1	Included in the base product.
z/VSE V3 R1	PTFs.
Linux on System z	Support delivered through IBM and distribution partners (for clear key RSA operations only).

## 5.6.3 Web deliverables

See the following sources for Web deliverables.

- ▶ For z/OS and z/OS.e Web deliverables, check the z/OS Web site at:  
<http://www.ibm.com/eserver/zseries/zos/downloads>
- ▶ For Linux on System z, support is delivered through IBM and distribution partners. For more information, check the DeveloperWorks Web site at:  
<http://www.ibm.com/developerworks/linux/linux390/>

## 5.6.4 z/OS ISCF FMIDs

Integrated Cryptographic Service Facility (ICSF) is a component of z/OS, and is designed to transparently use the available cryptographic functions, whether CPACF or Crypto Express2, to balance the workload and help address the bandwidth requirements of your applications.

FMID HCR7731 is available as a Web download for z/OS V1R6 and z/OS V1R7 and provides support for the PCI-X cryptographic coprocessor and accelerator functions as well as the CPACF AES, PRNG, SHA-256 support.



Table 5-4 lists the ICSF FMIDs and Web deliverables for z/OS V1.6 to V1.8.

Table 5-4 z/OS ICSF FMIDs

z/OS and z/OS.e	ICSF FMID <sup>a</sup>	Web deliverable name	supported function
V1.6	HCR770B	z990 and z890 Enhancements to Cryptographic Support	PCI-X Adapter Coprocessor only
	HCR7720	ICSF 64-bit Virtual Support for z/OS V1.6 and z/OS.e V1.6 <sup>b</sup>	PCI-X Adapter Coprocessor only
	HCR7730	Cryptographic support for z/OS V1.6/V1.7 and z/OS.e V1.6/V1.7	PCI-X Adapter Coprocessor and Accelerator CPACF Enhancements <sup>b</sup>
	HCR7731	Enhancements for Cryptographic support for z/OS and z/OS.e V1.6/V1.7	PCI-X Adapter Coprocessor and Accelerator CPACF Enhancements <sup>c</sup> Remote Key Loading ISO 16609 CBC Mode TDES MAC
V1.7	HCR7720	ICSF 64-bit Virtual Support for z/OS V1.6 and z/OS.e V1.6 (included in base)	PCI-X Adapter Coprocessor
	HCR7730	Cryptographic support for z/OS V1.6/V1.7 and z/OS.e V1.6/V1.7	PCI-X Adapter Coprocessor and Accelerator CPACF Enhancements <sup>c</sup>
	HCR7731	Enhancements for Cryptographic support for z/OS and z/OS.e V1.6/V1.7	PCI-X Adapter Coprocessor and Accelerator CPACF Enhancements <sup>c</sup> Remote Key Loading ISO 16609 CBC Mode TDES MAC
V1.8	HCR7731	Enhancements for Cryptographic support for z/OS and z/OS.e V1.6/V1.7 (included in base)	PCI-X Adapter Coprocessor and Accelerator CPACF Enhancements <sup>c</sup> Remote Key Loading and ISO 16609 CBC Mode TDES MAC

a. PTF information can be found in the PSP bucket '2094DEVICE'.

b. CPACF Enhancements include support for AES, PRNG and SHA-256

Archived



## Software support

This chapter lists the minimum operating system requirements and support considerations for the z9 EC and its features. It covers z/OS, z/VM, z/VSE, z/TPF, and Linux on System z. This information is subject to change; therefore, for the most current information, refer to the Preventive Service Planning (PSP) bucket for 2094DEVICE.

The System z9 EC functions that are supported will depend on the operating system version and release.

## 6.1 Operating systems summary

Table 6-1 summarizes the minimum operating system level required by the z9 EC.

Table 6-1 z9 EC minimum operating systems requirements

Operating systems	ESA/390 (31-bit Mode)	z/Architecture (64-bit mode)	Notes
z/OS V1R6 and later	No	Yes	
z/VM V5R1 and later	No	Yes	
Linux on System z	Yes <sup>a</sup>	Yes	Distributions of SUSE SLES 9 and later and Red Hat RHEL 4 and later
z/VSE V3R1	Yes	No	
TPF V4R1	Yes	No	
z/TPF V1R1	Yes	Yes	

a. Linux 64-bit distributions are supporting the IBM System z architecture. System z servers can also run code built for the 31-bit mainframe systems.

**Note:** Refer to the z/OS, z/VM, z/VSE, and z/TPF subsets of the 2094DEVICE Preventative Planning (PSP) bucket prior to installing an IBM System z9 EC.

Exploitation of some features depends on a particular operating system. In all cases, PTFs may be needed with the operating system level indicated. PSP buckets are continuously updated, and so should be reviewed regularly when planning for installation of a new server. They contain the latest information about maintenance.

Hardware and software buckets contain installation information, hardware and software service levels, service recommendations, and cross-product dependencies.

PSP buckets are organized by machine type. For a mainframe installation or upgrade, you may want to check:

- ▶ 2094DEVICE - For z9 EC

## 6.2 Support by operating system

In this section, we discuss support by operating system.

### 6.2.1 z/OS

Table 6-2 summarizes the z9 EC functions' support requirements for current z/OS releases.

Table 6-2 z/OS support summary

Function	z/OS V1R9	z/OS V1R8	z/OS V1R7	z/OS V1R6
z9 EC	Supported	Supported	Supported	Supported
60 logical partitions	Supported	Supported	Supported	Supported
LPAR Group Capacity Limit	Supported	Supported	Not supported	Not supported
Separate LPAR management of PUs	Supported	Supported	Supported	Supported
63.75 K Subchannels	Supported	Supported	Supported	Supported
Multiple Subchannel Sets	Supported	Supported	Supported	Not supported
MIDAW facility	Supported	Supported	Supported	Supported
Request Node Identification Data	Supported	Supported	Supported	Supported
FICON link incident reporting	Supported	Supported	Supported	Supported
HyperSockets support of IPV6	Supported	Supported	Supported	Supported
Crypto Express2 configured as accelerator or coprocessor	Supported	Supported	Web <sup>a</sup> deliverable	Web <sup>a</sup> deliverable
Hardware Decimal Floating Point	Supported	Supported	Supported	Supported
CPACF	Supported	Supported	Supported	Supported
CPACF AES, PRNG, and SHA-256	Supported	Supported	Web <sup>a</sup> deliverable	Web <sup>a</sup> deliverable
FICON Express2	Supported	Supported	Supported	Supported
FICON Express4	Supported	Supported	Supported	Supported
VLAN management	Supported	Supported	Supported	Not supported
OSA-Express2 Gigabit and 100baseT Ethernet CHPID type OSN	Supported	Supported	Supported	Supported
OSA-Express or OSA-Express2 1000baseT CHPID type OSC	Supported	Supported	Supported	Supported
OSA-Express or OSA-Express2 Gigabit and 1000baseT Ethernet CHPID type OSD	Supported	Supported	Supported	Supported

Function	z/OS V1R9	z/OS V1R8	z/OS V1R7	z/OS V1R6
OSA-Express or OSA-Express2 1000baseT CHPID type OSE	Supported	Supported	Supported	Supported
OSA-Express2 10 Gigabit CHPID type OSD	Supported	Supported	Supported	Supported
OSA Layer 3 Virtual MAC	Supported	Supported	Not supported	Not supported
OSA-Express2 QDIO Diagnostic Synchronization	Supported	Supported	Not supported	Not supported
OSA-Express2 Network Traffic Analyzer	Supported	Supported	Not supported	Not supported
OSA Dynamic LAN idle	Supported	Supported	Not supported	Not supported
OSA/SF enhancements for IP, MAC addressing (CHPID=OSD)	Supported	Supported	Supported	Supported
zIIP <sup>b</sup>	Supported	Supported	Supported	Supported
Server Time Protocol	Supported	Supported	Supported	Not supported <sup>c</sup>
System-initiated CHPID reconfiguration	Supported	Supported	Supported	Supported
Multipath IPL	Supported	Supported	Supported	Supported
CFCC level 15	Supported	Supported	Supported	Supported

a. Enhancements for Cryptographic Support for z/OS and z/OS.e V1R6/R7 Web deliverable. Support to include remote key loading, improved key exchange, and ISO 16609 CBC mode TDES for z/OS V1.6 and above. It replaces previous Web deliverable - Cryptographic Support for z/OS V1 R6/R7 and z/OS.e V1 R6/R7, which included support for CPACF enhancements and configuring of Crypto Express2 adapters.

b. System z9 Integrated Information Processors are designed to be exploited by DB2 V8 with enabling PTFs.

c. Although an STP-only CTN must consist of only z/OS V1.7 or later systems, a Mixed CTN can include z/OS V1.6, as long as the STP toleration PTFs are installed. For example, z/OS V1.8 can coexist with z/OS V1.6 in the same timing network in the same sysplex.

## 6.2.2 z/VM

Table 6-3 lists the z9 EC support requirements for z/VM.

Table 6-3 z9 EC support requirements for z/VM

Feature	Support on z9 EC requires at a minimum
Separate LPAR management of PUs	z/VM V5.1.
60 logical partitions	z/VM V5.1.
Hardware Decimal Floating Point	z/VM V5.2 (guest support).
CPACF	z/VM V5.1.
Enhancements to CPACF	z/VM V5.1.

<b>Feature</b>	<b>Support on z9 EC requires at a minimum</b>
Crypto Express2, compatibility support	z/VM V5.1 (guest support).
Crypto Express2, exploitation support when a PCI-X adapter is configured as an accelerator or a coprocessor	z/VM V5.1 (guest support).
Remote key loading for ATMs, ISO 16609 CBC mode TDES MAC	z/VM 5.1 (guest support).
63.75 K subchannels	z/VM V5.1.
MIDAW	z/VM V5.3 (guest support)
FICON Express2	z/VM V5.1.
FICON Express4	z/VM V5.1.
FICON Express2 and FICON Express4 CHPID type FCP support of SCSI disks	z/VM V5.1.
	z/VM V5.2 for performance assist for guests.
N_Port ID Virtualization for FICON CHPID type FCP	z/VM V5.1 and V5.2 provide for system usage of NPIV. <ul style="list-style-type: none"> <li>▶ z/VM V5.1 cannot be installed from DVD to SCSI disks when NPIV is enabled.</li> <li>▶ z/VM V5.2 support for guest operating systems and VM users to obtain virtual port numbers.</li> </ul>
FCP point-to-point attachments	z/VM V5.1 and V5.2 for guests. z/VM V5.3 for system and guest usage.
HiperSockets support of IPv6	z/VM V5.2.
VLAN management	z/VM V5.1. Support of guests is transparent to z/VM if the device is directly connected to the guest (pass through).
OSA Layer 3 Virtual MAC	z/VM V5.1 (guest support).
OSA-Express2 QDIO Diagnostic Synchronization	z/VM V5.1 (guest support).
OSA-Express2 Network Traffic Analyzer	z/VM V5.1 (guest support).
OSA Dynamic Lan Idle	z/VM V5.1 (guest support).
OSA-Express2 link aggregation support	z/VM V5.3.
OSA/SF enhancements IP and MAC addressing (CHPID type OSD)	z/VM 5.1.
z/VM integrated systems management	z/VM 5.3.
Program-directed re-IPL	z/VM 5.3 (guest support).
CFCC level 15	z/VM V5.1 (guest support).

Table 6-4 OSA-Express2 on z9 EC support requirements for z/VM

Feature	CHPID type OSD	CHPID type OSC	CHPID type OSE	CHPID type OSN
OSA-Express2 Gigabit Ethernet	z/VM V5.1			z/VM V5.1
OSA-Express2 1000BASE-T Ethernet	z/VM V5.1	z/VM V5.1	z/VM V5.1	z/VM V5.1
OSA-Express2 10 Gigabit Ethernet LR	z/VM V5.1			
OSA/SF enhancements for IP and MAC addressing	z/VM V5.1			

## 6.2.3 VSE/ESA and z/VSE

Table 6-5 lists z9 EC support requirements for VSE/ESA and z/VSE.

Table 6-5 z9 EC support requirements for VSE/ESA and z/VSE

Feature	Support on z9 EC requires at a minimum
Separate LPAR management of PUs	z/VSE V3.1
60 logical partitions	z/VSE V3.1
Crypto Express2 compatibility support	z/VSE V3.1
Crypto Express2, exploitation support when a PCI-X adapter is configured as an accelerator or a coprocessor.	z/VSE V3.1
FICON Express2	z/VSE V3.1
FICON Express4	z/VSE V3.1
FICON Express2 and FICON Express4 CHPID type FCP support of SCSI disks	z/VSE 3.1
N_Port ID Virtualization	z/VSE 3.1

Table 6-6 OSA-Express2 on z9 EC support requirements for VSE/ESA and z/VSE

Feature	CHPID type OSD	CHPID type OSC	CHPID type OSE	CHPID type OSN
OSA-Express2 Gigabit Ethernet	z/VSE V3.1			z/VSE 3.1
OSA-Express2 1000BASE-T Ethernet	z/VSE V3.1	z/VSE V3.1	z/VSE V3.1	z/VSE 3.1
OSA-Express2 10 Gigabit Ethernet LR	z/VSE V3.1			



## 6.2.4 Linux on System z

Table 6-7 lists z9 EC support requirements for Linux on System z.

Table 6-7 z9 EC support requirements for Linux on System z

Feature	Support on z9 EC requires at a minimum
Separate LPAR management of PUs	Distributions of SUSE SLES 9 and Red Hat RHEL 4
60 logical partitions	Distributions of SUSE SLES 9 and Red Hat RHEL 4
Program-directed re-IPL	SUSE SLES 9 SP3 <sup>a</sup>
CPACF	Distributions of SUSE SLES 9 and Red Hat RHEL 4
Enhancements to CPACF	SUSE SLES 9 SP3 <sup>b</sup> and RHEL 4 U3 <sup>b</sup>
Crypto Express2 compatibility support	Distribution of SUSE SLES 9
Crypto Express2 exploitation support when a PCI-X adapter is configured as an accelerator or a coprocessor	Distribution of SUSE SLES 9
Performance assists for z/VM guests	Note <sup>a</sup>
Multiple Subchannel Sets	Note <sup>a</sup>
63.75K subchannels for all channel types	Distributions of SUSE SLES 9 and Red Hat RHEL 4
FICON Express2	Distributions of SUSE SLES 9 and Red Hat RHEL 4
FICON Express4	Distributions of SUSE SLES 9 and Red Hat RHEL 4
FICON Express2 CHPID type FCP, support of SCSI disks	Distributions of SUSE SLES 9 and Red Hat RHEL 4
FICON Express4 CHPID type FCP, support of SCSI disks	Distributions of SUSE SLES 9 and Red Hat RHEL 4
Performance metrics for Linux on System z	Note <sup>a</sup>
N_Port ID Virtualization	SUSE SLES 9 SP3 <sup>a</sup>
FCP point-to-point attachments for FICON CHPID type FCP	Note <sup>a</sup>
System-initiated CHPID reconfiguration	Note <sup>a</sup>

a. IBM is working with its Linux distribution partners so that this function will be provided in future Linux on System z distribution releases or service updates.

b. IBM is working with its Linux distribution partners on kernel space exploitation.

Table 6-8 OSA-Express2 on z9 EC support requirements for Linux on System z

Feature	CHPID type OSD	CHPID type OSN
OSA-Express2 Gigabit Ethernet	Distributions of SUSE SLES 9 and Red Hat RHEL 4	SUSE SLES 9 SP3 and Red Hat RHEL 4 U3
OSA-Express2 1000BASE-T Ethernet	Distributions of SUSE SLES 9 and Red Hat RHEL 4	SUSE SLES 9 SP3 and Red Hat RHEL 4 U3
OSA-Express2 10 Gigabit Ethernet LR	Distributions of SUSE SLES 9 and Red Hat RHEL 4	

## 6.2.5 TPF and z/TPF

Table 6-9 lists z9 EC support requirements for TPF on System z9.

Table 6-9 z9 EC support requirements for TPF on System z

Feature	Support on z9 EC requires at a minimum
Separate LPAR management of PUs	TPF V4.1 and z/TPF V1.1
60 logical partitions	TPF V4.1 and z/TPF V1.1
FICON Express2	TPF V4.1 at PUT 16 and z/TPF V1.1
FICON Express4	TPF V4.1 at PUT 16 and z/TPF V1.1

Table 6-10 OSA-Express2 on z9 EC support requirements for TPF

Feature	CHPID type OSD	CHPID type OSN
OSA-Express2 Gigabit Ethernet	TPF V4.1 z/TPF V1.1.	TPF V4.1 z/TPF V1.1
OSA-Express2 1000BASE-T Ethernet	TPF V4.1 z/TPF V1.1.	TPF V4.1 z/TPF V1.1
OSA-Express2 10 Gigabit Ethernet LR	TPF V4.1 z/TPF V1.1	

## 6.3 Support by function

In this section, we discuss support by function.

### 60 logical partitions

This feature in the System z9 EC allows the system to be configured with up to 60 logical partitions. Since the limitation is 15 logical partitions per Channel Subsystem, it is necessary to configure four Channel Subsystems to reach 60 logical partitions.

Table 6-11 Minimum support requirements for 60 logical partitions

Operating system	Support requirements
z/OS V1R6 and above	Supported
z/VM V5 R1	Supported
z/VSE V4 R1	Supported
z/VSE V3 R1	Supported
z/TPF V1 R1	Supported
TPF V4 R1	Supported
Linux for System z	Distributions of SUSE SLES 9 and Red Hat RHEL 4

## Single system image

A single system image can control several processors (CPs, zIIPs, zAAPs, or IFLs, as appropriate). Table 6-12 shows the maximum number of processors supported for each operating system image.

Table 6-12 Single system image software support

Operating system	Maximum number of CPs+zIIPs+zAAPs <sup>a</sup> or IFLs per system image
z/OS V1R6 and above	32
z/VM V5 R3	32
z/VM V5 R1, R2	24
z/TPF V1 R1	54
Linux on System z	SUSE SLES 9 and Red Hat RHEL 3, Up to 32

a. Total CPs, zIIPs, and zAAPs refers to the sum of these PU characterizations.

## Separate LPAR management of PUs

The System z9 uses separate PU pools for each optional PU type. The separate management of PU types enhances and simplifies capacity planning and management of the configured logical partitions and their associated processor resources.

Table 6-13 Minimum support requirements for separate LPAR management of PUs

Operating system	Support requirements
z/OS V1R6 and above	Supported
z/VM V5R1 and above	Supported
z/VSE V3 R1 and above	Supported
z/TPF V1 R1	Supported
TPF V4 R1	Supported
Linux on System z	Distributions of SUSE SLES 9 and Red Hat RHEL 4

## 63.75 K subchannels

Servers prior to the z9 EC reserved 1024 subchannels for internal system use out of the potential maximum of 64 K subchannels. The z9 EC has reduced the reserved number to 256 subchannels, thus increasing the number of subchannels available. Reserved subchannels exist only in subchannel set 0; no subchannels are reserved in subchannel set 1.

The informal name 63.75 K represents  $63 \times 1024 + 0.75 \times 1024 = 65280$  subchannels.

Table 6-14 Minimum support requirements for 63.75 K subchannels

Operating system	Support requirements
z/OS V1R6 and above	Supported
z/VM V5R1	Supported
Linux on System z	Distributions of SUSE SLES 9 and Red Hat RHEL 4

## Multiple Subchannel Sets (MSS)

Multiple subchannel sets in z9 EC provide a mechanism for addressing more than 63.75 K I/O devices and aliases for ESCON (CHPID type CNC) and FICON (CHPID types FCV and FC) on the z9 EC.

Multiple subchannel sets are not supported for z/OS running as a guest under z/VM.

Table 6-15 lists the minimum operating systems level required on the z9 EC.

Table 6-15 Minimum software requirement for MSS

Operating system	Support requirements
z/OS V1R7 and above	Included.
Linux on System z	IBM is working with its Linux distribution partners so that this function will be provided in future Linux on System z distribution releases or service updates.

Exploitation of Multiple Subchannel Sets is *not* supported in z/OS V1R6. However, creation of a z9 EC IOCDS with Multiple Subchannel Sets defined is possible with z/OS V1R6 if the small program enhancements (SPEs) for HCD and IOCP are installed.<sup>1</sup>

## MIDAW facility

The MIDAW facility in z9 EC provides a more efficient replacement for data chained CCWs.

The MIDAW facility is not supported when running as a z/OS guest under z/VM prior to z/VM 5.3.

Table 6-16 lists the minimum support requirements for MIDAW.

Table 6-16 Minimum support requirements for MIDAW

Operating system	Support requirements
z/OS V1R6 and above	Supported
z/VM 5.3	Supported

## Request Node Identification Data (RNID)

Request Node Identification Data (RNID) for native FICON CHPID type FC allows isolation of cabling-detected errors on the z9 EC.

Table 6-17 lists the minimum support requirements for RNID.

Table 6-17 Minimum support requirements for RNID

Operating system	Support requirements
z/OS V1R6 and above	Supported

## N\_Port ID virtualization

N\_Port ID virtualization provides a way to allow multiple system images (in logical partitions or z/VM guests) to use a single FCP channel as though each were the sole user of the channel. Note that this feature may be used with earlier FICON features that have been carried forward from earlier servers.

<sup>1</sup> See OA08197 and OA07875 for IOCP and HCD SPEs.

Table 6-18 lists the minimum support requirements for NPIV.

Table 6-18 Minimum support requirements for NPIV

Operating system	Support requirements
z/VM V5R1	Provide for system usage of NPIV. z/VM V5.1 cannot be installed from DVD to SCSI disks when NPIV is enabled.
z/VM V5R2	Provide for system usage of NPIV. z/VM V5.2 support for guest operating systems and VM users to obtain virtual port numbers. z/VM 5.2 support for installing from DVD to SCSI disks when NPIV is enabled.
z/VSE V4 R1	Supported.
z/VSE V3 R1	Supported.
Linux on System z	SUSE SLES 9 SP3 <sup>a</sup>

a. IBM is working with its distribution partners to provide this function in future distribution releases or service updates.

### FICON link incident reporting

FICON link incident reporting is designed to allow an operating system image (without operator intervention) to register for link incident reports.

Table 6-19 Minimum support requirements for link incident reporting

Operating system	Support requirements
z/OS V1R7 and above	Supported

### Program directed re-IPL

Program directed re-IPL is designed to allow an operating system on a z9 EC to re-IPL without operator intervention. This function is supported for both SCSI and ECKD™ devices.

Table 6-20 Minimum support requirements for Program directed re-IPL

Operating system	Support requirements
Linux on System z	SUSE SLES 9 SP3 <sup>a</sup>

a. IBM is working with its distribution partners to provide this function in future distribution releases or service updates.

For more information about Linux on System z, see the developerWorks® Web site at:

<http://www.ibm.com/developerworks/linux/linux390/>

### OSA-Express2 1000BASE-T Ethernet

This adapter can be configured in:

- ▶ QDIO mode, with CHPID type OSD or OSN
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode with CHPID type OSC

Table 6-21 shows the minimum support requirements for OSA-Express2 1000BASE-T.

Table 6-21 Minimum support requirements for OSA-Express2 1000BASE-T

Operating system	CHPID type OSC	CHPID type OSD	CHPID type OSE
z/OS V1R6 and above	Supported	Supported	Supported
z/VM V5 R1 and above	Supported	Supported	Supported
z/VSE V3R1 and above	Supported	Supported	Supported
z/TPF V1R1	Not supported	Supported	Not supported
TFP V4R1	Not supported	PUT 13 plus PTFs	Not supported
Linux on System z	Not supported	SUSE SLES 9 and above Red Hat RHEL 4 and above	Not supported

### OSA-Express2 GARP VLAN Registration Protocol (GVRP)

GVRP support allows an OSA-Express2 port to register or de-register its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change.

Table 6-22 Minimum support requirements for GVRP

Operating system	Support requirements
z/OS V1R7 and above	Supported
z/VM V5R1 and above	Supported
Linux on System z	Support will be delivered through IBM and distribution partners

### OSA-Express2 OSN support

Channel Data Link Control (CDLC), when used with the Communication Controller for Linux, emulates selected functions of IBM 3745/NCP operations. The port used with the OSN support appears as an ESCON channel to the operating system. This support may be used with any OSA-Express2 feature, except for the 10 GbE LR feature.

Table 6-23 shows the minimum support requirements for OSA-Express2 OSN.

Table 6-23 Minimum support requirements for OSA-Express2 OSN

Operating system	OSA-Express2 OSN
z/OS V1R6 and above	Supported
z/VM V5 R1 and above	Supported
z/VSE V3 R1 and above	Supported
z/TPF V1 R1	Supported
TPF V4 R1	At PUT 13 plus PTFs
Linux on System z	Distributions of SUSE SLES 9 and Red Hat RHEL 4

## OSA-Express2 10 Gigabit Ethernet LR

Table 6-24 lists the minimum support requirements for OSA-Express2 10 Gigabit (CHPID type OSD).

Table 6-24 Minimum support requirements for OSA-Express2 10 Gigabit (CHPID type OSD)

Operating system	Support requirements
z/OS V1R6 and above	Supported
z/VSE V3.1 and above	
TPF 4.1	At PUT 13 with PTFs
z/TPF 1.1	
Linux on System z	Distributions of SUSE SLES 9 and Red Hat RHEL 4 (including Checksum Offload support) and above

## HiperSockets IPv6

IPv6 is expected to be a key element in future networking. The IPv6 support for HiperSockets permits compatible implementations between external networks and internal HiperSocket networks.

Table 6-25 lists the minimum support requirements for HiperSockets IPv6 (CHPID type IQD).

Table 6-25 Minimum support requirements for HiperSockets IPv6 (CHPID type IQD)

Operating system	Support requirements
z/OS V1R7 and above	Supported
z/VM V5 R2 and above	Supported

## VLAN management enhancements

Table 6-26 lists VLAN management enhancements for the OSA-Express2 and OSA-Express features (CHPID type OSD).

Table 6-26 Enhanced performance assists for VLAN management enhancements

Operating system	Support requirements
z/OS V1R7 and above	Supported.
z/VM V5R1 and above	Support of guests is transparent to z/VM if the device is directly connected to the guest (pass through).

## Cryptography

Detailed software support information for cryptography functions is provided in 5.6, "Software support" on page 161.

## CFCC

Generally, when you change Coupling Facility Code (CFCC) levels, the Coupling Facility structure sizes may change. If you run with a higher CFCC level on a Coupling Facility on your z9 EC, you may have larger structure size than you did previously. If your CFCC levels are identical, then there are no expected changes in structure sizes when moving from a previous server to a z9 EC.

If you are moving your Coupling Facilities to a z9 EC, and the CFCC levels are different than what they previously were, run the CFSIZER tool, as it may be necessary to increase CF structure sizes. Make the necessary changes as indicated by the CFSIZER tool.

See the IBM mainframe Web site at:

<http://www.ibm.com/servers/eserver/zseries/cfsizer/>

### 6.3.1 ICKDSF

ICKDSF Release 17 is required on all systems that share disk subsystems with a z9 EC processor.

ICKDSF supports a modified format of the CPU information field, which contains a two-digit logical partition identifier. ICKDSF uses the CPU information field instead of CCW reserve/release for concurrent media maintenance. It prevents multiple systems from running ICKDSF on the same volume, and at the same time allows user applications to run while ICKDSF is processing. In order to prevent any possible data corruption, ICKDSF must be able to determine all sharing systems that may potentially run ICKDSF; therefore, this support is required for z9 EC.

**Important:** The need for ICKDSF Release 17 applies even to systems that are not part of the same sysplex, or that are running a non-z/OS operating system, such as z/VM.

## 6.4 Software licensing considerations

The IBM System z9 mainframe software portfolio includes operating system software (that is, z/OS, z/TPF, z/VM, z/VSE, and VSE/ESA) and middleware that runs on these operating systems.

In this section, we discuss some of the current usage-based pricing options that you may want to take advantage of.

### 6.4.1 Workload License Charges

Workload License Charges (WLC) is a software license charge method. It has been enhanced with the Select Application License Charges (SALC); see 6.4.2, “Select Application License Charges (SALC)” on page 179.

WLC requires z/OS operating systems in 64-bit mode. Any mix of z/OS, z/VM, Linux, VM/ESA, VSE/ESA, and TPF images is allowed.

There are two WLC license types:

- ▶ Flat WLC (FWLC): Software products licensed under FWLC are charged at the same flat rate, no matter what capacity (MSUs) the server is.
- ▶ Variable WLC (VWLC): VWLC software products can be charged in two different ways:
  - Full-capacity: The server’s total number of MSUs is used for charging. Full-capacity is applicable when the server is not eligible for Sub-capacity.
  - Sub-capacity: Software charges are based on the logical partition’s utilization where the product is running.



WLC Sub-capacity allows software charges based on logical partition utilizations instead of the server's total number of MSUs. Sub-capacity removes the dependency between software charges and server (hardware) installed capacity.

Sub-capacity is based on the logical partition's rolling 4-hour average utilization. It is *not* based on the utilization of each product,<sup>2</sup> but on the utilization of the logical partition or partitions where it runs. The VWLC licensed products running on a logical partition will be charged by the maximum value of this partition's rolling 4-hour average utilization within a month.

The logical partition's rolling 4-hour average utilization can be limited by a *Defined Capacity* definition on the partition's image profiles. This activates the *Soft Capping* function of PR/SM, avoiding 4-hour average partition utilizations above the defined capacity value. Soft capping controls the maximum rolling 4-hour average utilization (the "last" 4-hour average value at every five minutes interval), but does *not* control the maximum "instantaneous" partition utilization.

Even using the soft capping option, the partition's utilization can reach up to its maximum share based on the number of logical processors and weights in the image profile. Only the rolling 4-hour average utilization is tracked, allowing utilization peaks above the defined capacity value.

As with the Parallel Sysplex License Charges (PSLC) software license charge type, the aggregation of servers' capacities within the same Parallel Sysplex is also possible in WLC, following the same prerequisites.

For further information about WLC and details on how to combine logical partitions utilization, see *z/OS Planning for Workload License Charges*, SA22-7506.

## 6.4.2 Select Application License Charges (SALC)

Select Application License Charges (SALC) applies to WebSphere MQ for System z only. It is designed to allow a WLC customer to licence MQ under product utilization rather than the sub-capacity pricing provided under WLC.

WebSphere MQ is typically a low usage product that runs pervasively throughout the customer environment. Clients who run WebSphere MQ at a very low usage may benefit from SALC. Alternatively, you can still choose to license WebSphere MQ under WLC.

A reporting function, which IBM provides in the operating system IBM Software Usage Report Program, is used to calculate the daily MSU number. The rules to determine the billable SALC MSUs for WebSphere MQ use the following algorithm:

- ▶ Determine the highest daily usage of a program<sup>3</sup> family, which is the highest of 24 hourly measurements recorded each day.
- ▶ Determine the monthly usage of a program<sup>1</sup> family, which is the fourth highest daily measurement recorded for a month.
- ▶ Use the highest monthly usage determined for the next billing period.

For additional information about SALC, see the announcement information at:

[http://www.ibm.com/common/ssi/rep\\_ca/3/897/ENUS205-183/](http://www.ibm.com/common/ssi/rep_ca/3/897/ENUS205-183/)

<sup>2</sup> With the exception of SALC products.

<sup>3</sup> *Program* refers to all active versions of MQ.

## 6.5 Concurrent upgrade considerations

Using Capacity Upgrade on Demand (CUoD), On/Off Capacity on Demand (On/Off CoD), Customer Initiated Upgrade (CIU), or Capacity BackUp (CBU), you can concurrently upgrade the z9 EC from one model to another, either temporarily or permanently.

Enabling and using the additional processor capacity should be transparent to most applications. However, there may be some programs that depend on processor model-related information. You need to consider the effect on the software running on a z9 EC when performing any of these configuration upgrades.

### Processor identification

There are two instructions used to obtain processor information: Store System Information instruction (STSI) and Store CPU ID instruction (STIDP).

STSI reports the processor model and model capacity identifier. It fully supports the concurrent upgrade functions and is the preferred way to request processor information.

STIDP is provided for purposes of backward compatibility.

### Store CPU ID instruction

The STIDP instruction provides information about the processor type, serial number, and logical partition identifier, as shown in Table 6-27. The logical partition identifier field is a full byte to support greater than 15 logical partitions.

The STIDP instruction also provides a 1-byte hexadecimal version code, which is always x'00' for a System z9 server.

Table 6-27 STIDP output for z9 EC

	Version code	CPU identification number		Machine type number	logical partition 2-digit indicator
Bit position	0-7	8-15	16-31	32-48	48-63
Value	x'00' <sup>a</sup>	logical partition ID <sup>b</sup>	6-digit number derived from the CPC serial number	x'2094'	x'8000' <sup>c</sup>

a. Version code is zero for System z9 processors.

b. The logical partition identifier is a two-digit number in the range from '00' to '3F'. It is assigned by the user on the image profile through the Support Element or HMC.

c. High order bit on indicates the logical partition ID value returned in bits 8-15 is a two-digit value.

When issued from an operating system running as a guest under z/VM, the result depends on whether the SET CPUID command has been used.

- ▶ Without the use of the **set CPUID** command, bits 0–7 are set to 'FF' by z/VM, but the remaining bits are unchanged, meaning they are exactly as they would have been without running as a z/VM guest.
- ▶ If the **set CPUID** command has been issued, bits 0–7 are set to 'FF' by z/VM and bits 8–31 are set to the value entered in the **set CPUID** command. Bits 32–63 are the same as they would have been without running as a z/VM guest.

Table 6-28 shows the possible output returned to the issuing program for an operating system running as a guest under z/VM.

Table 6-28 STIDP output for z9 EC, VM guest

	Version code	CPU identification number		Machine type number	logical partition 2-digit indicator
Bit position	0–7	8–15	16–31	32–48	48–63
Without set CPUID command	x'FF'	logical partition ID	4-digit number derived from the CPC serial number	x'2094'	x'8000'
With set CPUID command	x'FF'	6-digit number as entered by the command SET CPUID = nnnnnn		x'2094'	x'8000'

### STSI Store System Information instruction

The STSI instruction returns the processor model and model capacity identifier, in two 16-byte character fields. It also returns the same processor type that is returned by the STIDP instruction and the full serial number information.

The STSI instruction always returns the latest processor information, including the model capacity identifier after a concurrent upgrade has occurred. This is key to the functioning of CUoD, On/Off CoD, CIU, and CBU.

### Channel-to-channel links

There is extra planning required with CTCs if used with CUoD. After a concurrent upgrade, the channel CPC Node-Descriptor (NED)<sup>4</sup> information is not updated until the next Power On Reset (POR). The scenarios that we have seen in the field are:

- ▶ Customer does a concurrent upgrade of their server.
- ▶ Months later they do a POR.
- ▶ After the POR CTC devices became boxed.

The reason the CTC devices become boxed is that the NED information (machine model) has changed. With the z9 EC, the model number will only change in the above scenario with a concurrent book upgrade, for example, a S18 to S28 upgrade (see Figure 6-1 on page 182).

Currently with the z9 EC, z9 BC, z990, or z890, the level of exposure will depend on the APARs applied to the z/OS systems. Please check the following:

- ▶ OW53688: With this APAR, the model number for the CTC device is not included in its node descriptor.
- ▶ 0A099001: FICON spanned channel FCTC support.
- ▶ 0A110113: FICON CTC support for CPU upgrades from PRE z990 to z9 EC.

The alternative to the above APARs is to be prepared for the boxed CTCs to occur during the next POR of the upgraded system. In most cases, using the UNCOND option of the VARY ONLINE command will un-box the CTCs in a nondisruptive manner.

The implications of boxed CTCs should be investigated during the planning process prior to a concurrent upgrade.

<sup>4</sup> NED information, which includes serial number, machine type, and model, is exchanged between systems on the CTC link.

```
IEE174I 11.25.26 DISPLAY M 520
PROCESSOR STATUS
ID  CPU                      SERIAL
00  +                        21991E2094
01  +                        21991E2094
02  -
03  -

CPC ND = 002094.S18.IBM.02.000000002991E
CPC SI = 2094.712.IBM.02.0000000000002991E
CPC ID = 00
CPC NAME = SCZP101
LP NAME = A21          LP ID = 21
CSS ID = 2
MIF ID = 1
```

Figure 6-1 Node descriptor information

## 6.6 References

For the most current planning information, check the Support Web page for each operating system:

- ▶ z/OS  
<http://www.ibm.com/systems/support/z/zos/>
- ▶ z/VM  
<http://www.ibm.com/systems/support/z/zvm/>
- ▶ z/TPF  
<http://www.ibm.com/software/http/tpf/pages/maint.htm>
- ▶ z/VSE  
<http://www.ibm.com/servers/eserver/zseries/zvse/support/preventive.html>
- ▶ Linux on System z  
<http://www.ibm.com/systems/z/os/linux/>



## Sysplex functions

This chapter describes the capabilities of the z9 EC to support coupling functions, including Parallel Sysplex, Geographically Dispersed Parallel Sysplex™ (GDPS), and Intelligent Resource Director.

The following topics are included:

- ▶ 7.1, “Parallel Sysplex” on page 184
- ▶ 7.2, “Coupling Facility considerations” on page 185
- ▶ 7.3, “System-managed CF structure duplexing” on page 195
- ▶ 7.4, “Intelligent Resource Director” on page 197

## 7.1 Parallel Sysplex

Figure 7-1 illustrates the components of a Parallel Sysplex as implemented within the System z architecture. The figure is intended only as an example. It shows one of many possible Parallel Sysplex configurations; many other possibilities exist.

Shown is a z990 ICF (CF01) connected to two z9 ECs running in Sysplex. There is a second Integrated Coupling Facility (CF02) defined within one of the z9 ECs, containing Sysplex logical partitions running z/OS.

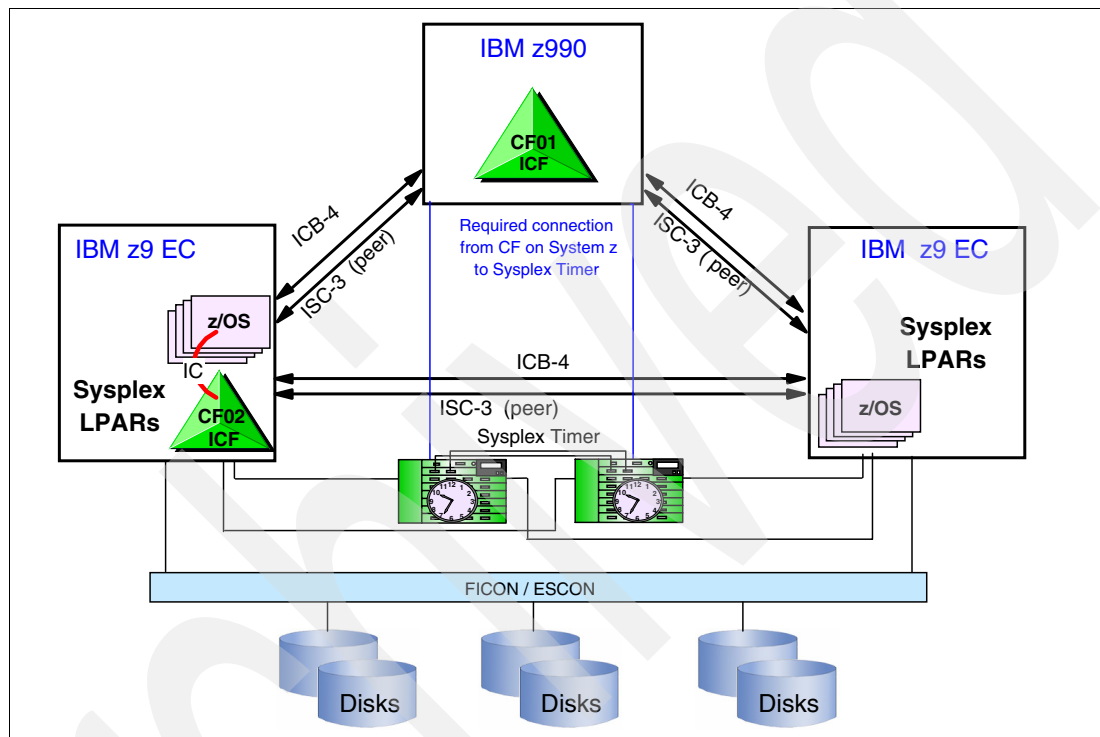


Figure 7-1 Sysplex hardware overview

Also shown is the required connection between the Coupling Facility (CF01) defined on a z990 or any z9 EC, and the Sysplex Timer, to support Message Time Ordering.

Parallel Sysplex technology is an enabling technology, allowing highly reliable, redundant, and robust System z technology to achieve near-continuous availability. A Parallel Sysplex comprises one or more z/OS operating system images coupled through one or more Coupling Facilities. The images can be combined together to form clusters. A properly configured Parallel Sysplex cluster is designed to maximize availability.

- ▶ **Continuous (application) availability:** The Parallel Sysplex cluster environment is composed of multiple images that provide concurrent access to all critical applications and data. You can introduce changes (such as software upgrades) one image at a time, while remaining images continue to process work. Note that in order to fully realize the benefits of continuous application availability, some application deployment and functional changes may be required (see *Parallel Sysplex Application Considerations*, SG24-6523).
- ▶ **High capacity:** The Parallel Sysplex environment can scale, in a nearly linear fashion, from two to 32 images. The aggregated capacity of this configuration meets every processing requirement known today.

- ▶ **Dynamic workload balancing:** The entire Parallel Sysplex cluster can be viewed as a single logical resource to users and business applications. Work can be directed to any like operating system image in a Parallel Sysplex cluster having available capacity. Workload management permits you to run diverse applications across a Parallel Sysplex cluster while maintaining the response levels critical to your business.
- ▶ **Systems management:** The Parallel Sysplex architecture provides the infrastructure to satisfy a customer requirement for continuous availability, while providing techniques for achieving simplified systems management consistent with this requirement.
- ▶ **Resource sharing:** A number of base z/OS components exploit Coupling Facility shared storage, providing an excellent medium for sharing component information for the purpose of multi-image resource management. This exploitation enables sharing of physical resources with significant improvements in cost, performance, and simplified systems management.
- ▶ **Single system image:** The collection of system images in the Parallel Sysplex appears as a single entity to the operator, the user, the database administrator, and so on. A single system image ensures reduced complexity from both operational and definition perspectives.

Through this state-of-the-art cluster technology, the power of multiple z/OS images can be harnessed to work in concert on common workloads. The System z Parallel Sysplex cluster takes the commercial strengths of the z/OS platform to improved levels of system management, competitive price/performance, scalable growth, and continuous availability.

## 7.2 Coupling Facility considerations

Described here are the supported Parallel Sysplex configurations, required setup information when connected to a Sysplex Timer, different forms of Coupling Facilities (CFs) supported on the z9 EC, CFRM policy considerations, and ICF processor assignments. The z9 EC models support both Central Processors (CPs) and Internal Coupling Facility (ICF) processors.

The z9 EC family of servers does not have a special model for a CF-only processor. However, a z9 EC can be configured as a stand-alone Coupling Facility with up to a maximum of 16 PUs defined as ICFs.

### 7.2.1 Sysplex configurations and Time Synchronization

Parallel Sysplex configurations can have system images and Coupling Facilities located across multiple servers. However, the z9 EC, like the z990, has some additional configuration considerations.

#### Message Time Ordering Facility

As server and coupling link technologies have improved over the years, the synchronization tolerance between operating systems in a Parallel Sysplex has become more rigorous. In order to ensure that any exchange of time stamped information between operating systems in a sysplex involving the Coupling Facility observe the correct time ordering, time stamps are included in the message-transfer protocol between the server operating systems and the Coupling Facility. This is known as Message Time Ordering.

Message Time Ordering requires a connection between the server and the Sysplex Timer whenever a Coupling Facility logical partition is located on a z9 EC, z9 BC, z990 or z890.

In a Mixed or STP-only CTN in a Parallel Sysplex configuration, all servers must support Message Time Ordering.

Even though multiple servers can connect to only one Sysplex Timer unit, the typical configuration is usually connected to two different Sysplex Timer units in an Expanded Availability configuration. Refer to *IBM System z Connectivity Handbook*, SG24-5444, for IBM 9037 Sysplex Timer connectivity information.

### External Time Reference ID

A Sysplex Timer unit is assigned a unique two-digit ID at installation time referenced as an External Time Reference ID (ETR ID) in the output of the z/OS command D ETR and Support Element panels.

A function was implemented in the server's Support Element code, which requires the ETR Network ID of the attached Sysplex Timer Network to be manually set in the Support Element at installation time. This function checks that the ETR Network ID being received in the timing signals through each of the server's two ETR ports matches the ETR Network ID manually set in the server's Support Element (SE).

Up to two Sysplex Timer units can be configured in an Expanded Availability configuration, each one with a unique ETR ID. When in Expanded Availability configuration, a network ID (Net ID) is also assigned at installation time to identify that these two Sysplex Timer units belong to the same Sysplex Timer configuration.

The ETR Network ID can be set by using the System (Sysplex) Timer task, located on the CPC configuration task list. When this task is invoked, the System (Sysplex) Time window is displayed. Figure 7-2 shows an example for a server with both ETR ports and STP feature installed. The ETR Configuration window, shown in Figure 7-3 on page 187, contains the Sysplex Timer configuration information; the ETR Network ID (0–31) is entered on this configuration window.

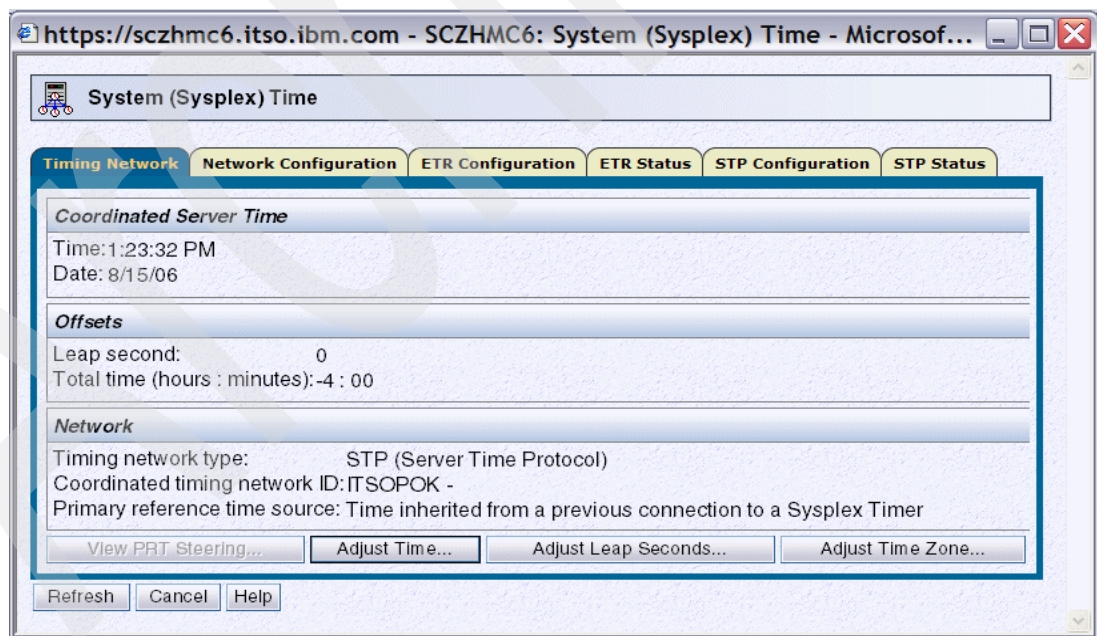


Figure 7-2 HMC workplace - System (Sysplex) Time, Timing network

The network ID configured on the z9 EC must match the actual Sysplex Timer network ID to which the server is connected. If the network ID entered on the window does not match the network ID that was assigned to the Sysplex Timer, the timer port enters a semi-operational



state. In this state, the port is disabled from stepping, but still receives configuration data from the attached Sysplex Timer.

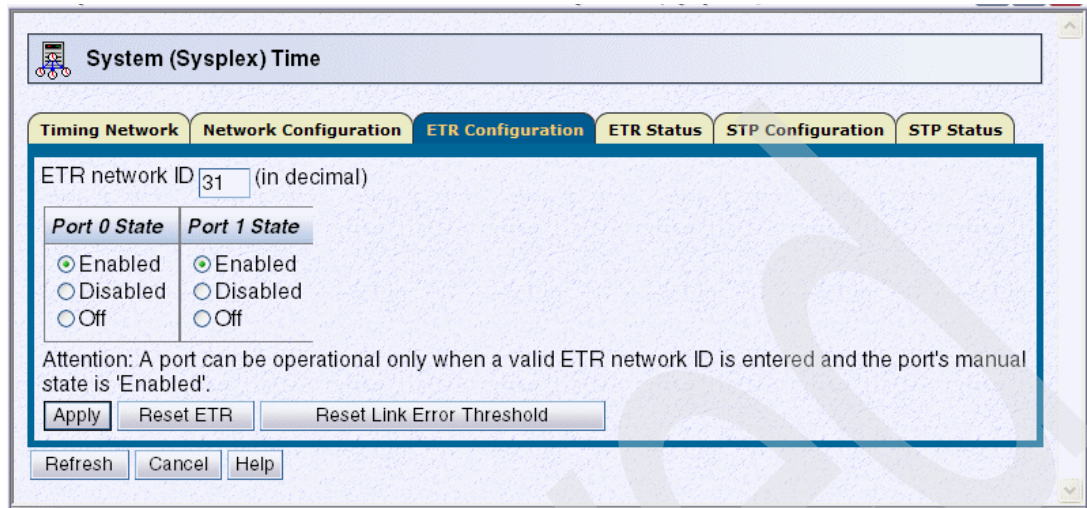


Figure 7-3 HMC workplace - System (Sysplex) Time, ETR configuration

After IPL, the configuration can be identified to any attached z/OS image by issuing the Display ETR command. The command output identifies the ETR configuration, including ETR ID and ETR Network ID; see Example 7-1.

Example 7-1 Display ETR in a Mixed Coordinated Timing Network in ETR timing mode

```

-D ETR
IEA282I 09.30.52 TIMING STATUS 031
SYNCHRONIZATION MODE = ETR
  CPC PORT 0 <== ACTIVE      CPC PORT 1
OPERATIONAL                  OPERATIONAL
ENABLED                      ENABLED
ETR NET ID=31                ETR NET ID=31
ETR PORT=21                  ETR PORT=21
ETR ID=13                    ETR ID=12
THIS SYSTEM IS PART OF TIMING NETWORK ITSOPK - 31
  
```

### Coordinated Timing Network ID

A Coordinated Timing Network (CTN) contains a collection of servers that are time synchronized through the Server Time Protocol. They are time synchronized to a time value called Coordinated Server Time (CST). The CST represents the time for the entire network of servers.

The servers that make up a CTN are all configured with a common identifier, referred to as a the Coordinated Timing Network ID. Only servers with the same CTN ID are allowed to become members of the same CTN. All servers in a CTN maintain an identical set of time-control parameters that are used to coordinate the TOD clocks.

A CTN can be configured as either:

- ▶ STP-only CTN

STP-only CTN is a timing network in which all servers are configured to be in STP timing mode. It can only be configured with STP-capable servers, and none of the servers can be in ETR timing mode.

- ▶ Mixed CTN

Mixed CTN allows the coexistence of servers and Coupling Facilities (CFs) synchronized in an ETR network, with servers and CFs that are synchronized with Coordinated Server Time (CST). The Sysplex Timer provides the timekeeping information in a Mixed CTN.

### CTN ID

The CTN ID is an identifier that is used to indicate whether the server has been configured to be part of a CTN and, if so configured, identifies the CTN. The CTN ID is made up of two fields:

- ▶ A field that defines the STP network ID
- ▶ A field that defines the ETR network ID

The *Hardware Management Console* (HMC) is used to define the CTN ID; see Figure 7-4. When both fields are specified, the CTN is referred to as a Mixed CTN. When the STP network ID field is specified, but the ETR network ID field is not specified (set to null or blank), then the CTN is referred to as an STP-only CTN.

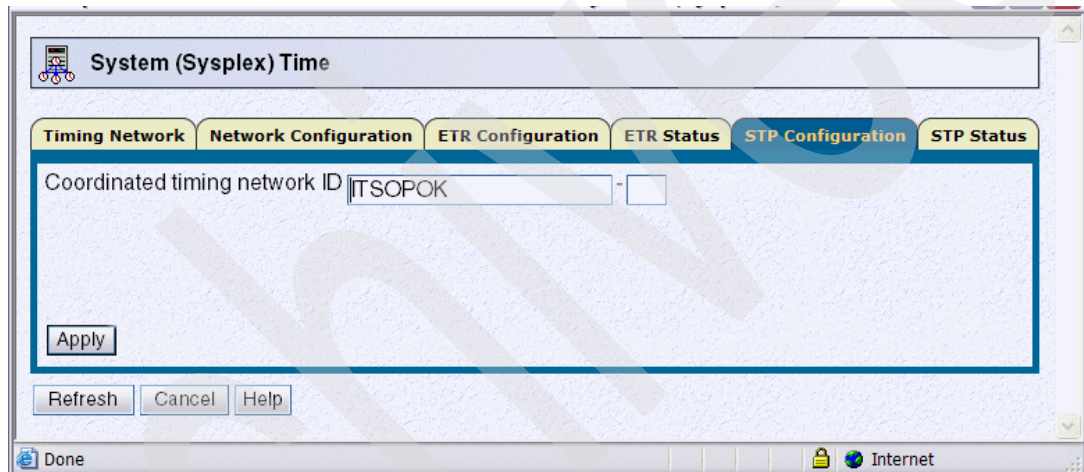


Figure 7-4 HMC workplace - System (Sysplex) Time, Coordinated Timing Network ID

## 7.2.2 Coupling Facility and CFCC considerations for z9 EC

Coupling Facility connectivity to a z9 EC server must be in peer mode. Servers not capable of peer mode connectivity cannot connect directly to a z9 EC CF or z/OS logical partition. z/OS images on a 9672 G5 or G6 server may participate in the same Parallel Sysplex as a z9 EC, but in this type of configuration, a Coupling Facility logical partition cannot reside on z9 EC or 9672 G5/G6. If the z9 EC is in a Mixed or STP-only CTN, it is not possible to have a 9672 G5/G6 in a Parallel Sysplex.

The CFCC logical partition must reside on a server that provides both peer and compatibility mode coupling links. A zSeries server will provide the intermediate CFs that can act as a bridge to 9672 G5/G6 and z9 EC logical partitions in the same Sysplex, until all the z/OS images can be migrated off the 9672 servers.

The level of Coupling Facility Control Code (CFCC) must also be considered. See Table 7-1 for Coupling Facility Control Code requirements when the Coupling Facility resides on a non-z9 EC and is connected to a z/OS image on a z9 EC, or when CF duplexing is used and one Coupling Facility resides on the z9 EC.

The initial support of the CFCC on the z9 EC was level 14, current level is 15.

Table 7-1 z9 EC CF code level considerations

CF with connections to z9 EC z/OS image or z9 EC (CF duplexing)	Connected to a z9 EC
z990 or z890	CFCC level 13 or later
z9 EC	CFCC level 14 or later

To support migration from one CFCC level to the next, you can run different levels of CFCC concurrently as long as the Coupling Facility logical partitions are running on different servers (CF logical partitions running on the same server share the same CFCC level).

### 7.2.3 CFCC enhanced patch apply

This method of patch application enables you to:

- ▶ Apply the patch on one of the available Coupling Facilities. A good place would be the test Coupling Facility of a test sysplex if one is available. If the test of the CFCC code is successful, it can be applied to the production Coupling Facility on the same server. To use the updated CFCC code to a CF logical partition, simply deactivate and reactivate the partition. When the CF comes up, it displays its version on the OPRMSG window for that partition.
- ▶ Continue to run other logical partitions on the server where a disruptive CFCC patch is applied without being impacted by the application of the patch.

Table 7-2 CFCC level supported on a z9 EC

CFLEVEL	Software support <sup>a</sup>
CFLEVEL 14 ▶ Improvements to the CF dispatcher and internal serialization mechanisms designed to better manage coupled workloads	▶ Any in-service z/OS release is required to fully exploit the functions. <ul style="list-style-type: none"> <li>– Optional APAR OA08742 to allow sysplex connectors to request structure allocation in a level 14 CF</li> </ul> ▶ Any in-service z/VM release for virtual CF support.
CFLEVEL 15 ▶ Allowable tasks in the CF increased from 48 to 112.	

a. Always consult the latest PSP bucket for 2094DEVICE and the appropriate subset for the latest maintenance information.

**Note:** When migrating to a new CFCC level, lock, list, and cache structure sizes will typically increase to support new functions. This adjustment can have an impact when the system allocates structures or copies structures from one Coupling Facility to another at different CFCC levels. The Coupling Facility structure sizer tool can size structures for you and takes into account the amount of space needed for the current CFCC levels.

For additional details on CF code levels, see the Parallel Sysplex Web site at:

<http://www-03.ibm.com/systems/z/pso/cftable.html>

For additional details regarding CF configurations, see the document *Coupling Facility Configuration Options*, GF22-5042, also available from the Parallel Sysplex Web site.

The CFSIZER tool can be found at:

<http://www.ibm.com/servers/eserver/zseries/cfsizer>

## 7.2.4 Coupling link connectivity

The type of coupling links you can use to connect a CF to an operating system logical partition is important because of the impact of the link performance on response times and coupling overheads. For configurations covering large distances, the time spent on the link can be the largest part of the response time.

The types of links that are available to connect an operating system logical partition to a Coupling Facility are:

- ▶ **IC:** Licensed Internal Code-defined links to connect a CF to a z/OS logical partition in the same z9 EC. IC links require two CHPIDs to be defined and can only be defined in Peer mode. The link bandwidth is greater than 2 GBps. A maximum of 32 IC links can be defined per z9 EC.
- ▶ **ICB-4:** Connects a z9 EC to z9 EC, z9 BC, z990 or z890. The maximum distance between the two servers is 7 meters (maximum cable length is 10 meters). The link bandwidth is 2 GBps. ICB-4 links can only be defined in Peer mode. The maximum number of ICB-4 links is 16 per z9 EC. ICB-4 supports transmission of STP timekeeping information.
- ▶ **ICB-3:** ICB-3 links are available to connect a z9 EC to z900 or z800; the maximum distance between the two servers is 7 meters (maximum cable length is 10 meters). The link bandwidth is 1 GBps. ICB3 links can only be defined in Peer mode. The maximum number of ICB-3 links is 16 per z9 EC. ICB-3 supports transmission of STP timekeeping information. Although ICB-3 links can be used to connect z9 EC to any System z server, it is not recommended to use it for other than z800 and z900 servers connectivity.
- ▶ **ISC-3:** The ISC-3 feature is available in Peer mode only. ISC-3 links can be used to connect to other System z servers. They are fiber links that support a maximum distance of 10 km, 20 km with RPQ 8P2197, and 100 km with Dense Wave Division Multiplexing (DWDM). ISC-3s operate in single mode only. Link bandwidth is 200 MBps for distances up to 10 km, and 100 MBps when RPQ 8P2197 is installed. Each port operates at 2 Gbps. Ports are ordered in increments of one. The maximum number of ISC-3 links per z9 EC is 48. ISC-3 supports transmission of STP timekeeping information.

Table 7-3 shows the z9 EC coupling link maximums.

Table 7-3 z9 EC coupling link maximums

Link type	STP supported	z9 EC max
IC	No	32
ISC-3	Yes	48
ICB-3	Yes	16
ICB-4	Yes	16
Maximum number of external and internal coupling links combined per z9 EC		64

Table 7-4 lists the coupling link connectivity options for the various servers.

Table 7-4 z9 coupling link connectivity

Connectivity options	z9 ISC-3	z9 ICB-3	z9 ICB-4
z9, z890, and z990 ISC-3	2 Gbps Peer Mode <sup>a</sup>	N/A	N/A
z900 and z800 ICB-3	N/A	1 GBps, Peer mode	N/A
z9, z990, and z890 ICB-3	N/A	1 GBps, Peer mode, recommendation use ICB-4	N/A
z9, z990, and z890 ICB-4	N/A	N/A	2 GBps Peer Mode

a. 1 Gbps when 20km RPQ 8P2197 is installed.

### z9 EC Peer mode links

The z9 EC only supports peer mode links; compatibility mode links are not supported. See Figure 7-5 for peer mode support to System z servers.

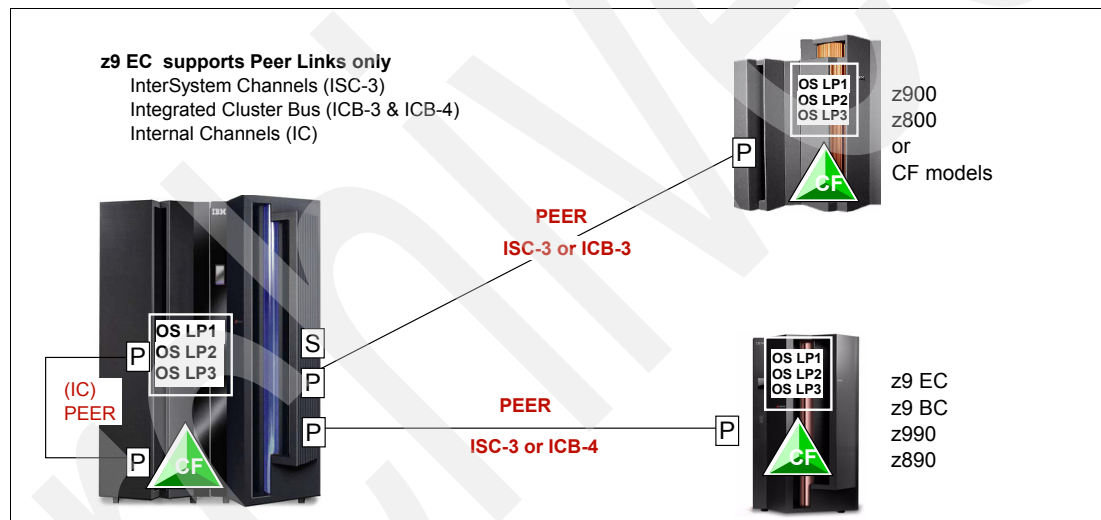


Figure 7-5 z9 EC CF connectivity options

z/OS and Coupling Facility (CF) images may be running on the same or on separate servers. There must be at least one CF connected to all z/OS images, although there can be other CFs which are connected only to selected z/OS images. Two Coupling Facility images are required for system-managed CF structure duplexing and each z/OS image has to be connected to both duplexed CFs in this case.

For availability reasons, there should be at least:

- ▶ Two coupling links between z/OS and Coupling Facility images.
- ▶ Two Coupling Facility images not running on the same server.
- ▶ One stand-alone Coupling Facility; if using system-managed CF structure duplexing or running with *Resource Sharing* only, then a stand-alone Coupling Facility is not mandatory.

## Coupling links and Server Time Protocol

Server Time Protocol (STP) is a message-based protocol in which STP timekeeping information is passed over data links between servers. The timekeeping information is transmitted over externally defined coupling links. Coupling links that can be used to transport STP messages are the InterSystem Channel-3 (ISC-3) links in Peer mode, the Integrated Cluster Bus 4 (ICB-4) links, or ICB-3 links.

There are advantages to using the ISC-3, ICB-4, and ICB-3 links to exchange STP message, as described here:

- ▶ By using the same links to exchange timekeeping information and Coupling Facility messages in a Parallel Sysplex, STP can scale with distance. Servers exchanging messages over short distances, such as ICB-3 and ICB-4 links, can meet more stringent synchronization requirements than servers exchanging messages over long ISC-3 links (distances up to 100 km). This is an enhancement over the Sysplex Timer implementation, which does not scale with distance.
- ▶ Coupling links also provide the connectivity needed in a Parallel Sysplex. Therefore, there is a potential benefit of minimizing the number of cross-site links required in a multi-site Parallel Sysplex.

### ***Coupling link redundancy for STP***

Between any two servers that are intended to exchange STP messages, it is recommended that each server be configured such that at least two coupling links exist for communication between the servers. This prevents the loss of one link causing the loss of STP communication between the servers. If a server does not have a CF logical partition, timing-only links can be used to provide STP connectivity.

The maximum number of attached servers supported by any STP-configured server in a Coordinated Timing Network is equal to the maximum number of coupling links supported by the server in the configuration. On the z9 EC, this value is equal to 64, which is the maximum number of combined ISC-3, ICB-3, and ICB-4 links.

For more details, refer to *Server Time Protocol Planning Guide*, SG24-7280, and *Server Time Protocol Implementation Guide*, SG24-7281.

## 7.2.5 ICF processor assignments

One advantage of using ICF processors instead of CPs for Coupling Facility images is that software licenses are not charged for ICF processors.

CPs are Processor Units used to process z/OS, CFCC, z/VM, Linux, TPF, VSE/ESA, or z/VSE instructions. The logical partition can use dedicated *or* shared CPs. However, it is not possible to have a logical partition with dedicated *and* shared CPs at the same time.

ICFs are PUs dedicated to process the CF Control Code (CFCC) on a Coupling Facility image, which is always running on a logical partition. A CF image can use dedicated *and* shared ICFs. It can also use dedicated *or* shared CPs. With Dynamic ICF expansion, a Coupling Facility image can also use dedicated ICFs and shared CPs.

The z9 EC can have ICF processors defined to CF images.

A Coupling Facility image can have one of the following combinations defined in the image profile:

- ▶ Dedicated ICFs
- ▶ Shared ICFs
- ▶ Dedicated *and* shared ICFs
- ▶ Dedicated CPs
- ▶ Shared CPs
- ▶ Dedicated ICFs *and* shared CPs

Shared ICFs add flexibility. However, running with shared Coupling Facility Processor Units (ICFs or CPs) only is not a recommended production configuration.

In Figure 7-6, the server on the left has two environments defined (production and test), each having one z/OS and one Coupling Facility image. The Coupling Facility images are sharing the same ICF processor. The logical partition processing weights are used to define how much processor capacity each Coupling Facility image can have. The *Capped* option can also be set for the Test Coupling Facility image, to protect the production environment. Connections between these z/OS and Coupling Facility images can use IC channels to avoid the use of real (external) coupling channels and to get the best link bandwidth available.

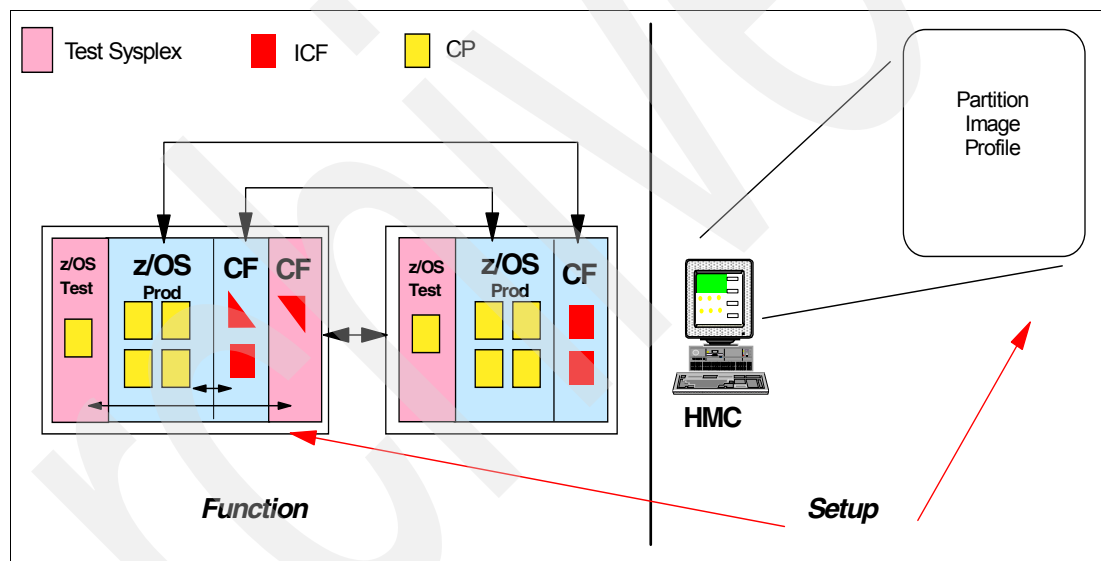


Figure 7-6 ICF options - Shared ICFs

## 7.2.6 Dynamic CF dispatching and Dynamic ICF expansion

The CF Control Code (CFCC), the *CF Operating System*, is implemented using the *Active Wait* technique. This means it is always running (processing or searching for service) and never enters into a wait state. This also means that it gets all the processor capacity (cycles) available for the Coupling Facility logical partition. If this logical partition uses only dedicated processors (CPs or ICFs), this is not a problem. But this may not be desirable when it uses shared processors (CPs or ICFs).

Dynamic CF dispatching provides the following function on a Coupling Facility: If there is no work to do, it enters into a wait state (by time). After an elapsed time, it wakes up to see if there is any new work to do (requests in the CF Receiver buffer). If there is no work, it will sleep again for a longer period of time. If there is new work, it enters into the normal Active Wait until there is no more work, starting the process all over again. This saves processor

cycles and is an excellent option to be used by a production backup CF or a testing environment CF. This function is activated by the CFCC command DYNDISP ON.

The CPs can run z/OS operating system images and CF Images. For software charge reasons, it is better to use ICF processors to run Coupling Facility images.

Figure 7-7 shows the dynamic CF dispatching.

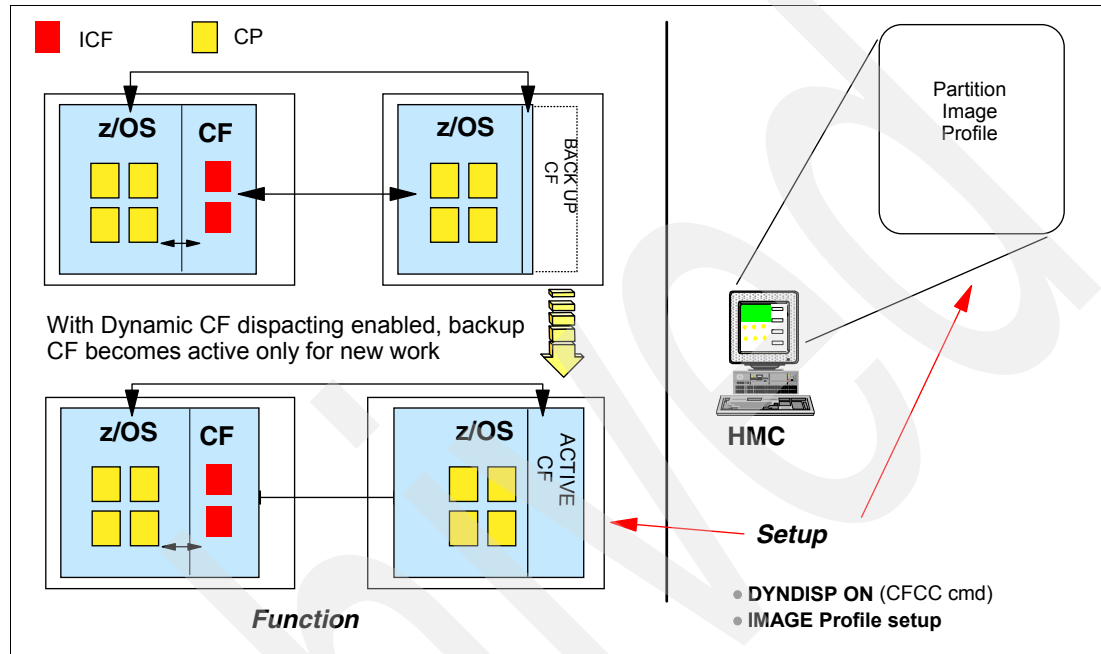


Figure 7-7 Dynamic CF dispatching (shared CPs or shared ICF PUs)

With Dynamic ICF expansion, a Coupling Facility image using one or more *dedicated* ICFs can also use one or more *shared* CPs of this same server. The Coupling Facility image uses the shared CPs only when needed, that is, when its workload requires more capacity than its dedicated ICFs have. This may be necessary during peak periods or during recovery processes.

Care must be exercised when defining CFs with a mix of shared and dedicated processors because in some cases it is possible that serialization held on a shared processor may interfere with the operation on dedicated processors. The recommended production configuration is to have one or more dedicated processors in the CF images.

Figure 7-8 on page 195 shows an example where the server on the left has a production and a test Coupling Facility that has dedicated and shared ICF PUs. This configuration enables the Coupling Facilities to utilize the shared ICF PUs when workload becomes excessive. Additionally, if the alternate production Coupling Facility goes down (for maintenance, for example) and the allocated ICF's capacity on the left server is not big enough to maintain its own workload plus that of the other Coupling Facility, then with Dynamic ICF expansion, the remaining Coupling Facility image can be expanded over shared ICF PUs.

Dynamic ICF expansion can also be configured using dedicated ICF PUs and shared CPs from the z/OS image. The z/OS image *must* have all CPs defined as *shared* and the Dynamic CF Dispatch function must be activated. Dynamic ICF expansion is available on models that have at least one ICF.

Dynamic ICF expansion requires that Dynamic CF Dispatching be activated (DYNDISP ON).



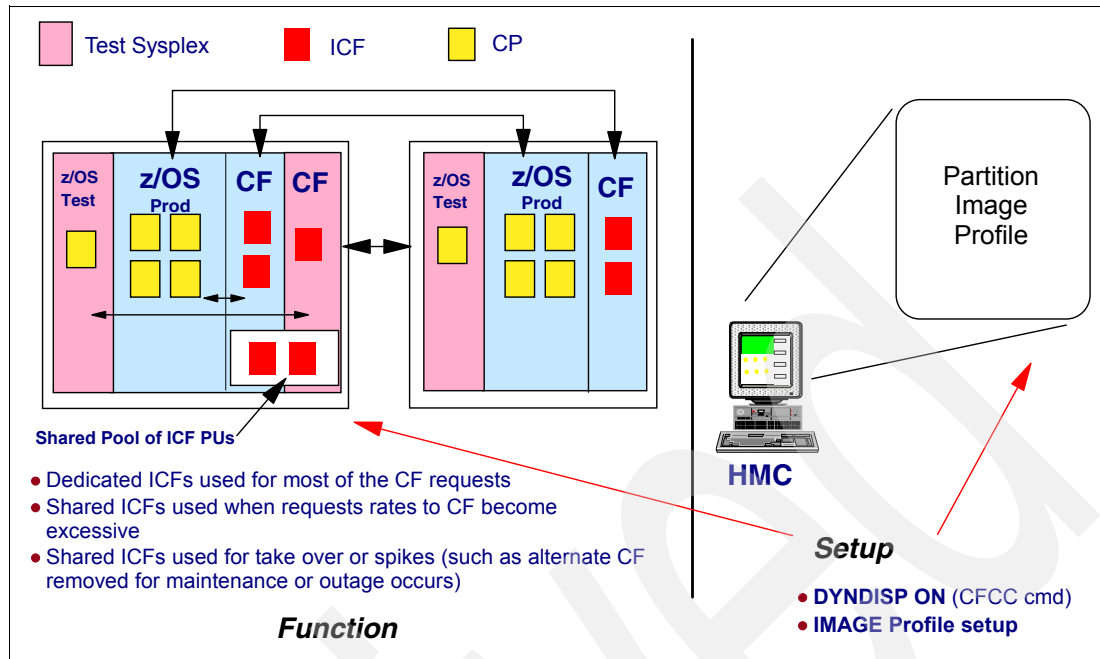


Figure 7-8 CF options - Dynamic ICF expansion

## 7.3 System-managed CF structure duplexing

System-managed Coupling Facility structure duplexing provides a general purpose, hardware-assisted, easy-to-exploit mechanism for duplexing CF structure data. This provides a robust recovery mechanism for failures, such as loss of a single structure or Coupling Facility, or loss of connectivity to a single Coupling Facility, through rapid fail-over to the other structure instance of the duplex pair.

System-managed CF structure duplexing provides:

- ▶ **Improved availability:** Faster recovery of structures is provided by having the data already in the second Coupling Facility when a failure occurs.
- ▶ **Simplified manageability and usability:** Are achieved by a consistent procedure to set up and manage structure recovery across multiple exploiters.
- ▶ **Cost benefits:** Are realized by enabling the use of non-stand-alone Coupling Facilities for all resource sharing and data sharing environments.

System-managed Coupling Facility structure duplexing creates a duplexed copy of the structure in advance of any failure, providing a robust failure recovery capability through fail-over to the unaffected structure instance. This results in:

- ▶ An easily exploited common framework for duplexing the structure data contained in any type of CF structure, with installation control over which structures are duplexed
- ▶ Minimized overhead of duplexing during mainline operation through hardware-assisted serialization and synchronization between the primary and secondary structure updates
- ▶ Maximized availability in failure scenarios by providing a rapid failover to the unaffected structure instance of the duplexed pair, with very little disruption to the ongoing execution of work by the exploiter and applications

Structure failures, CF failures, or losses of CF connectivity can be handled by:

1. Hiding the observed failure condition from the active connectors to the structure, so that they do not perform unnecessary recovery actions
2. Switching over to the structure instance that did not experience the failure
3. Re-establishing a new duplex copy of the structure if appropriate as the Coupling Facility becomes available again, or on a third CF in the Parallel Sysplex

System messages are generated as the structure falls back to simplex mode for monitoring and automation purposes. The structure operates in simplex mode until a new duplexed structure can be established, and can be recovered using whatever existing recovery techniques are supported by the exploiter.

### Configuration planning

A connectivity requirement for system-managed CF structure duplexing is that there must be bi-directional CF-to-CF connectivity between each pair of CFs in which duplexed structure instances reside.

With peer links, connectivity can be provided by a single bi-directional link (two for redundancy). CF-to-CF coupling links can either be dedicated or shared through MIF. They can be shared with coupling links between z/OS and Coupling Facility images in the pair of servers they connect. When planning sharing links, remember that peer links can only be shared by one Coupling Facility partition. Figure 7-9 gives an overview of system-managed CF structure duplexing.

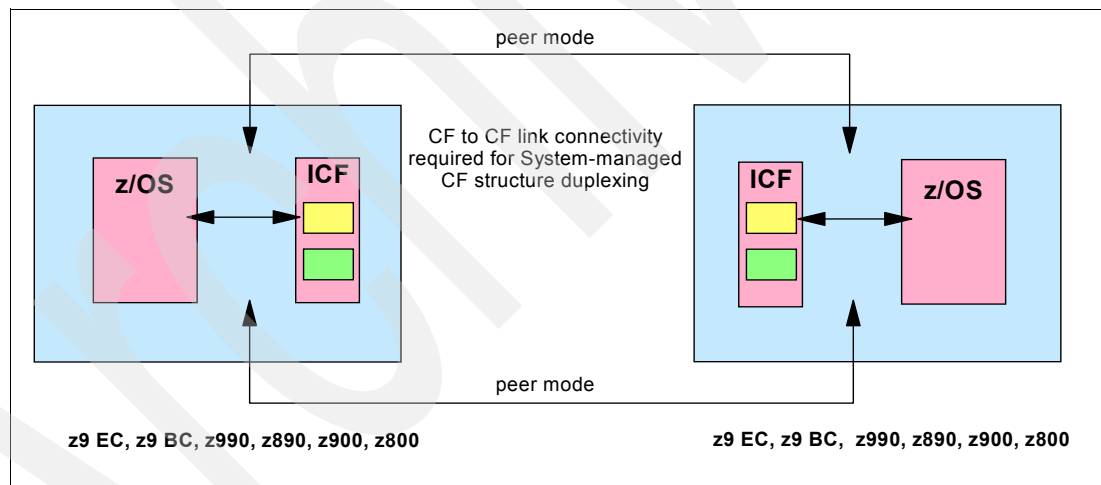


Figure 7-9 System-managed CF structure duplexing

A technical paper on system-managed CF structure duplexing is available at:

<http://www-1.ibm.com/servers/eserver/zseries/library/techpapers/gm130103.html>

## 7.4 Intelligent Resource Director

Intelligent Resource Director (IRD) is only available on System z running z/OS. IRD is a function that optimizes processor CPU and channel resource utilization across logical partitions within a single System z server.

### IRD overview

The Intelligent Resource Director (IRD) is a z/OS feature, extending the concept of goal-oriented resource management by allowing you to group system images that are resident on the same System z server running in LPAR mode, and in the same Parallel Sysplex, into an *LPAR cluster*. This gives Workload Management the ability to manage resources, both processor and I/O, not just in one single image, but across the entire cluster of system images.

Figure 7-10 shows an LPAR cluster. It contains three z/OS images, and one Linux image managed by the cluster. Note that included as part of the entire Parallel Sysplex is another z/OS image, as well as a Coupling Facility image. In this example, the scope that IRD has control over is the defined LPAR cluster.

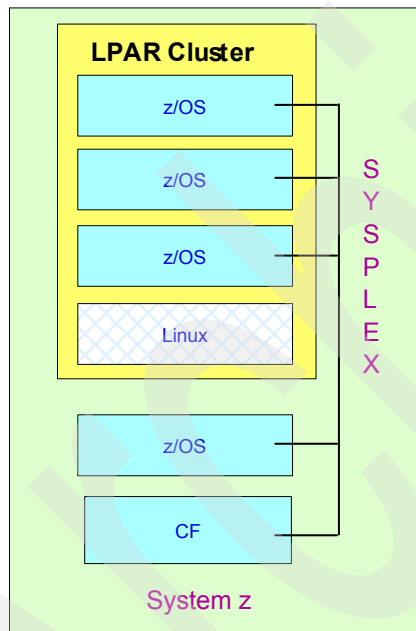


Figure 7-10 IRD LPAR cluster example

IRD addresses three separate but mutually supportive functions:

- ▶ LPAR CPU management

WLM dynamically adjusts the number of logical processors within a logical partition and the processor weight based on the WLM policy. The ability to move the CPU weights across an LPAR cluster provides processing power to where it is most needed, based on WLM goal mode policy.

- ▶ Dynamic channel path management (DCM)

DCM moves ESCON channel bandwidth between disk control units to address current processing needs. The z9 EC supports DCM within a Channel Subsystem.

- ▶ Channel Subsystem Priority Queuing

This function on the System z allows the priority queuing of I/O requests in the Channel Subsystem and the specification of relative priority among logical partitions. WLM in goal mode sets the priority for a logical partition and coordinates this activity among clustered logical partitions.

## 7.4.1 LPAR CPU management

LPAR CPU management allows WLM working in goal mode to manage the processor weighting and logical processors across an LPAR cluster.

LPAR CPU management dynamically manages non-z/OS operating systems, such as Linux and z/VM. This function allows z/OS WLM to manage the CPU resources given to these partitions based on their relative importance compared to the other workloads running in the same LPAR cluster.

Workload Manager distributes processor resources across an LPAR cluster by dynamically adjusting the logical partition weights in response to changes in the workload requirements. When important work is not meeting its goals, WLM will raise the weight of the partition where that work is running, thereby giving it more processing power. As the logical partition weights change, the number of online logical CPUs may also be changed to maintain the closest match between logical CPU speed and physical CPU speed.

Enabling LPAR CPU management involves defining the Coupling Facility structure and then performing several operations on the hardware management console: Defining logical CPs, and setting initial, minimum, and maximum processing weights for each logical partition.

CPU resources are automatically moved toward logical partitions with the most need by adjusting the partition's weight. The sum of the weights for the participants in an LPAR cluster is viewed as a pooled resource that can be apportioned among the participants to meet the goal mode policies. The installation can place limits on the processor weight value.

WLM will also manage the available processors by varying off unneeded CPs (more logical CPs implies more parallelism, and less weight per CP).

The benefits of CPU management include:

- ▶ Logical CPs perform at the fastest uniprocessor speed available.

This results in the number of logical CPs tuned to the number of physical CPs of service being delivered by the logical partition current weight. If the logical partition is getting four equivalent physical CPs of service and has eight logical CPs online to z/OS, then each logical CP only gets half of an equivalent physical CP. For example, if a CP delivers 200 MIPS, half of it will deliver 100 MIPS. This occurs because each logical CP gets fewer time slices.

- ▶ Reduced PR/SM overhead.

There is a PR/SM overhead for managing a logical CP. The higher the number of logical CPs in relation to the number of equivalent physical CPs, the higher the PR/SM overhead. This is because PR/SM has to do more processing to manage the number of logical CPs that exceeds the number of equivalent physical CPs.

- ▶ z/OS gets more control over how CP resources are distributed.

Using CPU management, z/OS is able to manage CP resources in relation to WLM goals for work. This was not possible in the past when a logical partition had CP resources assigned and used these as best it could in one logical partition. z/OS is able to change the assigned CP resources (logical partition weights) and place them where they are required for the work. CPU management does the following:

- Identifies what changes are needed and when.
- Projects the likely results on both the work it is trying to help and the work that it will be taking the resources from.
- Performs the changes.
- Analyzes the results to ensure the changes have been effective.

There is also the question of the speed at which an operator can perform these actions. WLM can perform these actions every Policy Adjustment interval, which is normally ten seconds, as determined by WLM. It is not possible for an operator to perform all the tasks in this time.

For additional information about implementing LPAR CPU management under IRD, see the Redbooks publication *z/OS Intelligent Resource Director*, SG24-59522.

## 7.4.2 Dynamic Channel Path Management

There is no such thing as a *typical* workload. The requirements for processor capacity, I/O capacity, and other resources vary throughout the day, week, month, and year.

Dynamic Channel Path Management (DCM) provides the ability to have the system automatically manage the number of paths available to disk subsystems. By making additional paths available where they are needed, the effectiveness of your installed channels is increased, and the number of channels required to deliver a given level of service is potentially reduced.

DCM also provides availability benefits by attempting to ensure that the paths it adds to a control unit have as few points of failure in common with existing paths as possible, and configuration management benefits by allowing the installation to define a less specific configuration. On a z9 EC where paths can be shared by Multiple Image Facility (MIF), DCM will coordinate its activities across logical partitions within a single Channel Subsystem on a server within a LPAR Cluster.

Where several channels are attached from a CSS to a switch, they can be considered a resource pool for accessing any of the control units attached to the same switch. To achieve this without DCM would require deactivating paths, performing a dynamic I/O reconfiguration, and activating new paths. DCM achieves the equivalent process automatically, using those same mechanisms.

Channels managed by DCM are referred to here as *managed* channels. Channels not managed by DCM are referred to as *static* channels.

Workload Manager dynamically moves channel paths through the ESCON Director from one I/O control unit to another in response to changes in the workload requirements. By defining a number of channel paths as managed, they become eligible for this dynamic assignment.

By moving more bandwidth to the important work that needs it, your disk I/O resources are used much more efficiently. This may decrease the number of channel paths you need in the first place, and could improve availability because, in the event of a hardware failure, another channel could be dynamically moved over to handle the work requests.

Dynamic Channel Path Management runs on a System z server in z/Architecture mode, in both basic and LPAR mode. The participating z/OS system images can be defined as XCFLOCAL, MONOPLEX, or MULTISYSTEM.

If a system image running Dynamic Channel Path Management in LPAR mode is defined as being part of a multisystem sysplex, it also requires a CF level 9 Coupling Facility structure, even if it is the only image currently running on the system.

Dynamic Channel Path Management operates in two modes:

- ▶ **Balance mode:** In balance mode, DCM will attempt to equalize performance across all of the managed control units.
- ▶ **Goal mode:** In goal mode, which is available only when WLM is operating in goal mode on systems in an LPAR cluster, DCM will still attempt to equalize performance, as in balance mode. In addition, when work is failing to meet its performance goals due to I/O delays, DCM will take additional steps to manage the channel bandwidth accordingly, so that important work meets its goals.

Enabling Dynamic Channel Path Management involves defining managed channels and control units through HCD. On the Hardware Management Console, you then need to ensure that all of the appropriate logical partitions are authorized to control the I/O configuration.

For additional information about implementing Dynamic Channel Path Management under IRD, see *z/OS Intelligent Resource Director*, SG24-59522.

Dynamic Channel Path Management provides the following benefits:

- ▶ **Help in overall image performance:** Achieved by automatic path balancing (WLM compatibility and goal mode) and Service Policy (WLM goal mode).
- ▶ **Maximum utilization of installed hardware:** Channels will be automatically balanced, providing opportunities to use fewer I/O paths to service the same workload.
- ▶ **Simple I/O definition:** The connection between managed channels and managed control units is not explicitly defined.
- ▶ **Reduced skills required to manage z/OS:** Managed channels and control units are automatically monitored, balanced, tuned, and reconfigured.
- ▶ **Channel availability:** A failing or hung channel path will result in reduced throughput on the affected control unit. DCM will rapidly detect the symptom and augment the paths, automatically bypassing the problem. The problem will still have to be analyzed and corrected by site personnel.

DCM will automatically analyze and minimize bottlenecks on an I/O path by selecting appropriate paths. DCM is sensitive to bottlenecks, such as:

- ESCON channel cards
- Processor Self-Timed Interconnect
- MBA fanout cards
- Books
- Director port cards
- Control Unit I/O bay
- Control Unit Interface card
- ESCON Director

### 7.4.3 Channel Subsystem Priority Queuing

Channel Subsystem (CSS) Priority Queuing is a function available on System z processors in LPAR mode. It allows the z/OS operating system to specify a priority value when starting an I/O request. When there is contention causing queuing in the Channel Subsystem, the request is prioritized by this value.

If important work is missing its goals due to I/O contention on channels shared with other work, it will be given a higher Channel Subsystem I/O priority than the less important work. This function goes hand-in-hand with the Dynamic Channel Path Management described previously: As additional channel paths are moved to control units to help an important workload meet goals, Channel Subsystem Priority Queuing ensures that the important work receives greater access to additional bandwidth than less important work that happens to be using the same channel.

Channel Subsystem Priority Queuing runs on a System z server in z/Architecture mode. The participating z/OS system images can be defined as XCFLOCAL, MONOPLEX, or MULTISYSTEM. It is optimized when WLM is running in goal mode. It does not require a Coupling Facility structure.

Enabling Channel Subsystem Priority Queuing involves defining a range of I/O priorities for each logical partition on the hardware management console, and then turning on the “Global input/output (I/O) priority queuing” switch. You also need to specify YES for WLM I/O priority management setting.

z/OS will set the priority based on a goal mode WLM policy. This complements the goal mode priority management that sets I/O priority for IOS UCB queues, and for queuing in the 2105 ESS disk subsystem.

CSS Priority Queuing uses different priorities calculated in a different way from the I/O priorities used for UCB and control unit queuing.

The benefits proved by Channel Subsystem Priority Queuing include:

- ▶ Improved performance

I/O from work that is not meeting its goals may be given priority over I/O from work that is meeting its goals, providing Workload Manager with an additional method for adjusting I/O performance. Channel Subsystem Priority Queuing is complementary to UCB priority queuing and control unit priority queuing, each addressing a different queuing mechanism that may affect I/O performance.

- ▶ Reduced skills required to manage z/OS

Monitoring and tuning requirements are reduced because of the self-tuning abilities of the Channel Subsystem.

## 7.4.4 WLM and Channel Subsystem priority

WLM assigns the highest to lowest CSS priority, as given in Table 7-5. It assigns eight priority levels.

Table 7-5 WLM-assigned CSS I/O priorities

Workload type	Priority
System work.	FF
Importance of one and two missing goals.	FE
Importance of three and four missing goals.	FD
Meeting goals. Adjust by ratio of connect time to elapsed time.	F9-FC
Discretionary.	F8

Work that is meeting its WLM target is assigned CSS priorities between F9 and FC, depending on its execution profile. Work that has a light I/O usage has its CSS priority moved upwards.

When an I/O operation is started by a CP on the Server, it can be queued by the Channel Subsystem for several reasons, including Switch port busy, Control unit busy, Device busy, and All channel paths busy. Queued I/O requests are started or restarted when an I/O completes or the Control unit indicates the condition has cleared. Where two or more I/O requests are queued in the Channel Subsystem, the CSS LIC on the System z selects the requests in priority order. The LIC also ages requests to ensure that low priority requests are not queued for excessive periods.

In the logical partition image profile for the z/OS image, there are two specifications that relate to the Channel Subsystem I/O Priority Queuing. They are:

- ▶ The range of priorities that will be used by this image
- ▶ The default Channel Subsystem I/O priority

For images running operating systems that do not support Channel Subsystem priority, the customer can prioritize all the Channel Subsystem requests coming from that image against the other images by specifying a value for the default priority.

Within an LPAR cluster, the prioritization is managed by WLM goal mode and coordinated across the cluster. Hence, the range should be set identically for all logical partitions in the same LPAR cluster.

WLM sets priorities within a range of eight values that will be mapped to the specified range. If a larger range is specified, WLM uses the top eight values. If a smaller range is specified, WLM maps its values into the smaller range, retaining as much function as possible within the allowed range. Note that the WLM calculated priority is still a range of 8. The mapped priority is shown in Table 7-6 on page 203.

A range of eight values is recommended for CSS I/O priority-capable logical partitions. If the logical partition is run in compatibility mode or with I/O priority management disabled, the I/O priority is set to the middle of the specified range.



Table 7-6 WLM CSS priority range mapping with specified range less than eight

WLM CSS priorities (range width)	Calculated range (8)	Specified range (7)	(6)	(5)	(4)	(3)	(2)
System work.	FF	FF	FF	FF	FF	FF	FF
Importance of one and two missing goals.	FE	FE	FE	FE	FE	FE	FF
Importance of three and four missing goals.	FD	FE	FE	FE	FE	FE	FF
Meeting goals. Adjust by ratio of connect time to elapsed time.	FC-F9	FD-FA	FD-FB	FD-FC	FD	FE	FF
Discretionary.	F8	F9	FA	FB	FC	FD	FE

## 7.4.5 Special considerations and restrictions

To use Sysplex functions, there are several considerations and restrictions of which you must be aware.

### Unique LPAR cluster names

LPAR clusters running on a 2064, 2066, 2084, 2086, or 2094 server must be uniquely named. This is the sysplex name that is associated with the LPAR cluster. Managed channels have an affinity (are owned by) to a specific LPAR cluster. Non-unique naming creates problems in terms of scope of control.

### Disabling Dynamic Channel Path Management

To disable Dynamic Channel Path Management within an LPAR cluster running z/OS, turning off the function by using the SETIOS DCM=OFF command is not sufficient. Although a necessary step, this does not ensure that the existing configuration is adequate to handle your workload needs, since it leaves the configuration in the state it was at the time the function was disabled. During your migration to DCM, we recommend that you continue to maintain your old IODF until you are comfortable with DCM. This will allow you to back out of DCM by activating a known configuration.

### Automatic I/O interface reset

When going through all of the steps to enable Dynamic Channel Path Management, also ensure that the "Automatic input/output (I/O) interface reset" option is enabled on the Hardware Management Console. This will allow Dynamic Channel Path Management to continue functioning in the event that one participating system image fails.

### System automation - I/O operations

When using system automation, take care when using PROHIBIT or BLOCK on a port that is participating in Dynamic Channel Path Management.

When blocking a managed channel port, configuring the CHPID OFFLINE to all members of the LPAR Cluster is all that is required. Dynamic Channel Path Management will ensure that if the CHPID is configured to managed subsystems, then the CHPID will be de-configured from all subsystems to which it is currently configured.

When blocking a port connected to a managed subsystem, the port must first be disabled for Dynamic Channel Path Management usage. This is done using the VARY SWITCH command to take the port OFFLINE to Dynamic Channel Path Management. This command should be issued on all partitions that are running DCM. Disabling the port for Dynamic Channel Path

Management usage will de-configure all managed channels that are connected to the subsystem through that port. Once the port is disabled to Dynamic Channel Path Management, it can then be blocked.

When prohibiting a set of ports, if any of the ports are connected to managed subsystems, then the PROHIBIT operation must be preceded by the VARY SWITCH commands to disable the managed subsystem ports to Dynamic Channel Path Management. As in the blocking case, this will cause any managed channels currently connected to the subsystem ports to be de-configured. Once the subsystem ports are disabled to Dynamic Channel Path Management, the PROHIBIT function can be invoked. This must then be followed by the VARY SWITCH commands to re-enable the prohibited subsystem ports to Dynamic Channel Path Management.

When ports are unprohibited or unblocked, these operations need to be followed, as necessary, by VARY SWITCH commands to bring ports ONLINE to Dynamic Channel Path Management.

## Concurrent upgrades and availability

This chapter describes features for availability and capacity upgrades.

The z9 EC is focused on providing higher availability and reducing planned and unplanned outages, which, when properly configured, may be accomplished with improved nondisruptive replace, repair, and upgrade functions for memory, books, and I/O, as well as extending nondisruptive capability to download Licensed Internal Code updates. In most cases, a z9 EC capacity upgrade can be *nondisruptive*, without a system outage.

The following sections are included:

- ▶ 8.1, “Availability enhancements” on page 206
- ▶ 8.2, “Concurrent upgrades” on page 207
- ▶ 8.3, “Enhanced Book Availability (EBA)” on page 232
- ▶ 8.4, “Enhanced Driver Maintenance (EDM)” on page 242
- ▶ 8.5, “Nondisruptive upgrades” on page 244

## 8.1 Availability enhancements

The following functions are unique to the System z9:

- ▶ **Enhanced Book Availability (EBA):** The z9 EC is designed to allow a single book, in a multibook server, to be concurrently removed from the server and reinstalled during an upgrade or repair action. Enhanced Book Availability is an extension of the support for Concurrent Book Add (CBA) delivered on z990. CBA is designed to allow you to concurrently upgrade a z9 EC by integrating a second, third, or fourth book into the server without affecting application processing.
- ▶ **Concurrent memory upgrade or replacement:** Memory can be upgraded concurrently using LIC-CC if physical memory is available on the books. If the physical memory cards need to be changed on a multiple books configuration, which would require the book to be removed, the Enhanced Book Availability function can prove useful. It would require the availability of additional resources on other books or reducing the need for resources during this action. To help ensure that you have the appropriate level of memory in a multiple book configuration, you may want to consider the selection of the flexible memory option (FC 2802 through FC 2824) to provide additional resources, to exploit EBA, when repairing a book or memory on a book, or when upgrading memory where larger memory cards might be required.
- ▶ **Enhanced Driver Maintenance (EDM):** One of the greatest contributors to downtime during planned outages is Licensed Internal Code (LIC) driver updates performed in support of new features and functions. When properly conditioned, the z9 EC is designed to support activating a selected new driver level concurrently.
- ▶ **Concurrent MBA fanout addition or replacement:** A Memory Bus Adapter (MBA) fanout card is designed to provide the path for data between memory and I/O using Self-Timed Interconnect (STI) cables. With the z9 EC, a hot-pluggable and concurrently upgradeable MBA fanout card is available. Up to eight MBA fanout cards are available per book for a total of up to 32 MBA fanout cards on the z9 EC when four books are installed. In the event of an outage, an MBA fanout card, used for I/O, may be concurrently repaired using Redundant I/O Interconnect.
- ▶ **Redundant I/O Interconnect:** Redundant I/O Interconnect helps maintain critical connections to devices. The z9 EC allows a single book, in a multibook server, to be concurrently removed and reinstalled during an upgrade or repair, continuing to provide connectivity to the server I/O resources using a second path from a different book.
- ▶ **Dynamic oscillator switch-over:** The z9 EC has two oscillator cards, a primary and a backup. In the event of a primary card failure, the backup card is designed to detect the failure, switch-over, and provide the clock signal to the server transparently.

Now let us look in more detail at the key On Demand capabilities for permanent, temporary, emergency, and disaster recovery increase in capacity of the z9 EC.

## 8.2 Concurrent upgrades

The z9 EC has the capability of concurrent upgrades, providing additional capacity with no *server* outage. In most cases, with prior planning and operating system support, a concurrent upgrade can also be nondisruptive to the operating system.

Given today's business environment, the benefits of the concurrent capacity growth capabilities provided by the z9 EC are plentiful, and include:

- ▶ Enabling exploitation of new business opportunities
- ▶ Supporting the growth of e-business environments
- ▶ Managing the risk of volatile, high growth, and high volume applications
- ▶ Supporting 24x365 application availability
- ▶ Enabling capacity growth during “lock down” periods

This capability is based on the flexibility of the z9 EC design and structure, which allows configuration control by the Licensed Internal Code (LIC) and concurrent hardware installation.

The sub-capacity models add to the configuration granularity within the family. This added granularity is available for models configured with up to eight CPs and provides 24 capacity settings. Sub-capacity models provide for CP capacity increase in two ways that can be used together to deliver configuration granularity. The first way is by adding CPs, the second way is by changing the capacity identifier of the CPs currently installed.

### ***Licensed Internal Code (LIC) based upgrades***

The LIC - Configuration Control (LIC-CC) provides for server upgrade with no hardware changes by enabling the activation of additional, previously installed capacity. Concurrent upgrades through LIC-CC can be done for:

- ▶ Processors (CPs, IFLs, ICFs, zAAPs, and zIIPs), when spare PUs are available on the installed books or the capacity identifier can be increased
- ▶ Memory, when spare capacity is available on the installed memory card
- ▶ I/O cards ports (ESCON channels and ISC-3 links), when ports are available on the installed I/O cards

### ***Concurrent hardware installation upgrades***

Configuration upgrades can also be concurrent by installing additional:

- ▶ Books (which contain processors, memory, and STIs), when book slots are available in the CEC cage.
- ▶ MBA fanouts.
- ▶ STI-A8, STI-A4, and STI-MP cards.
- ▶ I/O cards, when slots are still available on the installed I/O cages. I/O cages *cannot* be installed concurrently.

The concurrent upgrade capability can be better exploited when a future target configuration is considered in the initial configuration. Using the Plan Ahead concept, the required number of I/O cages for concurrent upgrades, up to the target configuration, can be included in the z9 EC initial configuration.

## Concurrent PU conversions

The z9 EC support concurrent conversion between different PU types, providing flexibility to meet changing business environments.

These LIC-CC based PU conversions, as listed in Table 2-11 on page 73, require that at least one PU, either CP, ICF, or IFL, remains unchanged; otherwise, the conversion is disruptive. The PU conversion generates a new LIC-CC that can be installed concurrently in two steps. First, the assigned PU is removed from the z9 EC configuration. Second, the newly available PU is activated as the new PU type.

Logical partitions may also need to “free” PUs to be converted, and the operating systems must have the *configure offline/online* support to make the PU conversion nondisruptively.

**Note:** Customer planning and operator action are required to exploit concurrent PU conversion. Note also that PU conversion:

- ▶ Is disruptive if *all* current PUs are converted to different types
- ▶ May require an individual logical partition outage if dedicated PUs are converted

Unassigned CP capacity is recorded by a capacity identifier feature. CP feature conversions change (increase or decrease) the capacity identifier feature.

## Model upgrades

The z9 EC have a machine type and model, 2094-Sxx, and a model capacity identifier 4xx, 5xx, 6xx, or 7xx.

The model indicates how many books are present on the configuration, while the capacity identifier describes how many CPs can be characterized and the capacity setting of the CPs.

A hardware configuration upgrade always requires physical hardware (books) addition. A server upgrade can change either, or both, the server model and the model capacity identifier:

- ▶ LIC-CC only upgrade:
  - May change the server capacity identifier 4xx, 5xx, 6xx, or 7xx if the number of CPs is changed or the capacity setting of the CPs changes.
  - Do not change the server model 2094-Sxx, as no additional books are included.
- ▶ Hardware installation upgrade:
  - Will change the server model 2094-Sxx, because additional books are included.
  - May change the server capacity identifier, 4xx, 5xx, 6xx, or 7xx if the number of CPs change or the capacity setting of the CPs changes.

Both the server model and the capacity can be concurrently upgraded. Concurrent upgrades can be accomplished in both *planned* and *unplanned* upgrade situations.

## Planned upgrades

Planned upgrades can be done using the Capacity Upgrade on Demand (CUoD), the Customer Initiated Upgrade (CIU), or the On/Off Capacity on Demand (On/Off CoD) functions.

CUoD and CIU are functions available to enable *concurrent and permanent* capacity growth. On/Off CoD function enables *concurrent and temporary* capacity growth.

CUoD can concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs), memory, and I/O ports, or change the model capacity identifier on an existing server. The upgrade may be LIC-CC only or LIC-CC along with the addition of hardware (books, memory, and I/O cards). CUoD requires IBM service personnel to perform the upgrade.

CIU can concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs) and memory, or change the model capacity identifier, up to the limit of the installed books on an existing server. CIU is initiated by the customer through the Web using IBM Resource Link, and makes use of CUoD techniques to ensure valid configurations. CIU requires a special contract, between the customer and IBM, where the terms and conditions are agreed to.

On/Off CoD is a function available on z9 EC that enables *concurrent* and *temporary* capacity growth of the server. On/Off CoD *can* be used for customer peak workload requirements, for any length of time, and has a daily hardware and software charge.

On/Off CoD can concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs) up to the limit of the installed books of an existing server, and is restricted to double the currently installed capacity. On/Off CoD uses the CIU ordering process, initiated by the customer through the Web using IBM Resource Link, and makes use of CUoD techniques to ensure valid configuration. On/Off CoD requires a special contract, between the customer and IBM, where the terms and conditions are agreed to.

### **Upgrades for Disaster Recovery**

Unplanned upgrades can be done by the Capacity BackUp (CBU) for emergency or disaster/recovery situations.

CBU is a *concurrent* and *temporary* activation of CPs, ICFs, IFLs, zAAPs, and zIIPs. CBU cannot be used for peak load management of customer workload. A CBU activation can last up to 90 days when a disaster/recovery situation occurs.

CBU features are optional and require uncharacterized PUs to be available on installed books of the back-up server. A CBU contract must be in place before the special code that enables this capability can be loaded on the server. The standard CBU contract provides for five 10 days tests and one 90 day disaster activation over a five year period. Contact your IBM Representative for details.

## Capacity upgrade functions

Table 8-1 summarizes the capacity upgrade functions available.

Table 8-1 Capacity upgrade functions summary

	Upgrades	Via	Type	Process
<b>CUoD</b>	CPs, IFLs, ICFs, zAAPs, zIIPs, Memory, and I/Os	LIC-CC and new hardware installation	Concurrent and permanent	Ordered as a normal upgrade, activated by IBM Service Personnel
<b>CIU</b>	CPs, IFLs, ICFs, zAAPs, zIIPs, and Memory	LIC-CC, no hardware can be added	Concurrent and permanent	Initiated through Web and activated by customer
<b>On/Off CoD</b>	CPs, IFLs, ICFs, zAAPs, and zIIPs	LIC-CC, no hardware can be added.	Concurrent and temporary (no time limit)	Initiated through Web and activated by customer
<b>CBU</b>	CPs, IFLs, ICFs, zAAPs, and zIIPs	LIC-CC, no hardware can be added	Concurrent and temporary (10 days per test and up to 90 days for disaster activation)	Initiated for backup testing or backup situations only and activated by customer

### 8.2.1 Capacity Upgrade on Demand (CUoD)

Capacity Upgrade on Demand (CUoD) is a function that enables *concurrent* and *permanent* capacity growth.

CUoD provides the ability to concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs), memory capacity, and I/O ports. In the case of the sub-capacity models, it provides the ability to concurrently adjust both the number of CPs and the capacity identifier. The concurrent upgrade can be done using Licensed Internal Code Configuration Control (LIC-CC) only, by installing additional books, adding I/O cards, or a combination:

- ▶ CUoD upgrades for processors are done by either:
  - LIC-CC assigning and activating spare PUs up to the limit of the current installed books
  - LIC-CC to adjust the number and type of PUs and or change the capacity identifier
  - Installing additional books and LIC-CC assigning and activating spare PUs on installed books
- ▶ CUoD upgrades for memory are done by either:
  - LIC-CC activating additional memory capacity up to the limit of the memory cards on the current installed books
  - Installing additional books and LIC-CC activating additional memory capacity on installed books
  - Using the EBA capability where possible on multi-book systems to add or change the memory cards.
- ▶ CUoD upgrades for I/O are done by either:
  - LIC-CC activating additional ports on already installed ESCON and ISC-3 cards
  - Installing additional I/O cards and supporting infrastructure if required on already installed I/O cages



**Important:** If the STI Rebalance feature (FC 2400) is selected at server upgrade configuration time, it will change the Physical Channel ID (PCHID) number of ICB-4 links, requiring a corresponding update on the server I/O definition through HCD or HCM.

CUoD is ordered as a “normal” upgrade, also known as Miscellaneous Equipment Specification (MES). CUoD requires IBM service personnel for the upgrade. In most cases, a very short period of time is required for the IBM personnel to install the LIC-CC and complete the upgrade.

To better exploit the CUoD function, an initial configuration should be carefully planned to allow a concurrent upgrade up to a target configuration.

You need to consider planning, positioning, and other issues to allow a CUoD *nondisruptive* upgrade. By planning ahead, it is possible to enable nondisruptive capacity and I/O growth for the z9 EC with no system power down and no associated POR or IPLs.

The model and model capacity identifier returned by the STSI instruction are updated coincident with the upgrade, but the channel CPC Node-Descriptor (NED) information is not updated until the next Power On Reset; see “Channel-to-channel links” on page 181.

The Plan Ahead feature involves pre-installation of additional I/O cages, as it is not possible to install an I/O cage concurrently.

**Note:** CUoD basically provides the “physical” upgrade, resulting in more enabled processors, different capacity setting for the CPs, additional memory, and I/O ports. Additional planning tasks are required for *nondisruptive* logical upgrades (see “Recommendations to avoid disruptive upgrades” on page 246).

### CUoD for processors

CUoD for processors can add, *concurrently*, more CPs, IFLs, ICFs, zAAPs, and zIIPs to a z9 EC by assigning available PUs that reside on the books through LIC-CC. Depending on the quantity of the additional CPs, IFLs, ICFs, zAAPs, or zIIPs in the upgrade, additional books may be required and can be concurrently installed before the LIC-CC enablement. With the sub-capacity models, additional capacity can be provided by adding CPs, by changing the capacity identifier on the current CPs or both.

**Note:** The sum of CPs, inactive CPs, IFLs, unassigned IFLs, ICFs, zAAPs, and zIIPs cannot exceed the maximum limit of PUs available for customer use. The number of zAAPs cannot exceed the number of purchased CPs on a z9 EC. The number of zIIPs cannot exceed the number of purchased CPs on a z9 EC. The combined number of zAAPs and zIIPs cannot exceed 2X the number of purchased CPs.

**Important:** CUoD for processors is not supported when CBU or On/Off CoD is *activated* on a z9 EC. CUoD for processors can be applied after the temporary capacity upgrade through CBU or On/Off CoD is deactivated.

Figure 8-1 is an example of CUoD for processors, showing the eight PUs in book 0 and ten PUs in book 1 that can be assigned as CPs, IFLs, ICFs, zAAPs, or zIIPs.

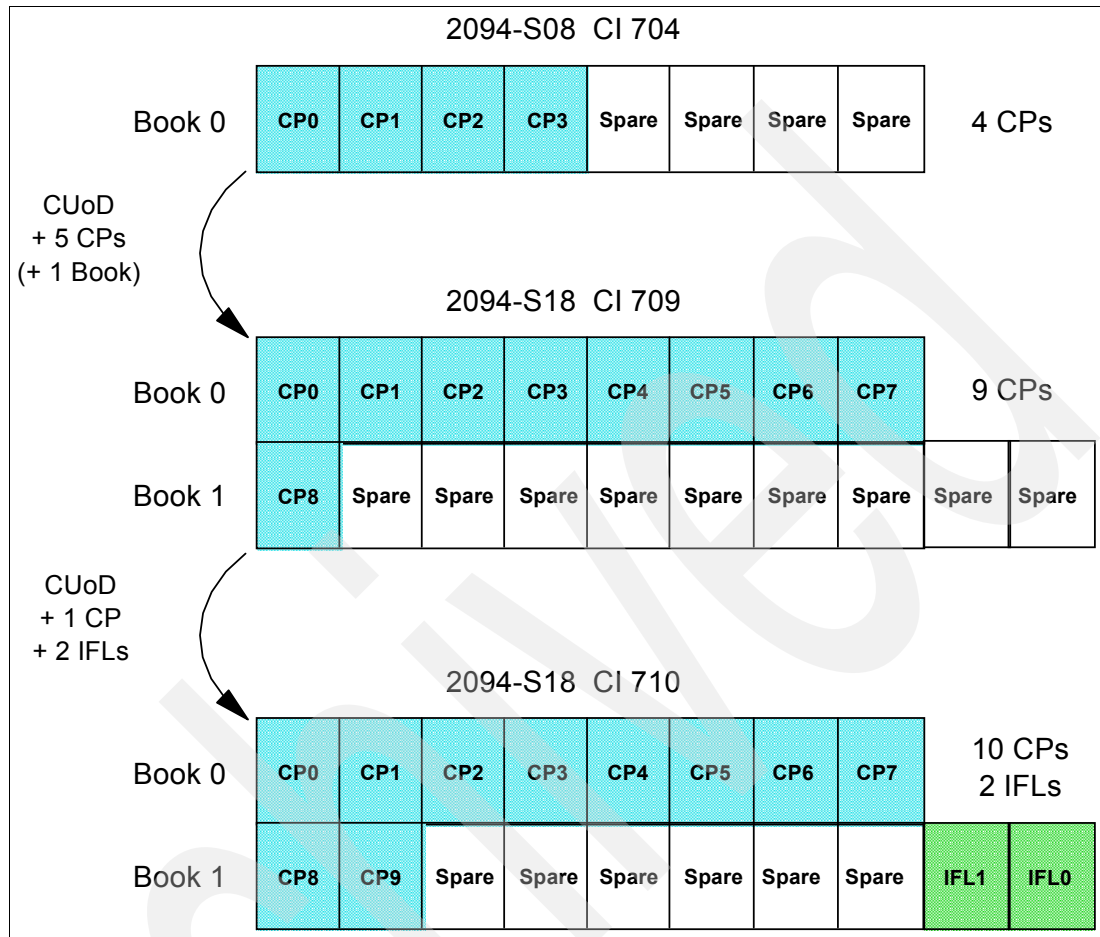


Figure 8-1 CUoD for processor example

An initial z9 EC Model S08 (one book), model capacity identifier 704 (four CPs), is concurrently upgraded to a z9 EC Model S18 (two books), with model capacity identifier 709 (nine CPs). The model upgrade requires adding a book and assigning and activating five PUs as CPs.

Then the z9 EC Model S18, capacity identifier 709, is concurrently upgraded to a capacity identifier 710 (10 CPs) with two IFLs by assigning and activating three more spare PUs (one as CP and two as IFLs).

Additional logical processors can be concurrently configured online to logical partitions by the operating system when reserved processors are defined in the image profile. The operating system must have the capability to concurrently configure more processors online.

If reserved CPs have not been defined to the logical partition, the logical partition will have to be deactivated, the image profile for that logical partition changed, and the logical partition reactivated to allow the additional CP resources to be available to the operating system. While these actions do not require a Power On Reset (POR), they are disruptive to the logical partition(s) requiring the change.

**Attention:** Up to 54 logical processors, including reserved processors, can be defined to a logical partition. You should not define more processors to a logical partition than the target operating system supports. V1R6 supports up to 32 processors, as a combination of CPs, zAAPs, and zIIPs. z/VM V5R2 supports up to 24 processors, which can be either all CPs or all IFLs. z/VM V5R3 extends this support to 32 PUs.

Software charges based on the total capacity of the server on which the software is installed would be adjusted to the maximum capacity after the CUoD upgrade.

Software products using Workload License Charge (WLC) may not be affected by the server upgrade, as their charges are based on partition utilization and not based on the server total capacity. Refer to 6.4.1, “Workload License Charges” on page 178 for more information about WLC.

### CUoD for memory

CUoD for memory can add, *concurrently*, more memory to a z9 EC by enabling, through LIC-CC, additional capacity up to the limit of the current installed memory cards, or by concurrently installing additional books and LIC-CC enabling memory capacity on the new books. If the z9 EC is a multiple book configuration, it may be possible to use the Enhanced Book Availability feature to remove a book and upgrade the memory cards to larger size and then LIC-CC enable the additional memory.

The memory card sizes on the z9 EC are 4, 8, or 16 GB, and each book has eight memory slots that are added in groups of four at a time. All memory cards in a book must have the same storage capacity.

**Note:** Upgrades requiring memory card changes can be concurrent using the Enhanced Book Availability feature. Planning is required to see if this is a viable option in your configuration. The use of the flexible memory option (FC 2802 through FC 2824) is the safest way to ensure EBA can work with the least disruption.

Table 2-2 on page 33 lists the range of system memory associated with a given memory card size and the number of memory cards for each server model. Figure 8-2 shows an example of CUoD for memory of a z9 EC Model S08 server with 48 GB of available memory.

The one-book z9 EC model has eight 8 GB memory cards, resulting in 64 GB of installed memory in total. Therefore, a concurrent memory upgrade within this Model S08 can be done up to the 64 GB limit through LIC-CC, but a memory upgrade to 80 GB would require the memory cards that are installed on the single book system to be replaced with eight 16 GB memory cards and it is *disruptive*.

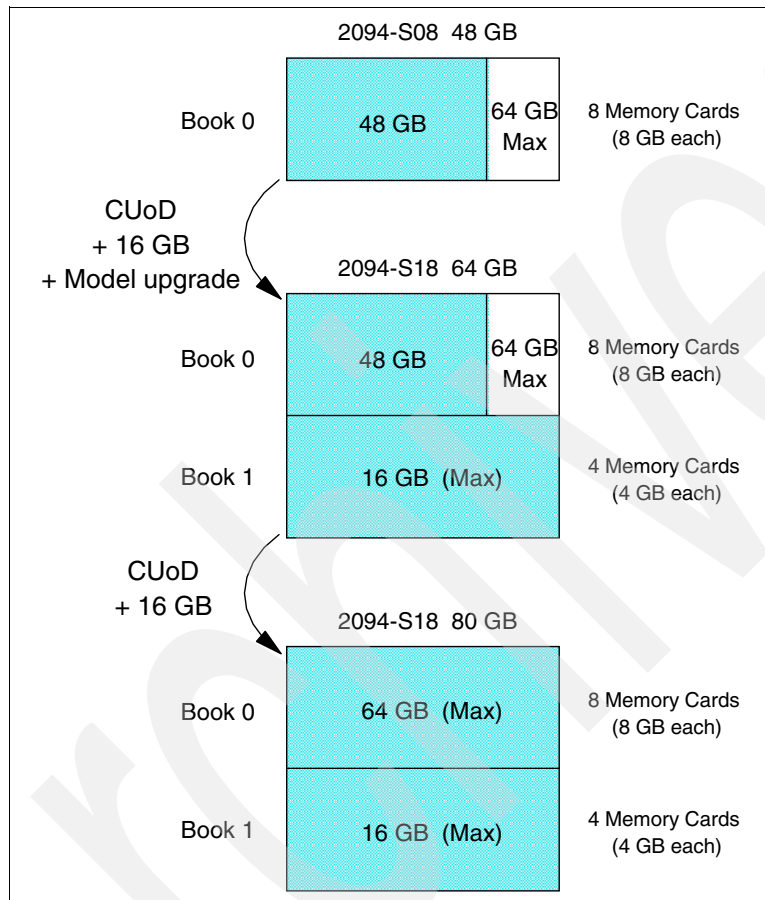


Figure 8-2 CUoD for memory example

However, as shown in the example, the upgrade of the Model S08 with 48 GB of memory to a Model S18 with 64 GB is *concurrent*, as the additional book comes with four memory cards (in this case, four 4 GB memory cards). The additional 16 GB memory capacity is enabled by LIC-CC on book 1.

In the last part of this example, the Model S18 is concurrently upgraded to 80 GB, by LIC-CC enabling all the installed memory.

A logical partition can dynamically take advantage of a memory upgrade if reserved storage has been defined to that logical partition. The reserved storage is defined to the logical partition as part of the image profile. Reserved memory can be configured online to the logical partition using the LPAR Dynamic Storage Reconfiguration (DSR) function. DSR allows a z/OS operating system image to add reserved storage to its configuration if any unused storage exists. If reserved storage has not been defined to the logical partition, the logical partition will have to be deactivated, the image profile changed, and the logical partition

reactivated to allow the additional storage resources to be available to the operating system image.

Concurrent memory upgrades also require that the memory must not be running in degraded mode.

### CUoD for I/O

CUoD for I/O can add, *concurrently*, more I/O ports to a z9 EC by either:

- ▶ Enabling additional ports on the already installed I/O cards through LIC-CC
  - LIC-CC-only upgrades can be done for ESCON channels and ISC-3 links, activating ports on the existing 16-port ESCON or ISC-3 daughter (ISC-D) cards.
- ▶ Installing additional I/O cards on an already installed I/O cage's slots
  - The installed I/O cages must provide the number of I/O slots required by the target configuration.

**Note:** I/O cages *cannot* be installed concurrently.

Figure 8-3 shows an example of CUoD for I/O through LIC-CC.

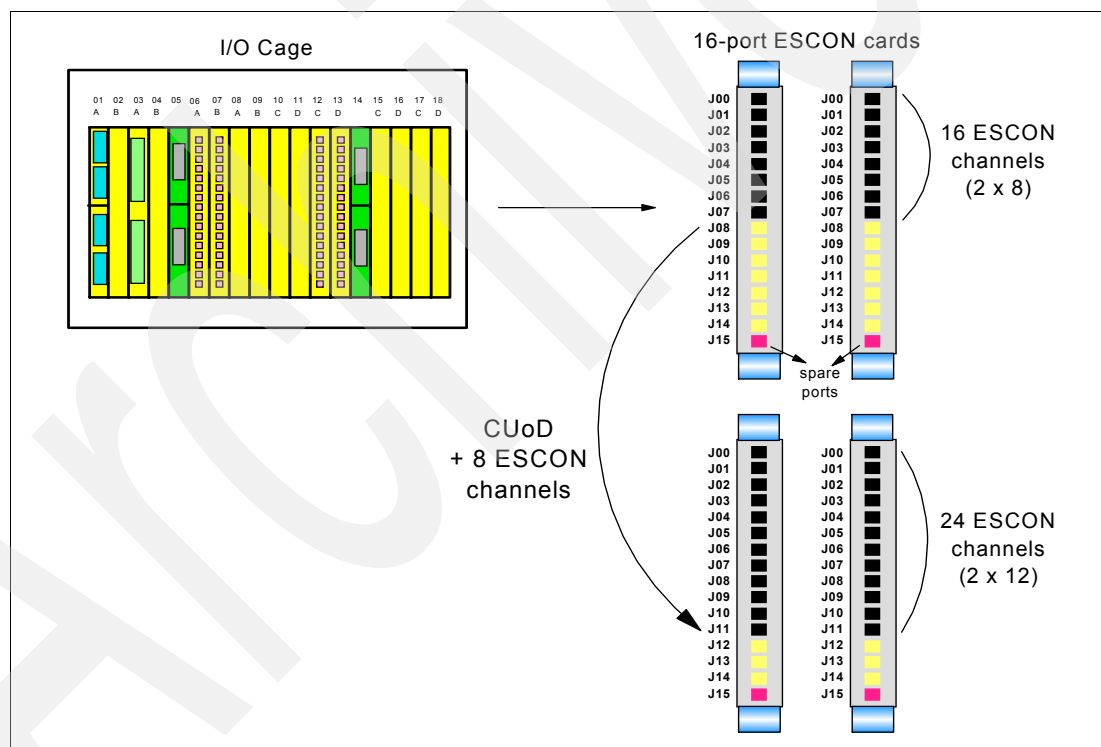


Figure 8-3 CUoD for I/O LIC-CC upgrade example

A z9 EC has 16 ESCON channels available, on two 16-port ESCON channel cards installed in an I/O cage. Each channel card has eight ports enabled. In this example, eight additional ESCON channels are concurrently added to the configuration by enabling, through LIC-CC, four unused ports on each ESCON channel card.

The additional channels installed concurrently to the hardware can also be concurrently defined in HSA and to an operating system using the Dynamic I/O configuration function. Dynamic I/O configuration can be used by z/OS or z/VM operating systems.

VSE, z/VSE, TPF, z/TPF, Linux, and CFCC do *not* provide Dynamic I/O configuration support. The installation of the new hardware is performed concurrently, but defining the new hardware in HSA and to the operating system requires an IPL.

To better exploit the CUoD for I/O capability, an initial configuration should be carefully planned to allow concurrent upgrades up to a target configuration. Plan Ahead concurrent conditioning process can include, in the initial configuration, the shipment of additional I/O cages required for future I/O upgrades.

### **Plan Ahead concurrent conditioning**

Concurrent Conditioning (FC 1999) and Control for Plan Ahead (FC 1995) features, together with the input of a future target configuration, allow upgrades to exploit the order process configurator for concurrent I/O upgrades at some future time.

The Plan Ahead feature identifies the content of the target configuration, which cannot be concurrently installed, avoiding any down time associated with feature installation. As a result, Concurrent Conditioning may include, in the initial order, additional I/O cages to support the future I/O requirements.

Accurate planning and definition of the target configuration is vital to maximize the value of this feature.

## **8.2.2 Customer Initiated Upgrade (CIU)**

Customer Initiated Upgrade (CIU) is the capability for the z9 EC *user* to initiate a *permanent* upgrade for CPs, ICFs, IFLs, zAAPs, zIIPs, or memory through the Web, using IBM Resource Link. CIU is similar to CUoD, but the additional resources are added by the customer. The customer also has the ability to unassign previously purchased CPs and IFLs processors through CIU.

The use of CIU requires that the CIU Enablement feature (FC 9898) be installed.

The CIU functions must be set up ahead of time. As part of the setup, the customer will need to register for Resource Link IDs. The Resource Link IDs will provide the customer access to the site where they will be able to configure and order their upgrades. In addition, the Web site contains customer education for use of the CIU and On/Off CoD.

After the order is placed and the customer receives notice that the order is ready to download, he or she will be able to download and apply the upgrade using functions available through the HMC, along with the Remote Support Facility. Once all the prerequisites are in place, the whole process from ordering to activation of the upgrade is performed by the customer. The actual upgrade process is fully automated and does not require any onsite presence of IBM service personnel.

CIU supports LIC-CC upgrades only; it does not support I/O upgrades. All additional capacity required by a CIU upgrade must be previously installed. Additional books or I/O cards cannot be installed through CIU.

The sum of CPs, unassigned CPs, IFLs, unassigned IFLs, ICFs, zAAPs and zIIPs cannot exceed the PU count of the installed books. The total number of zAAPs, or zIIPs, cannot exceed the number of owned CPs. The combined total of zAAPs and zIIPs cannot exceed twice the number of owned CPs.

**Important:** CIU for processors cannot be completed when CBU or On/Off CoD is *activated* on a z9 EC. In this case, the CIU for processors can be ordered and retrieved but *cannot* be applied until the temporary capacity upgrade through CBU or On/Off CoD is deactivated.

CIU may change the server model capacity identifier 4xx, 5xx, 6xx, or 7xx if additional CPs are requested or the capacity identifier is changed as part of the CIU, but it cannot change the *server* model, 2094-Sxx.

Additional logical processors can be concurrently configured online to logical partitions by the operating system when reserved processors are previously defined in the image profile for the logical partition. The operating system must have the capability to concurrently configure more processors online.

**Note:** CIU for processors can provide a *physical* concurrent upgrade, resulting in more enabled processors available to a server configuration. Thus, additional planning and tasks are required for *nondisruptive* logical upgrades. See “Recommendations to avoid disruptive upgrades” on page 246 for more information.

Software charges based on the total capacity of the server on which the software is installed are adjusted to the new capacity in place after the CIU upgrade. Software products using Workload License Charge (WLC) may not be affected by the server upgrade, as their charges are based on a logical partition utilization and not based on the server total capacity. See 6.4.1, “Workload License Charges” on page 178 for more information about WLC.

### **CIU registration and agreed contract for CIU**

To be able to use the CIU function, a customer has to be registered and the system set up. Once the CIU registration has been completed, the customer can access the CIU application through the Resource Link Web site. As part of the setup, the customer will provide one Resource Link ID for configuring and ordering CIU orders and, if required, a second ID as an approver. The IDs will be set up for access to the CIU support.

Using the CIU can provide benefits to the customer by allowing upgrades to be ordered and delivered much faster than the normal MES process.

Ordering and activation of the upgrade is accomplished by the customer logging on to the IBM Resource Link Web site and invoking the CIU application to upgrade a server for CPs, ICFs, IFLs, zAAPs, zIIPs, or memory. It is possible to request a customer order approval to conform to customer operation policies.

The Resource Link Web site also contains an education module for the CIU application. The customer can allow define additional IDs to be authorized to access CIU. Additional IDs can be authorized to enter or approve CIU orders, or only view existing orders.

Figure 8-4 illustrates the CIU ordering process on the IBM Resource Link Web site.

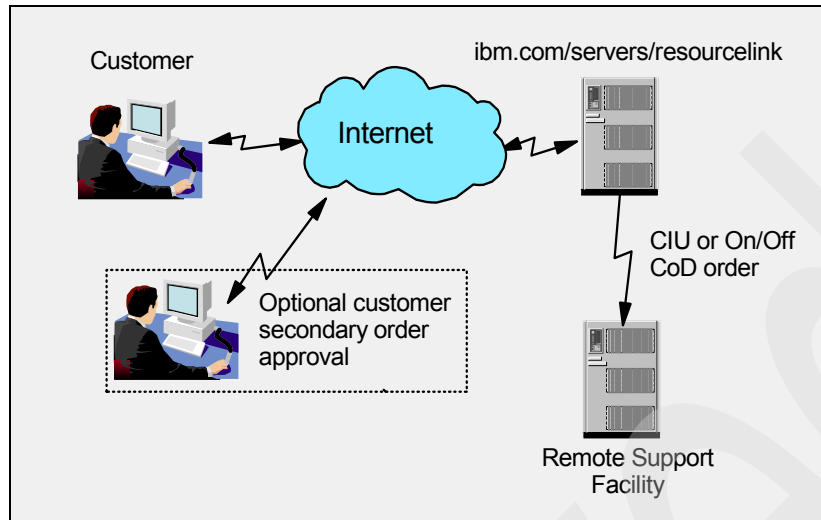


Figure 8-4 CIU ordering example

The following is a sample list of the windows a customer must follow on Resource Link to initiate an order:

1. Sign on to Resource Link.
2. Select the **CIU** option from the main Resource Link page.
3. Customer and server details associated with the user ID are listed.
4. The current configuration (PU allocation and memory) is shown for the selected server serial number.
5. Create a target configuration step-by-step for each upgradable option. Resource Link limits options to those that are valid/possible for this z9 EC configuration.
6. The target configuration is verified.
7. The customer has the option to accept or reject.
8. An order is created and verified against the pre-established agreement.
9. A price is quoted for the order; customer signals acceptance/rejection.
10. A customer secondary order approval is optional.
11. On confirmation of acceptance, the order is processed.
12. LIC-CC for the upgrade should be available within hours.



Figure 8-5 shows the CIU activation process. IBM Resource Link communicates with the Remote Support Facility to stage the CIU order and prepare it for download. The customer is automatically notified when the order is ready for download.

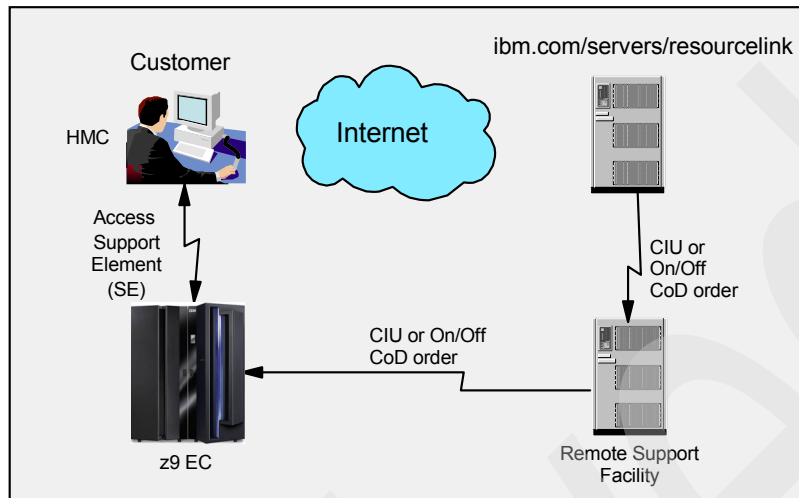


Figure 8-5 CIU activation example

### Order and fulfillment process

The CIU process allows the customer to order increased capacity for CPs, ICFs, IFLs, zAAPs, zIIPs, or memory. Resource Link is responsible for delivering the price or lease agreement to the customer. The interface handles the order differently based on whether the customer is leasing the server. The customer profile associated with the machine serial number will contain an indicator that Resource Link uses to make the determination. If the customer chooses to accept this agreement, then it will be forwarded to the correct billing system. Only Resource Link users who accept this feature will be able to access the CIU application.

The two major components in the process are *Ordering* and *Activation*.

#### Ordering

Resource Link provides the interface that allows the customer to order a dynamic upgrade for a specific server. The customer is able to create, cancel, and view the order. The customer also is able to view the history of orders that were placed through this interface. Configuration rules enforce only valid configurations being generated within the limits of the individual server. Warning messages are issued when invalid upgrade options are selected. The CIU application will allow only one order per server to be placed at a time.

Figure 8-6 shows a CIU order example.

The screenshot displays the IBM Machine profile interface. At the top, there is a navigation bar with the IBM logo, a search bar, and links for Home, Products, Services & solutions, Support & downloads, and My account. The main content area is titled "Machine profile" for machine 2094 - SCRNI - 1234567. It features a "Resource Link" sidebar on the left and a central table comparing "Current configuration" and "Ordered configuration".

	Current configuration	Ordered configuration
<b>Model Capacity:</b>	708 (8 CPs)	712 (12 CPs)
<b>ICF:</b>	0	0
<b>zAAP:</b>	2	2
<b>IFL:</b>	2	2
<b>SAP:</b>	4	4
<b>Memory:</b>	32	32
<b>CBU Capacity:</b>	712 (12 CPs)	714 (14 CPs)
<b>Unassigned IFLs:</b>	0	0

Additional sections include "Machine summary" (Type, serial no., System name, Model capacity downgraded from), "Customer summary" (Company name, Customer number, GEO, country), "About ordering upgrades" (Authorization to order and approve orders, Notes), and "Open orders" table.

Order number	Order summary	Date ordered	Order status
LT6F6Q35	Permanent upgrade 708 (8 CPs) to 712 (12 CPs)	08/11/2005 02:49:53 PM	Staging order

Figure 8-6 CIU order example

The number of CPs, ICFs, zAAPs, zIIPs, IFLs, SAPs, memory size, CBU features, unassigned CPs, and unassigned IFLs on the current configuration are displayed on the left side. On the right side are the corresponding updated values of the ordered configuration. This CIU example requests an upgrade from eight CPs (model capacity identifier 708) to twelve CPs (model capacity identifier 712) plus two additional CBU CPs.

Resource Link retrieves and stores relevant data associated with the processor configuration, like the number of CPs and installed memory cards. It allows you to select only those upgrade options that are deemed valid by the order process; it only allows upgrades within the bounds of the currently installed hardware.

### Activation

Once an order is placed, Resource Link dynamically enables the appropriate LIC-CC records and make them available to download.

When the order is available for download, the customer is sent a note containing an activation number for the order. The order can then be downloaded. To download the order, use any of the Hardware Management Consoles (HMC) attached to the system to be upgraded and perform a Single Object Operation into the Support Element (SE). From the SE, you can select the **Perform Model Conversion** task. Using the Model Conversion window, select the **CIU Options** to start the process; see Figure 8-7.

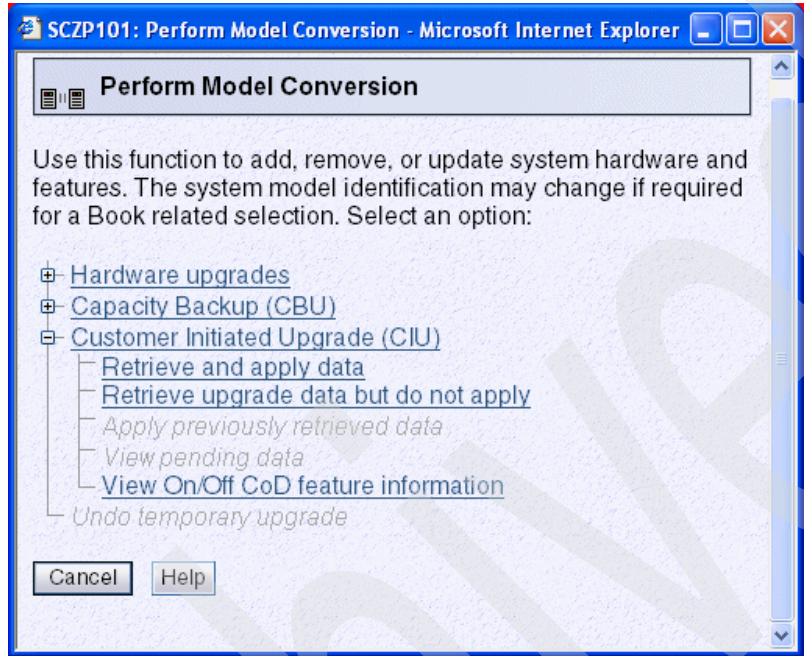


Figure 8-7 z9 EC Model Conversion window

The window provides several possible options:

- ▶ Retrieve and apply data.
- ▶ Retrieve upgrade data but do not apply.
- ▶ Apply previously retrieved upgrade.
- ▶ View pending data.

Selecting the **Retrieve and apply data** option prompts the customer to enter the order activation number to begin the code download process; see Figure 8-8. Once downloaded, the system will check if the upgrades can be applied.

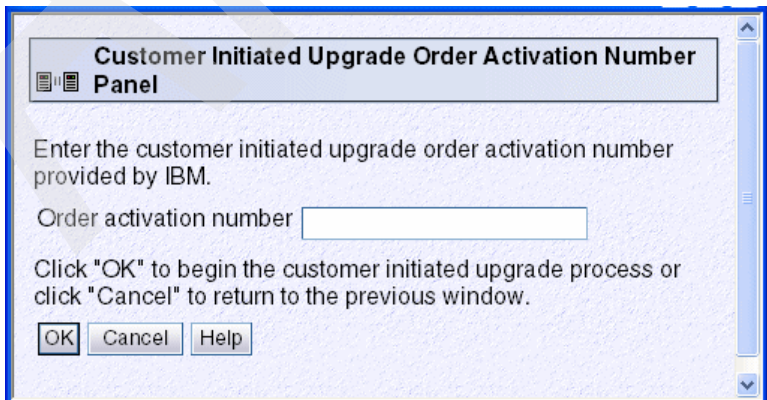


Figure 8-8 CIU upgrade selection window

A CIU upgrade for processors cannot be applied while On/Off CoD or CBU are active. In those cases, the requested upgrade can be retrieved, but cannot be applied until the On/Off CoD or CBU is deactivated.

### 8.2.3 On/Off Capacity on Demand (On/Off CoD)

The On/Off Capacity on Demand for z9 EC is the ability for the z9 EC installation to *temporarily* turn on unowned PUs, unassigned CPs, and unassigned IFLs available within the current model, or to change CI for CPs to help meet customers' peak workload requirements. On/Off CoD uses the Customer Initiated Upgrade (CIU) process to request the upgrade through the Web, using IBM Resource Link.

On/Off CoD requires the CIU Enablement feature (FC 9898) and the On/Off CoD Enablement feature (FC 9896) installed.

**Important:** The On/Off CoD capability can coexist with Capacity BackUp (CBU) enablement. Both On/Off CoD and CBU LIC-CC can be installed on a z9 EC, but the On/Off CoD activation and CBU activation are mutually exclusive.

The resources eligible for temporary use are CPs, ICFs, IFLs, zAAPs, and zIIPs. Temporary addition of memory and I/O ports is not supported. Spare PUs that are on the installed books can be temporarily and concurrently activated as CPs, ICFs, IFLs, zAAPs, or zIIPs through LIC-CC, up to double the current installed capacity, and up to the limits of the physical server size.

This means that On/Off CoD upgrade cannot change the *server* model 2094-Sxx; addition of new books is not supported. However, On/Off CoD may change the model capacity identifier 4xx, 5xx, 6xx, or 7xx.

The On/Off CoD upgrade features are:

- ▶ On/Off CoD Active CP Day (FC 9897)
- ▶ On/Off CoD Active IFL Day (FC 9888)
- ▶ On/Off CoD Active ICF Day (FC 9889)
- ▶ On/Off CoD Active zAAP Day (FC 9893)
- ▶ On/Off CoD Active zIIP Day (FC 9908)

You may concurrently install temporary capacity by ordering On/Off CoD as follows:

- ▶ On/Off CoD CP features equal to the capacity of installed CPs
- ▶ On/Off CoD IFL features up to the number of installed IFLs
- ▶ On/Off CoD ICF features up to the number of installed ICFs
- ▶ On/Off CoD zAAP features up to the number of installed zAAPs
- ▶ On/Off CoD zIIPs features up to the number of installed zIIPs

On/Off CoD can provide temporary capacity in two ways.

- ▶ By increasing the number of CPs.
- ▶ For subcapacity models, capacity can be added by increasing the number of CPs, by changing the capacity identifier of the CPs, or both. The CI for all CPs must be the same. If the On/Off CoD is adding CP resources that have a CI different than the installed CPs, then the base CPs will be changed to match.

On/Off CoD has limits associated with its use.

- The number of CPs cannot be reduced.
- The target configuration capacity is limited to twice the currently installed one for each individual processor type (CPs, IFLs, ICFs, zAAPs, and zIIPs). Table 8-2 shows the valid On/Off CoD configurations for CPs.

For example, for a z9 EC with capacity identifier 402, there would be two ways to deliver a capacity upgrade through on/off CoD. The first option is to add CPs of the same capacity setting. With this option, the capacity identifier could be changed to a 403, which would add one additional CP (making a 3-core) or to a 404, which would add two additional CPs (making a 4-core). The second option would be to change to a different capacity setting of the current CPs and increase the capacity identifier to a 502; the capacity setting of the CPs is increased but no additional CPs are added.

It is recommended that you use the Large System Performance Reference information to evaluate the capacity requirements according to your workload type. LSPR data for current IBM processors is available at this URL:

<http://www.ibm.com/servers/eserver/zseries/lspr/>

Table 8-2 Valid On/Off CoD upgrades for granular capacity models

Capacity identifier	On/Off CoD CP4	On/Off CoD CP5	On/Off CoD CP6	On/Off CoD CP7
401	402	501		
402	403, 404	502		
403	404, 405,406	503		
404	405 - 408	504		
405	406, 407, 408	505		
406	407, 408	506		
407	408	507		
408		508		
501		502	601	701
502		503, 504	602, 603	702, 703
503		504, 505, 506	603, 604, 605	703, 704
504		505 - 508	604 - 607	704, 705
505		506, 507, 508	605 - 608	705, 706
506		507, 508	606, 607, 608	706, 707, 708
507		508	607, 608	707, 708, 709
508			608	708 - 711
601		502	602	701
602		503, 504	603, 604	702, 703
603		505, 505, 506	604, 605, 606	703, 704, 705

Capacity identifier	On/Off CoD CP4	On/Off CoD CP5	On/Off CoD CP6	On/Off CoD CP7
604		505 - 508	605 - 608	704, 705, 706
605		507, 508	606, 607, 608	705 - 708
606		508	607, 608	706 - 710
607			608	707 - 712
608				708 - 714
701		502, 503	602	702
702		504	603, 604	703, 704
703		505, 506	604, 605, 606	704, 705, 706
704		507, 508	606, 607, 608	705 - 708
705			607, 608	706 - 711
706			608	707 - 713
707 to 753				Up to two times the base capacity

The On/Off CoD capacity is charged on a 24-hour basis; there is a sixty minute grace period within a 24 hour On/Off CoD day. This allows up to an hour before and after the 24-hour billing period to either change the On/Off CoD configuration for the next 24-hour billing period or deactivate the current On/Off CoD configuration without affecting the billing of the prior 24-hour period. The times when the capacity is activated and deactivated is maintained in the z9 EC and sent back to the support systems.

When the temporary capacity is no longer required, its removal is nondisruptive. While On/Off CoD is activated on a z9 EC, other permanent hardware upgrades or MES are restricted. With the exception of memory and channels, LIC-CC enabled features, such as CPs, ICFs, IFLs, zAAPs, and zIIPs can be ordered but not enabled until the On/Off CoD upgrade is deactivated.

If On/Off capacity is already active, additional On/Off capacity can be added without having to return the z9 EC to its purchase capacity. If the capacity is increased multiple times within a 24 hour period, the charges apply to the highest amount of capacity active in the period.

Additional logical processors can be concurrently configured online to logical partitions by the operating system when reserved processors are previously defined. The operating system must have the capability to concurrently configure more processors online.

**Note:** On/Off CoD provides a *physical* concurrent upgrade, resulting in more enabled processors available to a server configuration. Thus, additional planning and tasks are required for *nondisruptive* logical upgrades. See “Recommendations to avoid disruptive upgrades” on page 246.

To participate in this offering, customers must have accepted contractual terms for On/Off CoD (in addition to Customer Initiated Upgrade (CIU)), established a CIU profile, and installed an On/Off CoD *right to use* feature on the server. Subsequently, the customer may concurrently install temporary capacity up to the limits in On/Off CoD and use it for an indeterminate time. Monitoring will occur through the server call home facility and an invoice

will be generated if the capacity has been enabled during the calendar month. The customer will continue to be billed for use of temporary capacity until the server is returned to the original configuration. If the On/Off CoD support is no longer needed, the enablement code will need to be removed.

On/Off CoD orders can be pre-staged in Resource Link to allow multiple optional configurations. The pricing of the orders is done at the time of the order, and the pricing can vary from quarter to quarter. There can be staged orders with different pricing. When the order is downloaded and activated, the daily costs is based on the pricing at the time of the order. The staged orders do not have to be installed in order sequence. If a staged order is installed out of sequence and later an order that was staged that had a higher price is downloaded, the daily cost will be based on the lower price.

It is also possible to store up to one hundred On/Off CoD LICCC records on the Support Element with the same or different capacities at any given time, giving greater flexibility to quickly enable needed temporary capacity. Each record is easily identified with descriptive names, and users can select from a list of records that can be activated.

Before a customer can order On/Off CoD, there must be a signed agreement for the Customer Initiated Upgrade (CIU) facility. After signing a contract, an order is placed through CIU to install an On/Off CoD right to use feature. Once installed, the customer is free to order and activate temporary capacity.

### **Ordering**

Resource Link provides the interface that allows the customer to order a dynamic upgrade for a specific server. The customer is able to create, cancel, and view the order. Configuration rules are enforced, and only valid configurations are generated based on the configuration of the individual server. Once the prerequisites have been completed, orders for the On/Off CoD can be placed. The order process is similar to the CIU order process and uses the CIU process on Resource Link.

The customer can order temporary capacity for CPs, ICFs, IFLs, zAAPs, or zIIPs. Memory and channels are not supported on On/Off CoD. The amount of capacity is based on the amount of owned capacity for the different types of resources. A LIC-CC record is established and staged to RETAIN® for this order. The record, once activated, has no expiration date; however, an individual record can only be activated once. Subsequent sessions will require a new order to be generated producing a new LIC-CC record for that specific order.

### **On/Off CoD Testing**

There are two methods provided for testing the On/Off CoD function: a full function test and an administrative test. Each z9 EC that has On/Off CoD enabled has one full function test available and unlimited administrative tests. The administrative test has “zero” capacity associated with it. There are no additional charges for these tests.

The full function test allows for an On/Off CoD order that contains additional capacity to be placed, the order to be downloaded, and the capacity to be installed on the z9 EC. This allows the complete process to be validated. During the test, the process of ordering the temporary capacity, the notification process, the download, and the installation and activation is validated. In addition, the customer’s operations and support can validate their procedures for enabling the additional capacity. The enabling is more than just downloading and activating; there may be need to vary on CPs to logical partitions, change image profiles, or change weights on a logical partition to allow the additional resources to be fully tested. The full function test allows these processes to be tested and verified. The full function test can be enabled for up to 24 hours. If the support is enabled longer than the 24 hours, it will be billed as a normal On/Off CoD order. One full function test is provided.

The zero capacity test provides the ability to place orders, download, and activate On/Off Capacity orders for testing and operation training purposes. The test, which contains zero capacity, can be used as part of the setup of the process for On/Off CoD. During the planning phase of implementing On/Off CoD, this test would allow for testing of the procedures that will be used in the production environment as well as training operations and support staff in the use of the facilities. Once the process and procedures have been established, the zero capacity test can be used for ongoing training. The zero capacity test has no limit as to how many can be ordered or how long the orders can be active on the server.

Figure 8-9 shows a Resource Link Web page that displays an On/Off CoD order example.

The screenshot shows a web browser window titled "IBM Resource Link: Create machine upgrade order - Microsoft Internet...". The main heading is "Create On/Off Capacity on Demand order".

Customer and Machine Information:

- Customer number: 1234567
- Order number: LT6F6QQC
- Machine type: 2094
- Machine serial: SCR2

	Current configuration	Upgrade configuration	Upgrade price
Model Capacity:	708 (8 CPs)	709 (9 CPs) [dropdown]	\$0.00
ICF:	0	0 [dropdown]	\$0.00
zAAP:	2	[dropdown]	\$0.00
IFL:	2	2 [dropdown]	\$0.00
<b>Total daily price:</b>			\$0.00

Buttons: Submit, Cancel

Footer: Terms of use, Privacy, Close [x]

Figure 8-9 On/Off CoD order example

The order in Figure 8-9 is a On/Off CoD order for one CP. The maximum number of CPs, ICFs, IFLs, zAAPs, and zIIPs is limited by the current number of available unused PUs of the installed books. On/Off CoD can be accessed from the Resource Link Web site at:

<http://www.ibm.com/servers/resourceLink>

### Activation/Deactivation

When a previously ordered On/Off CoD is retrieved from Resource Link, it is downloaded and immediately activated. You cannot download the order and defer the installation. The process for downloading and activating the order uses the same facility as used for CIU.

When the customer has finished using temporary capacity, they must take an action to deactivate the temporary capacity. This deactivation uses the same facility as CBU called "undo temporary upgrade", accomplished from the Support Element. The deactivation is nondisruptive. Depending on how the additional capacity was added to the logical partitions, customers may be required to perform tasks at the logical partition level in order to remove the temporary capacity; for example, configure offline CPs that had been added to the partition.



On/Off CoD orders can be staged in Resource link so that multiple orders are available. A order can only be downloaded and activated one time. If a different On/Off CoD order is required, it can be downloaded and activated without having to restore the system to its original purchased capacity.

In support of automation, an API is provided that allows the activation of the On/Off CoD. The activation is performed from the HMC and requires entering of the order number. With this API, automation code can be used to send an activation command along with the order number to the HMC to enable the order.

### ***Termination***

A customer will be contractually required to terminate the On/Off CoD right to use feature whenever there is a transfer in asset ownership. A customer may also choose to terminate the On/Off CoD right to use feature without transferring ownership. Application of feature code 9898 will terminate the right to use On/Off CoD. This feature cannot be ordered if a temporary session is already active. Similarly, CIU cannot be removed if a temporary session is active. Anytime CIU is removed, the On/Off CoD right to use will be simultaneously removed. Reactivating the right to use feature will subject the customer to whatever terms and fees apply at that time.

### ***Upgrade Capability during On/Off CoD***

No upgrades involving physical hardware will be supported while an On/Off CoD upgrade is active on a particular z9 EC. LIC-CC only upgrades can be ordered and retrieved from Resource Link but not applied while an On/Off CoD upgrade is active. LIC-CC only memory upgrades can be retrieved and applied while a On/Off CoD upgrade is active.

### ***Repair capability during On/Off CoD***

If the z9 EC requires service while an On/Off CoD upgrade is active, the repair and verify code (R&V) will automatically deactivate the On/Off CoD upgrade. At the end of the repair, R&V will retrieve a new LIC-CC record from RETAIN to replace the record that was deactivated. The On/Off CoD upgrade will be activated to the state prior to R&V, including restoration of the original activation date.

### ***Monitoring***

When the customer activates an On/Off CoD upgrade, an indicator is set in Vital Product Data. This indicator is part of the call home data transmission, which is sent on a scheduled basis. A time stamp is placed into call home data when the facility is deactivated. At the end of each calendar month, the data will be used to generate an invoice for the On/Off CoD used during that month.

### ***Software***

Software PSLC customers will be billed at the MSU level represented by the combined permanent and temporary capacity. All PSLC products will be billed at the peak MSUs enabled during the month, regardless of usage. Customers with WLC licenses will be billed by product at the highest four-hour rolling average for the month. In this instance, temporary capacity will not necessarily increase the customer's software bill until that capacity is allocated to logical partitions and actually consumed.

Results from the STSI instruction will reflect the current permanent plus temporary CPs. See "STSI Store System Information instruction" on page 181 for more details.

## 8.2.4 Capacity BackUp (CBU)

Capacity BackUp (CBU) is offered with the z9 EC to provide reserved emergency backup processor capacity for unplanned situations where customers have lost capacity in another part of their establishment and want to recover by adding the reserved capacity on a designated z9 EC.

CBU is the quick, *temporary* activation of PUs, up to 90 days, in the face of a loss of customer processing capacity due to an emergency or disaster/recovery situation.

**Note:** CBU is for disaster/recovery purposes only and *cannot* be used for peak workload management.

**Important:** The CBU capability can coexist with On/Off CoD enablement. Both CBU and On/Off CoD LIC-CC can be installed on a z9 EC, but the CBU activation and On/Off CoD activation are mutually exclusive.

The CBU process allows for CBU to activate CPs, IFLs, ICFs, zAAPs, and zIIPs. You must order the quantity and type of PU you require. The feature codes are:

- ▶ For CPs
  - FC 7817 for CBU CP4
  - FC 7818 for CBU CP5
  - FC 7819 for CBU CP6
  - FC 7820 for CBU CP7
- ▶ FC 7821 for CBU IFL
- ▶ FC 7822 for CBU ICF
- ▶ FC 7824 for CBU zAAP
- ▶ FC 7825 for CBU zIIP

When the CBU activated capacity is no longer required, its removal is nondisruptive. If CBU is activated on a z9 EC, other hardware upgrades/MES are restricted. With the exception of memory and channels, LIC-CC enabled features, such as CPs, ICFs, IFLs, zAAPs, and zIIPs, can be ordered but not enabled until the CBU upgrade is deactivated.

The CPs that can be activated by CBU come from the available spare PUs on any installed book of the z9 EC. The number of CBU features that can be ordered is limited by the number of spare PUs on the server. For example:

- ▶ A z9 EC Model S18 server with eight CPs, no IFLs, ICFs, or zAAPs, has ten spare PUs available; it can have up to ten CBU features.
- ▶ A z9 EC Model S28 server with 12 CPs, four IFLs, and one ICFs has nine spare PUs available; it can have up to nine CBU features.

Sub-capacity makes a difference in the way the CBU features are done. On the standard models, the CBU features indicate the amount of additional capacity needed. If the amount of needed CBU capacity is equal to four CPs, then the CBU configuration would be four CBU CPs with the same capacity setting.

The sub-capacity models have multiple capacity settings 4xx, 5xx, 6xx, or 7xx. The capacity setting of all the sub-capacity CPs must be the same. The number of CBU CPs must be equal to, or greater, than the number of CPs in the base configuration, and all the CPs in the CBU configuration must have the same capacity setting.

For example, if the base configuration is a 2-core CP4, providing a CBU configuration of a four way of the same capacity setting requires two CBU feature codes. If the required CBU capacity changes the capacity setting of the CPs, and you go from model capacity identifier 402 to a CBU configuration of a 4-core 504, would require four CBU feature codes with a capacity setting of 5xx.

If the capacity setting of the CPs is changed, it requires more CBU features, not more physical PUs. This mean that your CBU contract requires more CBU features if the capacity setting of the CPs is changed.

Note that CBU can add CPs through LIC-CC only and the z9 EC must have the proper number of books installed to allow the required upgrade. CBU can change the model capacity identifier 4xx, 5xx, 6xx, or 7xx, but does not change the *server* model 2094-Sxx.

A CBU contract must be in place before the special code that enables this capability can be loaded on the server. CBU features can be added to an existing z9 EC nondisruptively.

The installation of the CBU code provides an alternate configuration that can be activated in the face of an actual emergency. Five CBU tests, lasting up to 10 days each, and one CBU activation, lasting up to 90 days for a real disaster/recovery, are typically allowed in a CBU contract.

A CBU system normally operates with a *base* PU configuration having a pre-configured number of additional spare PUs reserved for activation in case of an emergency. One CBU feature of the correct type (CP, IFL, ICF, zAAP, or zIIPs) is required for each *stand-by* PU type that can be activated. CBU activation enables the total number of CBU features installed.

The base server configuration must have sufficient memory and channels to accommodate the potential needs of the large CBU target server. When capacity is needed in an emergency, the customer can activate the emergency CBU configuration with the reserved spare PUs added into the configuration as CP, ICF, IFL, zAAPs, or zIIPs. It is important to ensure that all required functions and resources are available on the backup servers, including CF LEVELs for Coupling Facility partitions, memory, and cryptographic functions, as well as connectivity capabilities.

This upgraded configuration is activated *temporarily* and provides additional PUs above and beyond the server's original, *permanent* configuration. The number of additional PUs is predetermined by the alternate configuration, which has been stated in the CBU contract.

When the emergency is over (or the CBU test is complete), the server must be taken back to its original configuration. The CBU features can be deactivated by the customer at any time before the expiration date. Failure to deactivate the CBU feature before the expiration date will cause the system to slow down because the capacity is capped and will require a POR to return the system back to its original configuration.

**Note:** CBU for processors provides a physical concurrent upgrade, resulting in more enabled processors available to a server configuration. Thus, additional planning and tasks are required for *nondisruptive* logical upgrades. See "Recommendations to avoid disruptive upgrades" on page 246.

Software charges based on the total capacity of the server on which the software is installed would be adjusted to the maximum capacity after the CBU upgrade.

Software products using Workload License Charge (WLC) may not be affected by the server upgrade, as their charges are based on partition's utilization and not based on the server total capacity. See 6.4.1, "Workload License Charges" on page 178 for more information about WLC.

For detailed instructions, refer to the manual *System z Capacity on Demand User's Guide*, SC28-6846.

## **Activation/deactivation of CBU**

The activation and deactivation of the CBU function is a customer responsibility without the need for the on-site presence of IBM service personnel. The CBU function is activated and deactivated from the HMC, and in each case it is a nondisruptive task.

### ***CBU activation***

CBU is activated from the SE using the "Perform Model Conversion" task. In case of real disaster, use the Activate CBU option to activate the 90 day period. It uses RSF to trigger an automatic verification of CBU authentication at IBM. This will initiate an automatic sending of the authentication to the server, automatic unlocking of the reserved capacity, and activation of the CBU resources.

In situations where the RSF cannot be used, CBU can be activated through a password window. In this case, a request by telephone to the IBM Support Center will retrieve the password.

The CBU activation cannot be done when an On/Off CoD upgrade is already activated.

### ***Image upgrades***

After the CBU activation, the z9 EC can have more capacity, more active PUs, or both. The additional resources go into the resource pools and are available to the logical partitions. If the logical partitions need to increase their share of the resources, the logical partition weight can be changed or the number of logical processors can be concurrently increased by configuring reserved processors online. The operating system must have the capability to concurrently configure more processors online.

### ***CBU deactivation***

To deactivate the CBU, the additional resources need to be released from the operating systems. In some cases, this is a matter of varying the resources offline. In other cases, it could mean shutting down operating systems or deactivating logical partitions. Once the resources have been released, the same facility on the SE is used to turn off CBU; to deactivate CBU, select the **Undo temporary upgrade** option from the Perform Model Conversion task on the SE.

### ***CBU testing***

Test CBUs are provided as part of the CBU contract. CBU is activated from the SE using the Perform Model Conversion task. Select the test option to initiate a 10 day test period. There are five of this type of test with the standard contract. Test CBU has a ten day limit and must be deactivated in the same way as the real CBU, using the same facility through the SE. If the deactivation is not done before the ten days are up, the server will be capped and performance will be impacted. Testing can be accomplished by ordering a diskette, calling the support center, or using the facilities on the SE.

## Capacity BackUp example

Figure 8-10 shows an example of a z9 EC Model S18 capacity identifier 704 to a z9 EC Model S18 capacity identifier 712 Capacity BackUp operation.

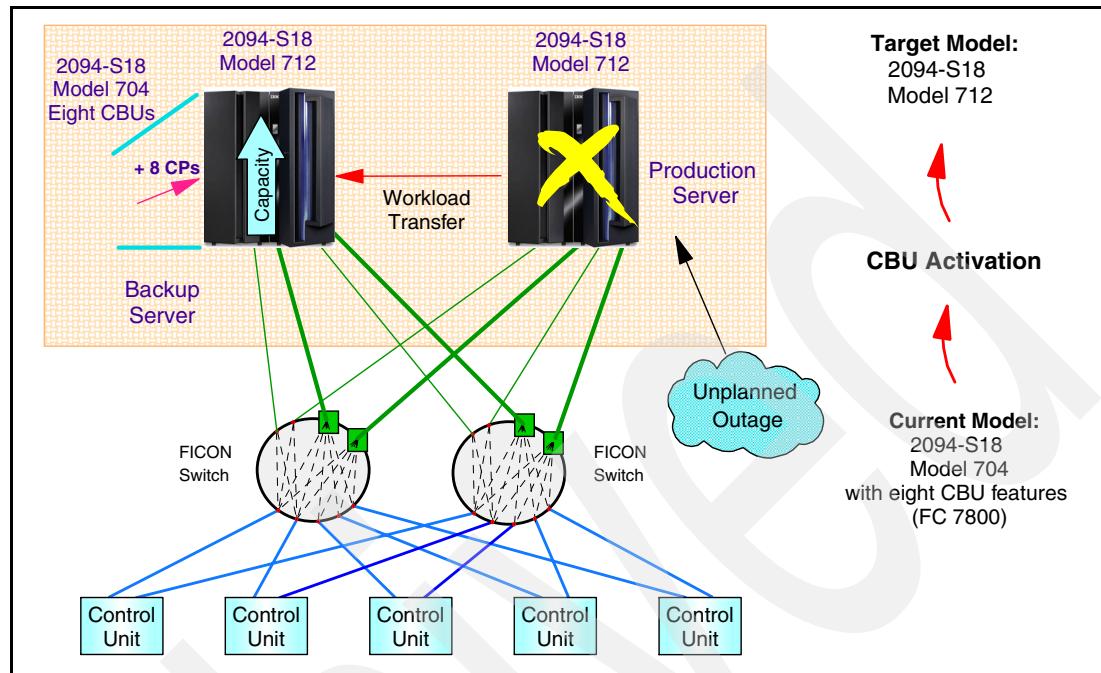


Figure 8-10 Capacity BackUp operation example

The PUs associated with Capacity BackUp are reserved for future use with CBU features (FC 7800) installed on the backup server. In this example, there should be eight CBU features installed on the backup z9 EC Model S18, model capacity identifier 704.

When the production z9 EC Model S18, capacity identifier 712 has an unplanned outage, the backup server can be temporarily upgraded to the target model planned, z9 EC Model S18, capacity identifier 712, to get the capacity to take over the workload on the failed production server.

Furthermore, customers can configure systems to back each other up. For example, if a customer uses two z9 EC models S08, capacity identifier 703 for the production environment, both can have three CBU features installed (or even more). If one server has a disaster, the other one can be upgraded up to approximately the total original CP capacity.

### Automatic CBU enablement for GDPS

The intent of the GDPS CBU is to enable automatic management of the reserved PUs provided by the CBU feature in the event of a server failure or a site failure. Upon detection of a site failure or planned disaster test, GDPS will concurrently add CPs to the servers in the take-over site to restore processing power for mission-critical production workloads. GDPS automation will:

- ▶ Perform the analysis required to determine the scope of the failure; this minimizes operator intervention and the potential for errors.
- ▶ Automate authentication and activation of the reserved CPs.
- ▶ Automatically restart the critical applications after reserved CP activation.
- ▶ Reduce the outage time to restart critical workloads from several hours to minutes.

## 8.3 Enhanced Book Availability (EBA)

The z9 EC is designed to allow a single book, in a multi-book server, to be concurrently removed from the server and reinstalled during an upgrade or repair action, while continuing to provide connectivity to the server I/O resources using a second path from a different book.

Enhanced Book Availability is an extension of the support for Concurrent Book Add (CBA) delivered on z990. With Enhanced Book Availability, and with proper planning to ensure that you still have all the resources available to run your critical applications in a (N-1) book configuration, you may be able to avoid planned outages. Consider also the selection of the flexible memory option (FC 2802 through 2824) to provide additional memory resources when replacing a book.

To minimize the impact on current workloads, you should ensure that you have sufficient inactive physical resources on the remaining books to complete a book removal. You may also consider non-critical system images that you are willing to give up, such as test or development logical partitions. After these non-critical logical partitions have been stopped and their resources freed, you may be able to find sufficient inactive resources to contain critical workloads while completing a book replacement.

It is recommended that the z9 EC be configured using guidance provided in the next section.

### 8.3.1 Planning consideration

In order to take advantage of the Enhanced Book Availability function, you need to configure enough physical memory and engines such that the loss of a single book does not result in any degradation to your critical workloads during:

- ▶ A degraded restart in the rare event of a book failure
- ▶ A book replacement for repair or physical memory upgrade

We recommend the following configurations that will enable exploitation of the Enhanced Book Availability function. Select PU/model configuration such that 100 percent of the customer-owned PUs can be activated even when one book within a model is fenced.

- ▶ A maximum of eight customer PUs are configurable on the S18.
- ▶ A maximum of 18 customer PUs are configurable on the S28.
- ▶ A maximum of 28 customer PUs are configurable on the S38.
- ▶ A maximum of 40 customer PUs are configurable on the S54.
- ▶ Requires no special feature codes for PU/model configuration.
- ▶ New Flexible Memory option: Deliver physical memory cards such that 100 percent of the purchased memory increment can be activated even when one book is fenced.

The main point here is that the server configuration should have sufficient *dormant* resources on the remaining books in the system for the *evacuation* of the book to be replaced or upgraded. Dormant resources could be:

- ▶ Unused PUs or memory not LIC-CC enabled
- ▶ Inactive resources that are LIC-CC enabled, that is, memory that is not being used by any activate logical partitions
- ▶ Memory purchased with the flexible memory option
- ▶ Additional books as discussed previously

The I/O connectivity must also support the removing of the book. The majority of the pathing to the I/O is covered by the design of the z9 EC with the redundant I/O interconnect support in the I/O cages that allow for connection through multiple STIs. You will need to ensure that all the ICBs have redundant paths from different books. (STI Rebalance may need to be ordered after a concurrent book add to provide even distribution of the connections across books; see “STI Rebalance feature” on page 96).

If sufficient resources are not present on the remaining books, you may have to deactivate some non-critical logical partitions, configure off CPs or specialty engines, or configure storage offline to reach the required level of available resources. Planning to address these possibilities should help to reduce operational errors.

**Note:** Single book systems like the S08 cannot make use of the EBA function.

The planning should be included as part of the initial installation and any follow-on upgrade that modifies the operating environment. The e-config report can be used to determine the number of books, active PUs, memory configuration, and the channel layout.

If the z9 EC is installed, you can use the Prepare For Enhanced Book Availability task of the EBA process (see Figure 8-11 on page 236) to determine what resources are required to support the removal of a book with acceptable degradation to the operating system images.

The EBA prepare process determines what resources like memory, CPs, and I/O paths will have to be freed to allow for the removal of a given book. This preparation can be run on each book to determine what resources changes would be needed and the result can be used as input to your planning. Doing this allows you to identify critical resources.

Using this information, you could look at the logical partitions configuration and workloads priorities to determine ahead of time how resources could be reduced to allow for the book to be removed. The facility is accessed through the SE under the Perform Model Conversion. The option is Prepare for Enhanced Book Availability. The process can be run for each of the books, multiple times, according to multiple scenarios. The results can be used to plan for Enhanced Book Availability.

The planning should include:

- ▶ Review of the z9 EC configuration to determine:
  - Number of books installed and the number of PUs enabled.
    - Use the e-config output or the HMC to determine the model and number and type of PUs (CPs, IFL, ICF, zAAP, and zIIP).
    - Amount of memory, both physically installed and LIC-CC enabled.
    - Work with your IBM service Personnel to determine the memory card size in each book. The memory card sizes and the number of cards installed for each installed book can be viewed from the SE under the CPC Configuration task list, using the View Hardware Configuration Option.
  - Channel Layouts, STI to Channel connections. Use e-config output to review channel config including the STI pathing. This is a normal part of the I/O connectivity planning. The alternate paths should be separated as far into the system as possible.
- ▶ Review of the system images configurations to determine the resources for each.
- ▶ Determine the importance and relative priority of each logical partition.

- ▶ Identify the logical partition or workloads and the actions to be taken:
  - Deactivate the whole logical partition.
  - Configure off CPs.
  - Reconfiguration of memory (may require use of RSU<sup>1</sup> value).
  - Vary off of channels.
- ▶ Review the channel layout and determine if any changes are needed to address single paths.
- ▶ Develop the plan to address the requirements.

By performing the review you can document what resources could be available if the EBA were to be used. The resources on the books are allocated at POR of the system and can change during a POR. The review should be done when changes are made to the z9 EC, like adding books, CPs, memory, or channels or when workloads are added or removed. The prepare function will be used ahead of any EBA action to determine that the system can be conditioned to allow for the removing of the book or what actions will be required to support the removal.

### 8.3.2 Enhanced Book Availability - Process

In order to use the EBA, certain conditions must be satisfied:

- ▶ Used processors (PUs) on the book to be removed need to be freed up.
- ▶ Utilized memory on the book must be freed up.
- ▶ For all I/O domains connected to this book, alternate paths must exist or the I/O needs to be placed offline.

The EBA process, this is the prepare phase, is started from the Support Element, either directly or using single object operation task from the HMC using the *Perform model Conversion* from CPC configuration task list; see Figure 8-11 on page 236.

#### **Processor availability**

Processor resource availability for reallocation/deactivation is affected by the type and quantity of resources in use.

- ▶ Total number of PUs LIC-CCed
- ▶ PU definitions in the profiles
  - Dedicated and dedicated reserved
  - Shared
- ▶ Active logical partitions with dedicated resources at the time of book repair or replacement

To maximize the PU availability option:

- ▶ Ensure that you have sufficient inactive physical resources on the remaining books to complete a book removal.
- ▶ For maximum availability, you may consider purchasing a z9 EC using the guidance in 8.3.1, “Planning consideration” on page 232.

<sup>1</sup> For additional information about Reconfigurable Storage Unit, see *System z9 Processor Resource/Systems Manager (PR/SM) Planning Guide*, SB10-7041.



### **Memory availability**

Memory resource availability for reallocation/deactivation depends on:

- ▶ Physically installed storage
- ▶ Image profile storage allocations
- ▶ Amount of LIC-CCed memory
- ▶ Flexible memory option

There are 23 maximum availability memory features available for models S18, S28, S38, and S54, delivering 32 GB to 384 GB in 16 GB increments (FC 2802 to 2824).

### **MBA to STI-MP connectivity requirements**

The optimum situation is to maintain maximum I/O connectivity during book removal.

- ▶ The Redundant I/O Interconnect (RII) function provides for redundant STI connectivity to all installed I/O domains in all I/O cages.
- ▶ However, ICB-3 and ICB-4 do not take advantage of the RII function. In this case, the associated CHPIDS need to be configured offline.
  - ICB-4s connect directly from the MBA port on the book to an STI on another server.
  - ICB-3s connect directly from the MBA port on the book to a card in the I/O cage; however, this card is not physically part of the I/O domain in which it is plugged. It receives its STI signalling directly from the MBA through the STI cable and converts this connection to two ICB-3 ports.
  - Always ensure there are redundant ICBs to the same CF or z/OS image, from different books. Anytime a single book in a multiple books configuration has all, or the vast majority of ICBs, the STI rebalance feature should be considered. See “STI Rebalance feature” on page 96.

### **Preparing for Enhanced Book Availability**

The Prepare for Concurrent Book replacement step validates that enough dormant resources exist for this operation. If enough resources are not available on the remaining books to complete the EBA process, it identifies the resources and guides the user through a series of steps to select and free up resources. The prepare step does not complete until all memory and I/Os conditions have been successfully resolved.

**Note:** The prepare step does not reallocate any resources; it is only used to record customer choices and produce a configuration file on the SE that will be used by the Perform concurrent Book replacement operation.

The prepare step can be done in advance. However, if there are any changes to the configuration between the time the prepare is done and the book is physical removed, the prepare phase will need to be rerun.

The process can be run multiple times, as it does not move any resources. The Support Element window also includes the option Display Previous Prepare Enhanced Book Availability Results to view the results of the last Preparation operation done.

The prepare step can be run several times without actually performing a book replacement. It allows you to dynamically adjust the operational configuration for book repair or replacement prior to CE activity. Figure 8-11 shows the options for the Prepare for Enhanced Book Availability.

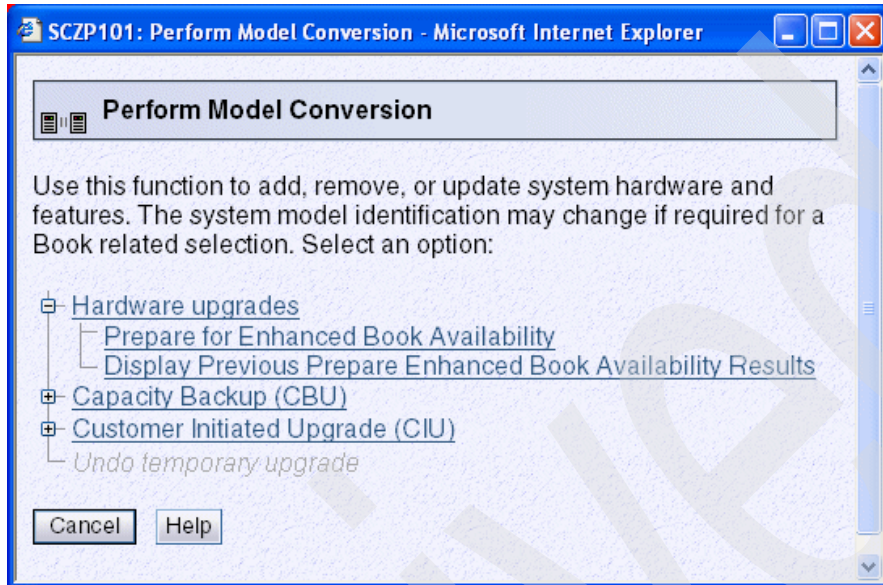


Figure 8-11 Perform Model Conversion, Prepare for Enhanced Book Availability

The next window, Figure 8-12, will be presented so you can select the book that is to be repaired or upgraded. Only one book can be checked at a time.

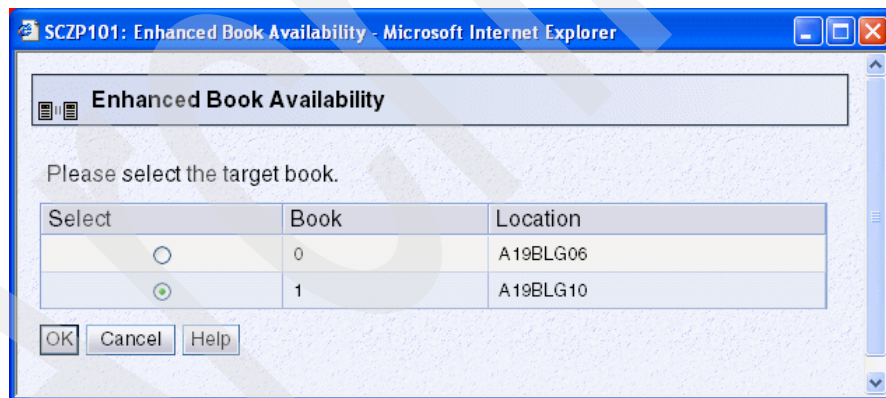


Figure 8-12 Enhance Book Availability, selecting the target book

The system will verify the resources required for the removal, determine the required actions needed, and present the results for review. Depending on the configuration, the task could take up to thirty minutes.

The prepare step determines the readiness of the system for the removal of the targeted book. The configured processors and the in-use memory will be evaluated against unused resources available across the remaining books.

If not enough resources are available, the conflicts will be identified so that you can take action to free up resources. I/O connections associated with the removal of the targeted book will be analyzed for any single path I/O connectivity.

There are three states that can result from the prepare option:

- ▶ The system is ready to perform the Enhanced Book Availability for the targeted book with the original configuration.
- ▶ The system is not ready to perform the Enhanced Book Availability due to conditions noted from the preparation step.
- ▶ The system is ready to perform the Enhanced Book Availability for the targeted book. However, processors were reassigned from the original configuration in order to continue. The results of this reassignment must be reviewed relative to the customer operation and business requirements. The reassignments can be changed on the final window that is presented. Before changes are made or the reassignment is approved, ensure that the correct level of support have reviewed based on the needs of their business and have been agreed to.

The results of the preparation will be presented in a tabbed format for review (see Figure 8-13). These are the conditions that prevent the EBA option from being performed. There are tabs on the resulting panel for processors, memory, and various single path I/O conditions. Only the tabs that have conditions that prevent the book from being removed are displayed. Each tab will indicate what the specific conditions are and possible options to correct the conditions. Possible tabs selections are:

- ▶ Processors
- ▶ Memory
- ▶ Single I/O
- ▶ Single Domain I/O
- ▶ Single Alternate Path I/O

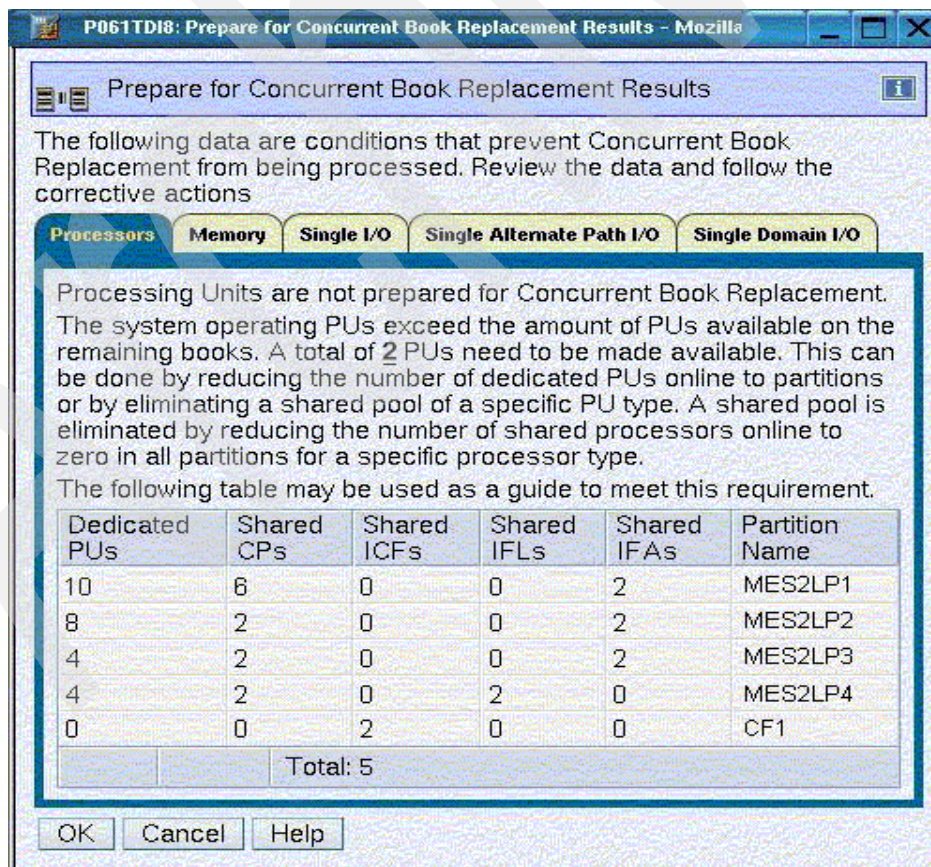


Figure 8-13 Prepare for Concurrent Book Replacement Results

The following are examples of the different windows that can be presented as you go through the preparation phase. These are presented if the condition is found that requires actions to condition the system for the book removal. Figure 8-14 shows the results of the preparation phase for removing Book 0. Across the top of the frame are multiple tabs. The tabs show the different conditions that were found in preparing the book to be removed. The Memory tab is selected and the text in the frame shows that the amount of memory in use and the amount available when the book is removed. The message provides the amount of memory that must be made available, in this case, 6144 MB. Below that is the breakdown of the in-use memory by partition name. Once the required amount of memory has been made available, rerun the preparation to verify.

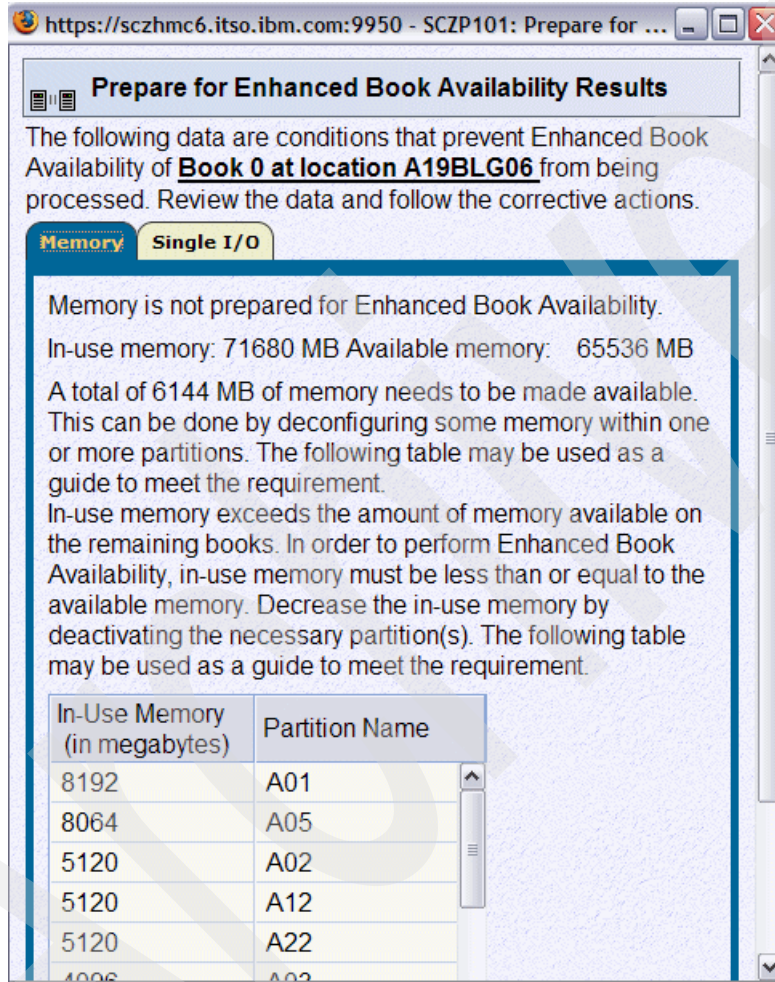


Figure 8-14 Prepare for EBA - memory conditions

The other tab is for Single I/O (see Figure 8-15). The preparation identified one single I/O path associated with the removal of Book 0. The path would need to be placed offline to perform the book removal. Once the condition has been addressed, rerun the prepare to ensure all the required conditions have been met.

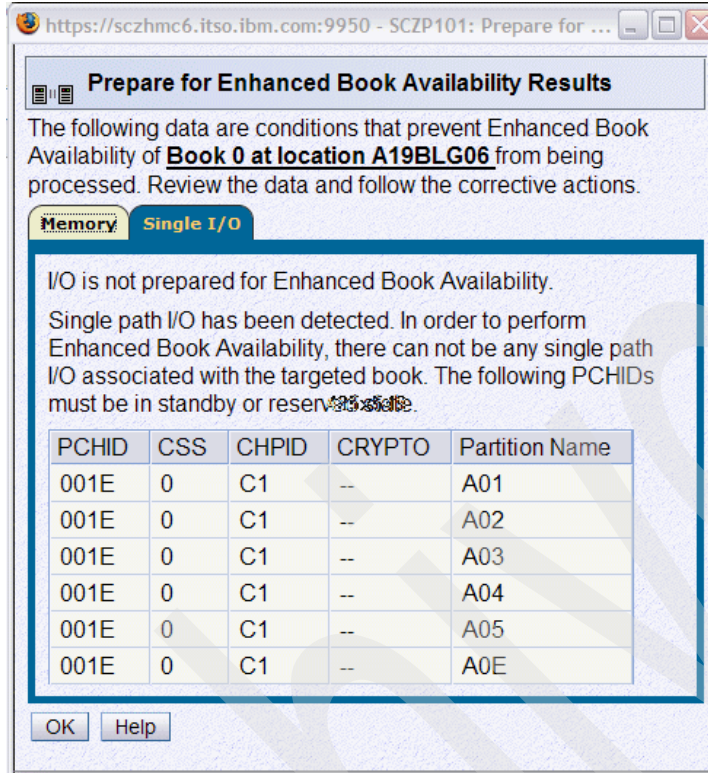


Figure 8-15 Prepare for EBA - single I/O result

### Getting the server ready to perform Enhanced Book Availability

During the preparation, the system determines the CP configuration that will be required to perform the book remove. Figure 8-16 shows the results and provides the option to change the assignment on non-dedicated processors.

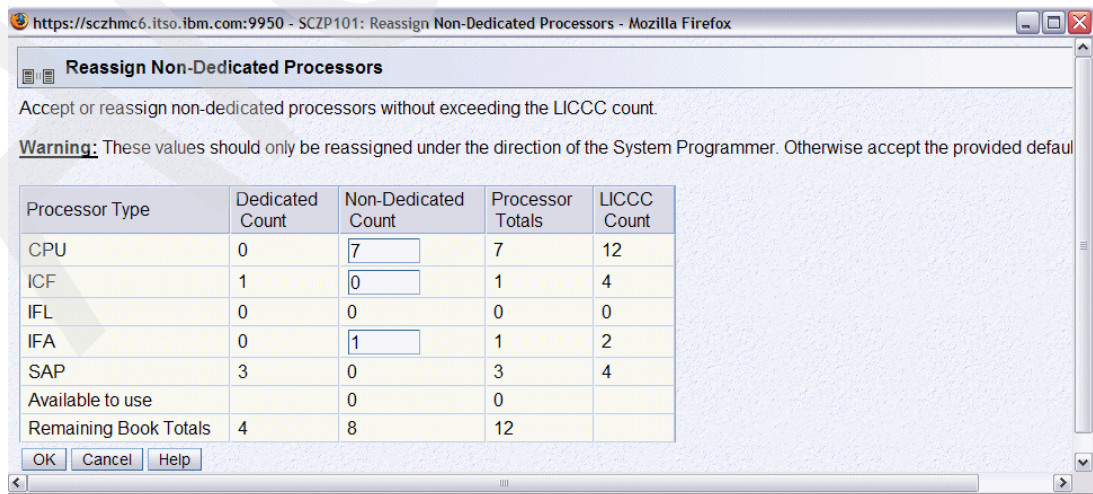


Figure 8-16 Reassign Non-dedicated Processors Results

It is important to understand the results of these changes relative to the operational environment. It requires that you understand the potential impact of making such operational changes. Changes to the PU assignment, although technically correct, could result in constraints for critical system images. In some cases, the solution may be to defer to a time frame that would have less impact on the production system images.

Once the reassign results have been reviewed, and adjusted if needed, selecting **OK** will present the final results of the reassignment, which would include changes made as a result of the review. Figure 8-17 shows the results; this will be the assignments when the book evacuation phase of the EBA is done.

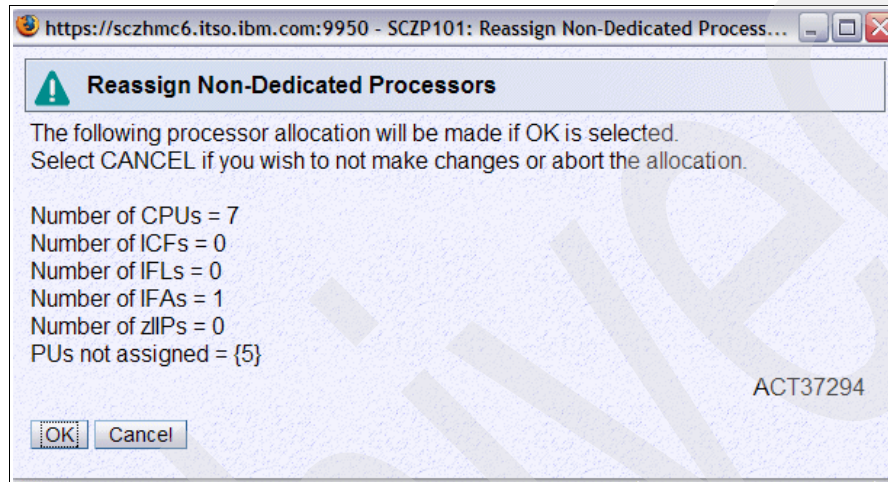


Figure 8-17 Reassign Non-Dedicated Processors - ACT37294

## Summary of the book removal process

The following section steps through the process of a concurrent book replacement.

To evacuate a book, the following resources must be moved to the remaining active books:

- ▶ Enough PUs must be available. This includes all types of characterizable PUs (CP, IFL, ICF, zAAP, zIIPs, and SAP).
- ▶ Enough installed memory must be available in the remaining active books.
- ▶ I/O connectivity: Alternate paths to other books must be available or the I/O will need to be taken offline.

It is important to understand both the server configuration and the LPAR allocation for memory, PUs, and I/O. This knowledge is required to make the best decision on how to free up the necessary resources to allow for a book evacuation.

The steps for concurrent book replacement are:

1. Run the preparation task to determine the needed resources (see “Summary of the book removal process” on page 240).
2. Review the results.
3. Determine the required actions needed to meet the required conditions for EBA.
4. When ready for the book removal, free up the resources indicated in the prepare step.
5. Re-run the step in Figure 8-11 on page 236 to ensure the required conditions are all satisfied.

6. When the completed successfully message is received (see Figure 8-18), the system is now ready for the removal of the book.

The preparation step can be run multiple times to ensure the conditions have been met. It does not reallocate any resources; all it does is produce a results report. The resources will not be reallocated until the perform part of the book removal.

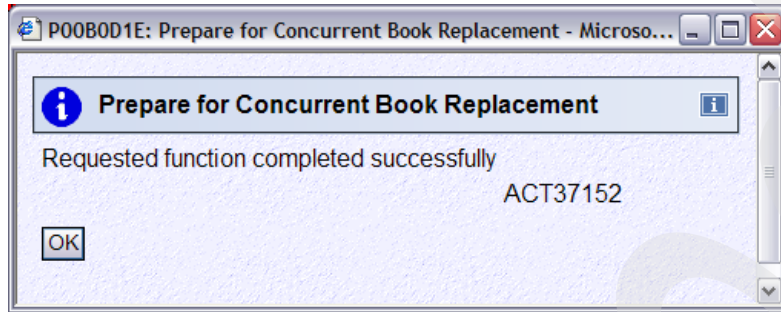


Figure 8-18 Prepare completed successfully - ACT37152

► Processor rules during EBA

All processors in any remaining books are available to be used during EBA. This includes the two spare PUs or any non LIC-CCed PU available.

The EBA process also allows conversion of one PU type to another PU type. One example would be turning a zAAP into a CP for the duration of the EBA function. The preparation for concurrent book replacement task will also indicate if any *SAPs* will need to be moved to the remaining books.

► Memory rules during EBA

All physical memory installed in the system, including flexible memory FC 28xx, is available for the duration of the EBA function. Any physical installed memory, whether purchased or not, is available to be used by the EBA function.

► Single I/O rules during EBA

Alternate paths to other books must be available or the I/O will need to be taken offline.

**Note:** ICB-4 and STI-3 cards that provide ICB-3 connectivity are connected directly to the STI in the physical book. Removal of the book requires that all STI cables be disconnected. Connectivity for ICB4 and ICB3 would be lost during this time. Make sure that there are redundant paths available.

Review the results. The result of the preparation task is a list of resources that need to be made available before the book replacement can take place.

At this stage, create a plan to free up these resources. To free up resources:

- Options available to free up PUs:
  - Vary CPs offline, reducing the number of CP in the shared CP pool.
  - Use any spare CP.
  - Deactivate logical partitions.
  - PU conversion: An example would be converting a zAAP to a CP.

- ▶ To free up memory
  - Deactivate a logical partition.
  - Vary offline some of the reserved (online) memory<sup>2</sup> using the command:

```
CONFIG_STOR(E=1), <OFFLINE/ONLINE>
```

This allows for a storage element to be taken offline. Note that the size of the storage element is dependant on your RSU value. The following command allows you to configure offline smaller amounts of storage than what you have set for your storage element.

```
CONFIG_STOR(nnM), <OFFLINE/ONLINE>
```

- A combination of both logical partition deactivation and vary offline command.

**Note:** If you plan to use the EBA function, we recommend that you set up reserved storage and set a RSU value. The RSU value allows you to specify the number of storage units that are to be kept free of long-term fixed storage allocations, allowing for storage elements (from the RSU value) to be varied offline.

## 8.4 Enhanced Driver Maintenance (EDM)

Enhanced Driver Maintenance, exclusive to the System z9, is another step in reducing the duration of a planned outage. One of the contributor's planned outages is Licensed Internal Code updates performed in support of new features and functions. When properly configured, the z9 EC is designed to support activating a select new LIC level concurrently. Concurrent activation of the select new LIC level is only supported at specific sync points (points in the maintenance process when LIC may be applied concurrently (MCL service level)). Sync points may exist throughout the life of the current LIC level. Once a sync point has passed, a user will be required to wait until the next sync point supporting concurrent activation of a new LIC level. Certain LIC updates will not be supported by this function.

The key points of EDM are:

- ▶ The ability to concurrently install and activate a new driver can eliminate planned outage.
- ▶ Select a window of opportunity within the code maintenance stream.
- ▶ Like some concurrent patches, you may need to vary off/on certain devices.
- ▶ The ability to concurrently move from one patch point on major driver level N to a patch point on major driver level N+1.
- ▶ It cannot move any-to-any; it must move from a specific *from* patch bundle to a specific *to* patch bundle.
- ▶ A limited number of specific crossover bundles will be defined by IBM for a driver.
- ▶ Disruptive driver upgrades are permitted at any time.
- ▶ Concurrent crossover from driver level N to driver level N+1, to driver level N+2 must be done serially; no composite moves.
- ▶ No concurrent back-off is possible. It must move forward to driver level N+1 once Enhanced Driver Maintenance is initiated. Catastrophic errors during update may lead to an outage.

<sup>2</sup> The amount of storage that will come offline will depend on how many long-term fixed pages are in use.



- ▶ The current plan is to try to have sync points generated at regular intervals, as shown in Figure 8-19.

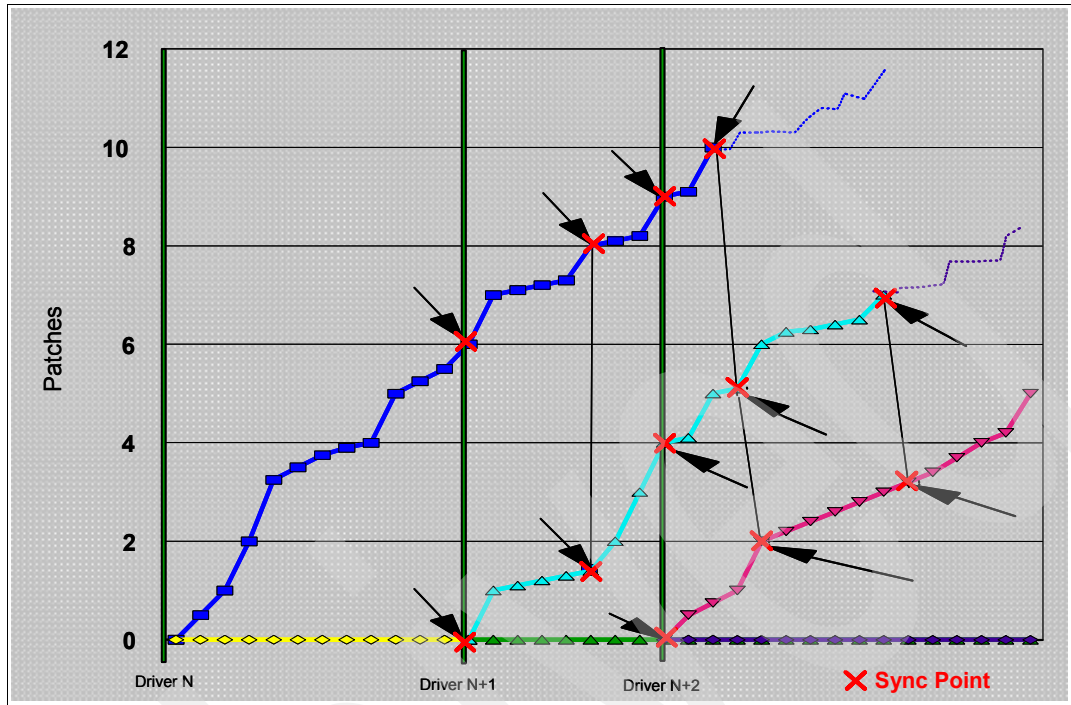


Figure 8-19 Driver levels and relationship to sync points

The EDM function does not completely eliminate the need for planned outages for driver level upgrades. Although very infrequent, there will be circumstances when the system will have to be scheduled for an outage. Below is a list of circumstances when such an outage may have to be planned for:

- ▶ Specific complex code changes may dictate a disruptive driver upgrade. This will be alerted in advance for planning purposes.
  - Design data fixes
  - CFCC level change
  - LPAR code fixes
- ▶ Non-QDIO OSA CHPID types will still require CHPID Vary OFF/ON in order to activate new code.
  - Adapter type: OSA-ICC (OSC)
  - Adapter type: Ethernet (OSE)
- ▶ Cryptographic code load will require a Config OFF/ON in order to activate new code.
- ▶ FICON and FCP code changes involving code loads will require CHPID *reset* to activate.

## 8.5 Nondisruptive upgrades

Continuous availability is an increasingly important requirement for most customers, and even planned outages are no longer acceptable. Although Parallel Sysplex clustering technology is the best continuous availability solution for z/OS environments, nondisruptive upgrades within a single server can avoid system outages and are suitable to further operating system environments.

The z9 EC allows *concurrent* upgrades, meaning it is possible to dynamically add more capacity to the server. If operating system images running on the upgraded server need no disruptive tasks to use the new capacity, the upgrade is also *nondisruptive*. This means that Power-on Reset (POR), logical partition deactivation, and IPL do not have to take place.

If the concurrent upgrade is intended to satisfy an *image upgrade* to a logical partition, the operating system running in this partition must also have the capability to concurrently configure more capacity online. z/OS operating systems have this capability. z/VM can concurrently configure new processors and I/O devices online, but it does not support dynamic storage reconfiguration.

Linux operating systems do *not* have the capability of adding more resources concurrently. However, Linux virtual machines running under z/VM can take advantage of the z/VM capability to nondisruptively configure more resources online (processors and I/O).

**Important:** Dynamic add/delete of a logical partition name allows reserved partition *slots* to be created in an IOCDS in the form of extra Channel Subsystem, Multiple Image Facility (MIF) image pairs, which can be later assigned a logical partition name for use (or later removed) through HCD concurrently.

**Important:** If the STI Rebalance feature (FC 2400) is selected at server upgrade configuration time, and effectively results in STI rebalancing for ICBs, it will also change the Physical Channel ID (PCHID) number of ICB-4 links, requiring a corresponding update on the server's I/O definition through HCD/HCM. The STI Rebalance feature is disruptive.

### Processors

CPs, IFLs, ICFs, zAAP, and zIIP processors can be concurrently added to a z9 EC if there are spare PUs available on any installed book. The number of zAAPs cannot exceed the number of CPs plus unassigned CPs on a z9 EC. The same holds true for the zIIPs.

Additional books can also be installed concurrently, allowing further processor upgrades.

A processor upgrade cannot be performed when CBU or On/Off CoD is activated.

Concurrent upgrades are not supported with CPs defined as additional SAPs.

If reserved processors are defined to a logical partition, then z/OS and z/VM operating system images can dynamically configure more processors online, allowing nondisruptive processor upgrades. The Coupling Facility Control Code (CFCC) can also configure more processors online to Coupling Facility logical partitions using the CFCC image operations window.

## Memory

Memory can be concurrently added to a z9 EC up to the physical installed memory limit. Additional books can also be installed concurrently, allowing further memory upgrades by LIC-CC enabling memory capacity on the new books.

Using the previously defined reserved memory, z/OS operating system images can dynamically configure more memory online, allowing nondisruptive memory upgrades.

## I/O

I/O cards can be added concurrently to a z9 EC if all the required infrastructure (I/O slots and STIs) is present on the configuration. The Plan Ahead process can assure that an initial configuration will have all the infrastructure required for the target configuration.

Also, I/O ports can be concurrently added by LIC-CC, enabling available ports on ESCON and ISC-3 daughter cards.

Dynamic I/O configurations are supported by some operating systems (z/OS and z/VM), allowing nondisruptive I/O upgrades. However, it is not possible to have dynamic I/O reconfiguration on a stand-alone Coupling Facility server, because there is no operating system with this capability running on this server. Dynamic I/O configurations require additional space in the HSA for expansion.

## PCI-X Cryptographic adapters

Crypto Express2 features can be added concurrently to a z9 EC if all the required infrastructure, I/O slots, and STIs are present on the configuration. The Plan Ahead process can assure that an initial configuration will have all the infrastructure required for the target configuration.

In order to make the addition of Crypto Express2 features nondisruptive, logical partitions must be predefined with the appropriate PCI-X cryptographic adapters selected in its candidate list on the partition image profile. To maximize concurrent upgrade capabilities, it is recommended that all eligible logical partitions define all possible PCI-X cryptographic adapters as candidates for the logical partition. This is possible even if there are no PCI-X cryptographic adapters currently installed on the server.

### 8.5.1 Planning for nondisruptive upgrades

CUoD, CIU, On/Off CoD, and CBU can be used to concurrently upgrade a z9 EC. But there are some situations that require a disruptive task to enable the new capacity just added to the server. Some of these can be avoided if planning is done in advance. Planning ahead is a key factor for nondisruptive upgrades.

#### Reasons for disruptive upgrades

These are the current main reasons for disruptive upgrades on the z9 EC, but with careful planning and by reviewing “Recommendations to avoid disruptive upgrades” on page 246, you can minimize the need for these outages:

- ▶ Changing the number of logical partitions defined to a z9 EC. The only way to add or delete a logical partition is by a POR using a new IOCDs, including or excluding the new partition.
- ▶ Changing the number of CSSs on a server.
- ▶ Changing the maximum number of subchannels supported on an CSS.
- ▶ Changing the number of Multiple subchannel set (MSS) within an CSS.

- ▶ Logical partition processor upgrades when reserved processors were not previously defined are disruptive to image upgrades.
- ▶ Logical partition memory upgrades when reserved storage was not previously defined are disruptive to image upgrades.
- ▶ Installation of I/O cages is disruptive.
- ▶ An I/O upgrade when the operating system cannot use the Dynamic I/O configuration function.  
Linux, z/VSE, TPF, z/TPF, and CFCC do not support Dynamic I/O configuration.
- ▶ Adding a Crypto Express2 PCI-X adapter to a logical partition, if not predefined with the appropriate adapter number selected in the PCI Cryptographic Candidate List of the logical partition's image profile.

### Recommendations to avoid disruptive upgrades

Based on the previous list of reasons for disruptive upgrades, here are some recommendations to avoid or at least minimize these situations, increasing the possibilities for nondisruptive upgrades:

- ▶ Define spare or reserved logical partitions.

A z9 EC can have up to 60 logical partitions defined. It is possible to define more partitions than you need in the initial configuration:

- If running a z/OS version prior to z/OS V1.6, include more partition names in the IOCP statement RESOURCE. The *spare partitions* do not need to be activated, so any valid partition configuration can be used during their definitions. The initial definitions (LPAR mode, processors, and so on) can be changed later to match the image type requirements.

The only resource that spare partitions will use is subchannels; keep in mind that z9 ECs can have up to 7665 K subchannels (127.75 K per logical partition \* 60 partitions) total in HSA.

- If you are running z/OS V1.6 or later, defining reserved logical partitions. A reserved partition is defined with the partition name placeholder “\*”.

No I/Os are assigned to the reserved logical partition. The dynamic logical partition name definition allows reserved partition *slots* to be created in an IOCDs in the form of extra Channel Subsystem, Multiple Image Facility (MIF) image pairs. These extra Channel Subsystem MIF image ID pairs (CSSID/MIFID) can be later assigned a logical partition name for use (or later removed) through dynamic I/O commands using the Hardware Configuration Definition (HCD). The dynamic activation that names the logical partition also adds all the channels that the logical partition will have access to based on the HCD definitions. The IOCDs still must have the extra I/O slots defined in advance, since structures are built in the Hardware System Area (HSA).

- ▶ Define an appropriate number of Channel Subsystems (CSSs).

Define the appropriate number of CSSs (maximum is four), based on the required number of logical partitions (maximum is 60) and the number of CHPIDs (maximum is 256 per image and per CSS), that a future configuration may have. Spare and reserved logical partitions, which can have partition name “\*” for future renaming, as described in the previous item, can help to define additional CSSs for future use.

Spanned channels can help spread logical partitions across CSSs while maintaining physical channels sharing for some channel types.

- ▶ Define the maximum supported number of subchannels on each CSS.  
The z9 EC can have up to 7665 K subchannels (127 K per logical partition \* 60 partitions) total in HSA.
- ▶ Configure as many Reserved Processors (CPs, IFLs, ICFs, zAAPs and zIIPs) as possible.  
Configuring Reserved Processors for all logical partitions *before* their activation enables them to be nondisruptively upgraded. The operating system running in the logical partition must have the ability to configure processors online. The total number of defined plus reserved CPs cannot exceed the number of CPs supported by the operating system. z/OS can support up to 32 CPs plus zAAPs and zIIPs.
- ▶ Configure Reserved Storage to logical partitions.  
Configuring Reserved Storage for all logical partitions *before* their activation enables them to be nondisruptively upgraded. The operating system running in the logical partition must have the ability to configure memory online. The amount of reserved storage can be above the book threshold limit (128 GB), even if no other book is already installed. The current partition storage limit is 128 GB.
- ▶ Consider the Flexible Memory option.  
Use a convenient entry point memory capacity and consider the Flexible memory option to allow future upgrades within the memory cards already installed on the books.
- ▶ Use the Plan Ahead concurrent condition for I/O.  
Use the Plan Ahead concurrent condition process to include in the initial configuration all the I/O cages required by future I/O upgrades, allowing the planned concurrent I/O upgrades.
- ▶ Crypto Express2 features requires careful planning to prepare for nondisruptive changes.  
Although the installation of a Crypto Express2 feature is concurrent, a change to a logical partition image profile to modify its cryptographic domain index(es) or Candidate list is disruptive to the partition. It requires a partition deactivation-activation to take effect.  
The cryptographic PCI-X adapter number coupled with the usage domain index must be unique across all *active* logical partitions. Within these limits, define all planned Crypto Express2 PCI-X adapters as candidates for eligible logical partitions.  
Each PCI-X adapter provides 16 domains, and up to 60 partitions can be defined and active. When all 60 logical partitions require *concurrent* access to cryptographic functions, the server must have at least two Crypto Express2 features installed (four PCI-X adapters with 16 domains per PCI-X adapter). More may be needed for redundancy.  
Note that the same PCI-X adapter number and usage domain index combination may be defined to more than one logical partition. In that case, only one of the logical partitions can be active at any one time. However, this may be valid to define a configuration for backup situation.  
For detailed planning and configuration information about Crypto Express2 features, see the Redbooks publication *IBM System z9 109 Configuration Setup*, SG24-7203.

### **Considerations when installing additional books**

During a z9 EC upgrade, additional books can be installed concurrently. Depending on the number of additional books in the upgrade and the customer's I/O configuration, a STI rebalancing for ICBs may be recommended for availability reasons. It will change ICB-4 PCHID numbers, requiring an I/O definition update.

Archived



## Environmental requirements

This chapter describes, in short, the z9 EC environmental requirements. We list its dimensions, weights, power, and cooling requirements as an overview of what is needed to plan for the installation of a z9 EC server.

For more comprehensive physical planning information, refer to *IBM System z9 Installation Manual for Physical Planning*, GC28-6844.

The following topics are covered:

- ▶ 9.1, “Power and cooling” on page 250
- ▶ 9.2, “Weights” on page 252
- ▶ 9.3, “Dimensions” on page 252

## 9.1 Power and cooling

The z9 EC is always a two-frame system. The frames are shipped separately and are fastened together when installed.

Installation of a z9 EC is always on a raised floor. The number of cables to be expected for most configurations may be so large that installation is only possible with space underneath. The dimensions and weight of a z9 EC are equal or differ slightly compared to a z990.

The z9 EC requires at least two power feeds and uses two redundant three-phase line cords, allowing the system to survive the loss of power to either one. In case of a power failure of one of the line cords, the other one is able to take over the entire load to keep the system operating without interruption. The z9 EC is installed with three-phase wiring and operates with 50/60Hz AC power, and voltages ranging from 200V to 480V. For ancillary equipment (like the Hardware Management Console, its display, and the modem), additional single-phase outlets are required.

### 9.1.1 Power consumption

Actual power consumption is dependent on the server configuration in terms of the number of books and the number of I/O cages installed. The figures listed in Table 9-1 assume the maximum configuration.

Table 9-1 Power consumption and heat load

Model	One I/O cage	Two I/O cages	Three I/O cages
z9 EC Model S08	6.3 kW	9.2 kW	12.1 kW
z9 EC Model S18	8.8 kW	11.8 kW	14.7 kW
z9 EC Model S28	10.9 kW	13.9 kW	16.9 kW
z9 EC Model S38	12.8 kW	15.7 kW	18.3 kW
z9 EC Model S54	12.8 kW	15.7 kW	18.3 kW

Input power in kVA is equal to the output power in kW. Heat output expressed in kBTU per hour is derived by multiplying the table entries by a factor of 3.4.

The maximum allowed circuit breaker rating is 60 Amps, which is to be used for both power feeds where 200-240V is applicable. For 380-414V, 32 Amps, and for 480 Volts, 30 Amps are recommended for both power feeds.

#### zPower Estimation Tool

The power consumption tool for System z9 is available through the IBM Resource Link Web site. This tool provides an estimate of the anticipated power consumption of a particular machine model and its associated configuration. A user will input the machine model, memory size, number of I/O cages, and quantity of each type of I/O feature card. The tool will output an estimate of the power requirements needed for this system:

<http://www.ibm.com/servers/resourceLink>

It is designed to help in power and cooling planning for new or currently installed IBM System z9 servers.



## 9.1.2 Internal Battery Feature

The optional Internal Battery Feature (IBF) provides sustained system operations for a relatively short period of time, allowing for orderly shutdown. In addition, an external UPS system can be connected to the System z9, allowing for longer periods of sustained operation.

The Internal Battery Feature, given that the batteries are not older than three years and have been discharged regularly, are capable of providing emergency power for the periods of time shown in Table 9-2.

Table 9-2 Internal Battery Feature emergency power times

Model	One I/O cage	Two I/O cages	Three I/O cages
z9 EC Model S08	9 minutes	13 minutes	10 minutes
z9 EC Model S18	13 minutes	11 minutes	12 minutes
z9 EC Model S28	11 minutes	13 minutes	10 minutes
z9 EC Model S38	14 minutes	10 minutes	9 minutes
z9 EC Model S54	14 minutes	10 minutes	9 minutes

## 9.1.3 Emergency power-off

On the front of frame A is an emergency power-off switch that will immediately disconnect utility and battery power from the server when activated. This causes all volatile data in the server to be lost.

In case a server is connected to a machine room emergency power-off switch, and the Internal Battery Feature is installed, the batteries will take over if the switch is engaged.

It is possible to connect the machine room emergency power-off switch to the server power-off switch. In that case, when the machine room emergency power-off switch is engaged, all power will be disconnected from the line cords and the Internal Battery Features. This causes all volatile data in the server to be lost.

## 9.1.4 Cooling requirements

The z9 EC requires chilled air from under the raised floor to fulfill the air-cooling requirements. The chilled air is usually provided through perforated floor tiles. The amount of chilled air needed in the computer room is indicated in the IMPP for a variety of underfloor temperatures.

At an underfloor temperature of 20° Celsius (68° Fahrenheit), the cooling airflow requirements in cubic feet per minute (CF/M) and cubic meters per minute (CM/M) with maximum populated I/O cages are listed in Table 9-3.

Table 9-3 Underfloor cooling airflow requirements (CF/M (CM/M))

Model	One I/O cage	Two I/O cages	Three I/O cages
z9 EC Model S08	821 (23.2)	1137 (32.2)	1453 (41.1)
z9 EC Model S18	1137 (32.2)	1453 (41.1)	1769 (50.1)
z9 EC Model S28	1453 (41.1)	1769 (50.1)	2085 (59.0)
z9 EC Model S38	1453 (41.1)	1769 (50.1)	2085 (59.0)

Model	One I/O cage	Two I/O cages	Three I/O cages
z9 EC Model S54	1453 (41.1)	1769 (50.1)	2085 (59.0)

## 9.2 Weights

Since there may be a large number of cables connected to a z990 installation, a raised floor is mandatory. In the IMPP, weight distribution and floor loading tables are published, to be used together with the maximum frame weight, frame width, and frame depth to calculate the floor loading for the z990 system.

Table 9-4 indicates the minimum and maximum system weights for all models. The weight ranges are based on configuration models with one and three I/O cages.

Table 9-4 System weights

Configuration	Weight in kg (lb) without IBF	Weight in kg (lb) with IBF
z9 EC Model S08	1210 (2668) to 1548 (3412)	1299 (2865) to 1726 (3806)
z9 EC Model S18	1312 (2892) to 1649(3636)	1490 (3286) to 1917 (4227)
z9 EC Model S28	1354 (2986) to 1692 (3730)	1533 (3380) to 1960 (4321)
z9 EC Model S38	1421 (3132) to 1734 (3824)	1689 (3723) to 2003 (4415)
z9 EC Model S54	1421 (3132) to 1734 (3824)	1689 (3723) to 2003 (4415)

## 9.3 Dimensions

The z9 EC always has two frames: Frame A and frame Z. The external dimensions of both frames of a z9 EC, with and without covers, are listed in Table 9-5.

Table 9-5 Frame dimensions

Frames	Width mm (in)	Depth mm (in)	Height mm (in)
Frame A without covers	750 (29.5)	1171 (46.1)	1921 (75.6)
Frame A with covers	767 (30.2)	1577 (62.1)	1941 (76.4)
Frame Z without covers	750 (29.5)	1171 (46.1)	1921 (75.6)
Frame Z with covers	767 (30.2)	1476 (58.1)	1941 (76.4)

**Note:** The total machine room area required is 2.49 square meters (26.78 square feet). With service clearance, 5.45 square meters (58.69 square feet) are needed.

## 9.4 Frame tie down for raised floor and non-raised floor

A Bolt-Down Kit for raised floor and non-Raised floor environments is available to be ordered for the System z frames. It provides hardware to make System z9 frames more rugged and to tie them down to a concrete floor beneath a 9- to 13-inch or 12- to a 22-inch raised floor or in a non-raised floor installation.

- ▶ (#7995): Bolt-Down Kit, High-Raised Floor 2094 feature provides frame stabilization and bolt-down hardware for securing a frame to a concrete floor beneath a 11.75- to 16.0-inch (298mm to 405mm) raised floor.
- ▶ • (#7996): Bolt-Down Kit, Low-Raised Floor 2094 feature provides frame stabilization and bolt-down hardware for securing a frame to a concrete floor beneath a 9.25- to 11.75-inch (235mm to 298mm) raised floor.

These are designed to help secure the frames and their contents from damage when exposed to shocks and vibrations such as those generated by a seismic event. The frame tie downs are intended for securing a System z frame weighing less than 3600 lbs per frame. For IBM System z9 EC, you need a quantity of two Bolt-Down kits.

## 9.5 Restriction of Hazardous Substances

In 2003, the European Union (EU) passed a Restriction of the use of certain Hazardous Substances in Electrical and Electronic Equipment Directive requiring that member states (EU member countries, Norway, and Switzerland) comply.

This directive restricts the use of lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls, and polybrominated diphenyl ether flame retardants in new electrical and electronic equipment put on the market in EU member countries as of July 1, 2006. However, the RoHS Directive contains several exemptions that will allow continued use of these materials in some applications after the phaseout date.

IBM System z9 Enterprise Class will fully comply with the EU RoHS requirements by July 1, 2006.

In this regard, IBM has undertaken an aggressive development program to release products that comply with the requirements of the Directive. IBM programs include an internal development program, as well as active involvement with the supply chain and various consortia, universities and national laboratories.

Customers with questions regarding a particular product or product line should contact their IBM client representative.

See also this Web site:

<http://www.ibm.com/ibm/environment/products/rohs.shtml>

Archived



## Hardware Management Console

The z9 EC Hardware Management Console (HMC) is a closed platform that only supports the HMC application and will not allow the installation of any other applications.

The z9 EC HMC does not support running the Sysplex timer or ESCON director application. If you currently have these applications running on a prior generation HMC, you will need to purchase a new console to control these devices or keep the prior generation HMC for the single task of managing these devices.

# HMC support for Server Time Protocol

With the Server Time Protocol functionality, the role of the HMC is extended to provide the user interface to manage the Coordinated Timing Network. Management of STP-configured servers requires the HMC application V2.9.1.

In a Mixed CTN, the HMC is used to do the following:

- ▶ Initialize or modify the CTN ID, and ETR port states.
- ▶ Monitor the status of the CTN.
- ▶ Monitor the status of the coupling links initialized for STP message exchanges.

In an STP-only CTN, the HMC is used to do the following:

- ▶ Initialize or modify the CTN ID.
- ▶ Initialize the time, manually or by dialing out to a time service, so that the Coordinated Server Time can be set to within 100 ms of an international time standard, such as UTC.
- ▶ Initialize the Time Zone offset, Daylight Saving Time offset, and Leap second offset.
- ▶ Schedule periodic dial-outs to a time service so that CST can be steered to the international time standard.
- ▶ Assign the roles of Preferred, Backup, and Current Time Servers, as well as Arbiter.
- ▶ Adjust time by up to +/- 60 seconds.
- ▶ Schedule changes to the offsets listed. STP can automatically schedule Daylight Saving Time, based on the selected Time Zone.
- ▶ Monitor the status of the Coordinated Timing Network.
- ▶ Monitor the status of the coupling links initialized for STP message exchanges.

## External Time Source

For specific requirements to provide accurate time relative to some external time standard for data processing applications, using an External Time Source (ETS) function may be considered.

In a Mixed CTN, the ETS function is provided by the Sysplex Timer. In an STP-only CTN, the ETS function is provided by dialing out from the HMC to a time service, such as the *Automated Computer Time Service (ACTS)* or an international equivalent. A modem attached to the HMC is required to perform this function. Time (CST) can be initialized for the STP-only CTN to within +/- 100 ms of an international time standard, such as UTC. After CST has been initialized to UTC, a periodic dial-out is needed to the time service, either manually or automatically, in order to maintain the accuracy of CST.

The HMC and SE consoles support automatic retrieval of the time from a time service and automatic update of CST on a scheduled basis. One of the following can be requested:

- ▶ An immediate dial-out
- ▶ A single scheduled operation at a specified date and time
- ▶ A recurring scheduled operation that occurs at a specified frequency

Setting up a schedule to dial out to the time service automatically is done using the Scheduled Operations option at the Support Element using Single Object Operations.

At the scheduled time, the SE requests the HMC to dial out to the time service; the HMC sends the information obtained from the time service to the SE, which in turn sends the update to STP. STP makes gradual adjustments by steering CST to the time obtained from the external time source.

### **Support Element (SE)**

In an STP-only CTN, dial-out is possible from the HMC to adjust Coordinated Server Time (CST) to a time standard such as UTC. The SE supports automatic retrieval of time from a time service according to a schedule set by the user. The dial-out schedule must be set up using the SE's scheduled operations task.

The SE's disk storage provides the means for STP to initialize the configuration and timing parameters. The SE keeps a record of changes to the CTN, such as network configuration, CTN ID, and time parameters. The updates are preserved through a Power-on-Reset.

## **Remote operations**

The z9 EC HMC application simultaneously supports one local user and any number of remote users. A remote operation is considered as any request coming from outside the local subnet.

Remote operations uses the same Graphical User Interface (GUI) used by a local HMC operator. There are two ways to perform remote manual operations:

- ▶ Using a remote HMC
- ▶ Using a Web browser to connect to a local HMC

The choice between a remote HMC and a Web browser connected to a local HMC is determined by the scope of control needed. A remote HMC defines a specific set of managed objects that will be directly controlled by the remote HMC, while a Web browser to a local HMC controls the same set of managed objects as the local HMC. An additional consideration is communications connectivity and speed. LAN connectivity provides acceptable communications for either a remote HMC or Web browser control of a local HMC, but dialup connectivity is only acceptable for occasional Web browser control.

Choosing the best option involves understanding your remote control needs and use patterns. Figure A-1 shows a sample configuration for each option.

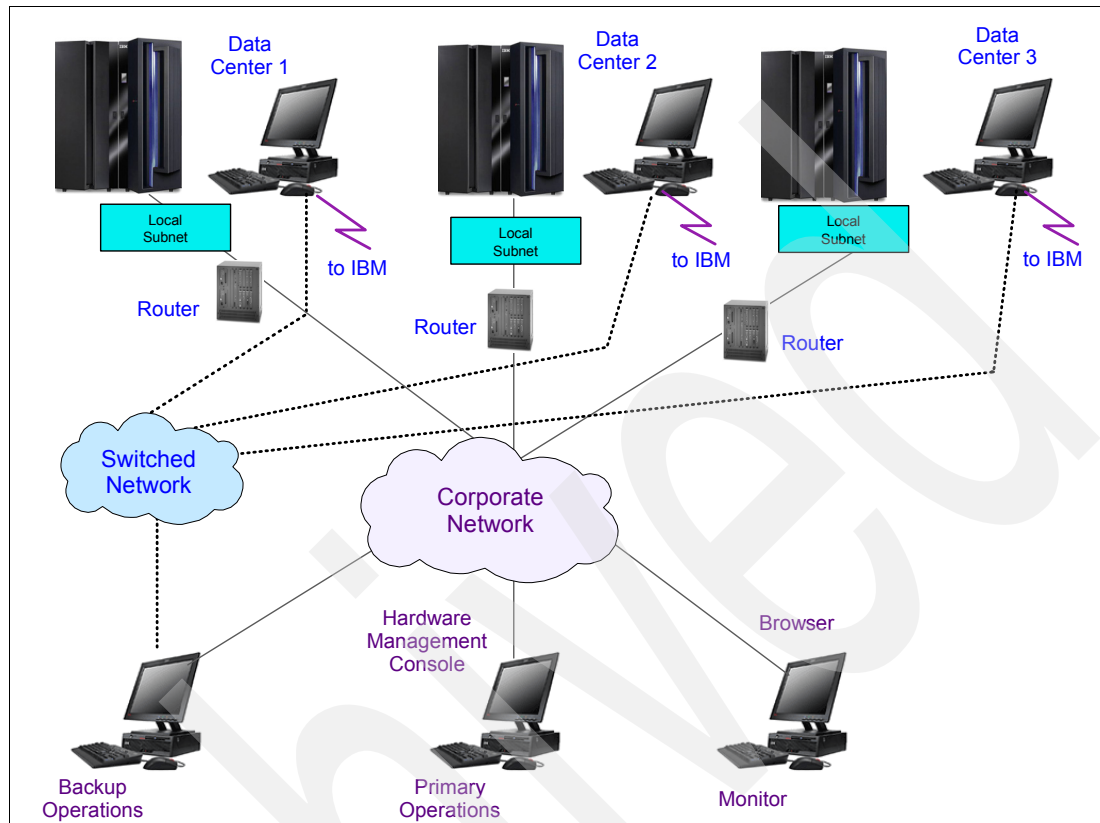


Figure A-1 Remote operation configuration example

Because of the functionality provided by the HMC, it is important that all servers and CFs in a Mixed or STP-only CTN be defined objects to any HMC that is to be used to manage the Coordinated Timing Network. Remote HMCs should also define all servers and coupling facilities in the CTN.

### Using a Hardware Management Console

A remote HMC gives the most complete set of functions because it is a complete Hardware Management Console; only the connection configuration is different from a local Hardware Management Console. As a complete HMC, it requires the same setup and maintenance as other HMCs. A remote HMC needs LAN TCP/IP connectivity to each Support Element to be managed. Therefore, any existing customer-installed firewall between the remote HMC and its managed objects must permit communications between the HMC and SE. The remote HMC also requires connectivity to IBM or another HMC with connectivity to IBM for service and support.

### Using a Web browser

Each HMC contains a Web server that can be configured to allow remote access for a specified set of users. When properly configured, an HMC can provide a remote user with access to all the functions of a local HMC except those that require physical access to the diskette or DVD media. The user interface on the remote HMC is the same as the local HMC and has the same constraints and availability as the local HMC.



The Web browser can be connected to the local HMC using either a LAN TCP/IP connection or a switched, dial, or network PPP TCP/IP connection. Both types of connection can only use encrypted (HTTPS) protocols, as configured in the local HMC. If a PPP connection is used, the PPP password must be configured in the local HMC and in the remote browser system. Logon security for a Web browser is provided by the local HMC user logon procedures. Certificates for secure communications are provided, and can be changed by the user.

### **Automated operations**

As an alternative to manual operations, it is possible to allow a computer to interact with the consoles through a programmable interface, or API. The automated interface allows a program to monitor and control the hardware components of the system in the same way a human can monitor and control the system. The HMC APIs provide monitoring and control functions through TCP/IP SNMP to an HMC. These APIs provide the ability to get and set a managed object's attributes, issue commands, receive asynchronous notifications, and generate SNMP traps. For additional information about APIs, see the *System z Application Programming Interfaces*, SB10-7030.

The automated interfaces are used by various automation products, including Tivoli® System Automation for z/OS - Processor Operations.

If your automation product is using the IBM-supplied APIs to communicate to the HMC, updates for these are found on the IBM resource link Web site:

<http://www.ibm.com/servers/resourceLink>

### **Remote support facility (RSF)**

There is a capability in the z9 EC HMC to use an Internet connection rather than an analog phone line for RSF connections. The benefit is a reliable high speed data transfer between the console and the IBM Service support system. This connection uses the customer supplied LAN infrastructure, and is an outbound only high grade SSL connection to port 443 of one of four IBM servers. This connection can pass through customer provided NAT (Network Address Translation) firewalls. When enabling this feature, there is a test button to check that the HMC has internet access to the IBM support system.

### **HTTPS proxy for network-based RSF connections**

The network-based Remote Support Facility (RSF) connection can optionally pass through a customer-supplied Hypertext Transfer Protocol over Secure Socket Layer (HTTPS) proxy system for even greater security. The Hardware Management Console (HMC) provides a choice of Secure Sockets Layer (SSL) (https) connections to IBM through a phone (modem) based RSP connection or a network based (Internet) connection.

## **HMC Console support**

In this section, we discuss HMC Console support.

### **HMC Integrated 3270 Console**

For z/VM, the requirement to invest in console hardware, such as the 2074, just to have a z/VM console, has been seen as overkill. Thus, a solution for those users has been to lower the cost of acquisition by offering an alternative solution. The integrated 3270 Console support meets that requirement.

On the HMC workplace, there is an icon to open this 3270 window. One HMC at a time can use the function; however, there is support to switch the function from one HMC to another.

With the function, there is also a highly customizable keyboard mapping support, to alleviate the ASCII-to-EBCDIC mapping. This function uses the SCLP hardware interface to connect to the operating system, and support for the function goes back to z/VM V4R4.

### HMC Integrated ASCII Console support

As for the 3270 integrated console support, a similar requirement exists for customers wanting to run Linux in a logical partition on a System z server. The Integrated ASCII Console support meets that requirement. One HMC at the time can use the function to have a Linux terminal running, and there is a support for switching the function from one HMC to another. HMC Integrated ASCII Console is also supported by z/VM V5R3.

### Customizable console data mirroring

If you require having uniform HMC information across your HMCs for data, such as passwords, password rules, and user settings, IBM has implemented a *data mirroring* function for the HMC setup. A user can customize an HMC, and then associate other HMCs in the configuration to the same customized data (see Figure A-2).

As a result, all associated HMCs will be started with the same data.

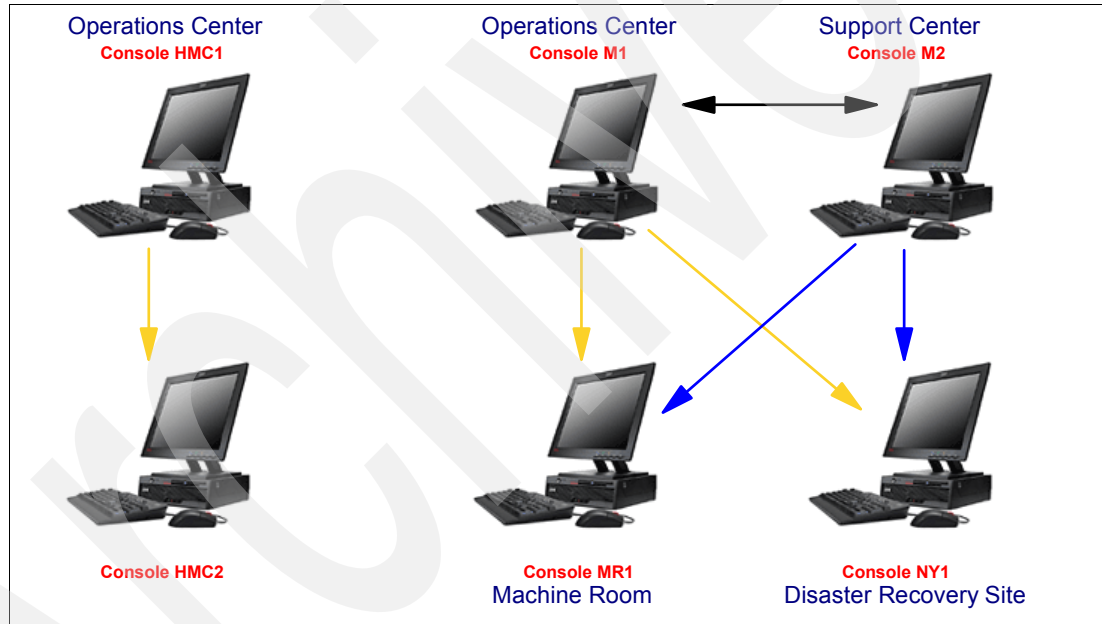


Figure A-2 HMC data replication options

## HMC application

In this chapter, we want to discuss some features in HMC application, which were not available before System z9 and which can be of interest for readers who are familiar with the previous HMC.

### Security

Security has been enhanced on the z9 EC HMC.

As access to the z9 EC HMC is provided by an HTML-based user interface, the HMC application has enabled current browser encryption techniques for enhanced security. In fact, all remote browser access is forced to use Secure Sockets Layer (SSL) encryption when

accessing the HMC. Only the local user interface is allowed to make use of non-encrypted access, since it is inherently a secure environment (closed platform).

Since SSL encryption is required for all remote access to the HMC, a certificate is required to provide the keys used for this encryption. The HMC provides a self-signed certificate that allows for this encryption to occur. The Certificate Management task is provided to manage the certificates used by the HMC. Additionally, a dynamic and integrated firewall controls and limits all communication and connectivity to the HMC application.

### **Lightweight Directory Access Protocol (LDAP) support**

LDAP support for HMC user authentication allows a user to configure their HMCs to use a LDAP server to perform user ID and password authentication at logon time. This function allows the use of the current user ID and password policy for HMC user IDs and passwords, and provides one centralized user ID and password control mechanism to help meet the user's corporate security guidelines.

The user ID is defined on the HMC along with the roles to be given to the user ID. HMC settings related to the user ID will continue to reside on the HMC, and the LDAP directory will be used to authenticate the user, thus eliminating the need to store the user ID's password locally. SSL and non-SSL connections to the LDAP server are supported. This function is designed to assist system administrators to easily create HMC user IDs matching existing company user names, thus eliminating the need to create and distribute passwords, since this is already being managed by the corporate control mechanism.

It is designed to improve management and audit functions for HMC user IDs and passwords.

### **System Activity Display with Power Monitoring**

Power Monitor is an additional function for SAD on HMC. It displays Watts and BTUs per hour as well as cooling air input temperature. It is designed to help verify power consumption for currently installed System z9 EC servers.

### **z/VM integrated systems management**

z/VM integrated systems management for the System z9 EC and z9 BC Hardware Management Console (HMC) provides out-of-the-box integrated GUI-based basic management of z/VM guests. The HMC will automatically detect z/VM images.

The z/VM integrated systems management capability supports the following image management functions: activate, deactivate, and display guest status.

### **Optional TCP/IP connection for the Application Programming Interface**

The System z9 HMC Application Programming Interface (API), when enabled, allows remote systems management of the System z9 hardware through systems management applications. This API uses the industry-standard Simple Network Management Protocol (SNMP) as the access mechanism. In the past, only User Datagram Protocol (UDP) has been supported for SNMP API communication. Today, both UDP and TCP/IP are supported. TCP/IP may be preferred where a firewall must be crossed or where a busy or unreliable network makes TCP/IP guaranteed delivery desirable.

Archived



## CHPID mapping tool

On the z9 EC, there is no default set of CHPID. CHPID mapping is performed using the Hardware Configuration Dialog or IOCP, and optionally the CHPID Mapping Tool (CMT).

The z9 EC flexible CHPID numbering helps you to keep I/O definitions the same on system upgrades that have different channel card placements or CHPID number assignments, if desired.

The CHPID Mapping Tool is intended for customer use. Since most of the functions and input require a high degree of knowledge regarding the environment, it is extremely difficult for an IBM representative to have the customer's detailed level of knowledge for the entire configuration.

## CMT requirements

The minimum requirements to run CMT are:

1. You must have access to the WWW with a browser that is at least:

- Internet Explorer® 5.0
- Netscape Navigator 4.7

Your browser must be configured with Java Script and have cookies enabled.

2. Resource Link ID

Before using CMT, the user must have a valid user ID and password for Resource Link. The URL for Resource Link is:

<http://www.ibm.com/servers/resourceLink>

There is an option on the Welcome window to obtain a user ID and password for Resource Link.

When you access the Resource Link Web site, log in and select **Tools** and the **CHPID Mapping Tool**. There are several security measures in use by Resource Link that protect your hardware configuration data.

3. CCN number for the new server

When the server to be mapped is configured and sent to the manufacturing database from the IBM configurator (e-config), a CCN number is associated with the configuration. This CCN number appears on the output listing of the configurator, and it *must* be used in order for the mapping function to be able to identify the appropriate server configuration. Users should ask their IBM representatives to verify that they have the latest CCN number associated with the machine order.

4. Java Runtime Environment (JRE™)

The stand-alone CMT requires a minimum runtime environment of Java 1.3.0. This can be verified through the following command from the command prompt:

```
java -version
```

The Java Runtime Environment is part of the CMT download.

## CMT purpose and description

The intent of the CMT is to ease the installation of z9 EC. It is also intended for making changes after installation to an already installed z9 EC, either to make slight changes to the mapping or as part of an MES action to add or remove channel features on the processor.

Advantages of CHPID mapping are:

- ▶ The existing IOCP definitions for CHPID assignments to control units can be maintained. This minimizes any changes that might need to be made to in-house documentation, cable labels, or HCD definitions.
- ▶ A numbering scheme can be implemented for associating different device types to ranges of CHPID addresses. For example, you might want to have all DASD devices within a certain CHPID range.

There are two different methods for using CMT for the z9 EC. These are manual mapping or availability mapping. They are discussed in more detail in the sections below.

## z9 EC CHPID mapping

The z9 EC does not have default CHPIDs assigned to ports as part of the initial configuration process. It is the customer's responsibility to perform these assignments by using HCD/IOCP definitions and, optionally, the CHPID Mapping Tool (CMT). One of the results of using CMT is an IOCP source that will map the defined CHPIDs to the corresponding PCHIDs of the server. There is no *requirement* to use CMT; you can assign CHPIDs to PCHIDs directly in an IOCP source or through HCD. However, this is a very cumbersome process for larger configurations. If you choose to do manual assignment of CHPIDs to PCHIDs (using HCD or IOCP), it is your responsibility to distribute CHPIDs among the physical channel card ports (PCHIDs) for availability and performance. The objective of CMT is to assist in performing these tasks.

Figure B-1 shows a diagram containing the suggested steps needed to define a new z9 EC I/O configuration.<sup>1</sup> This section focuses on describing the functions associated with CMT, the input that is needed, and the output that is created.

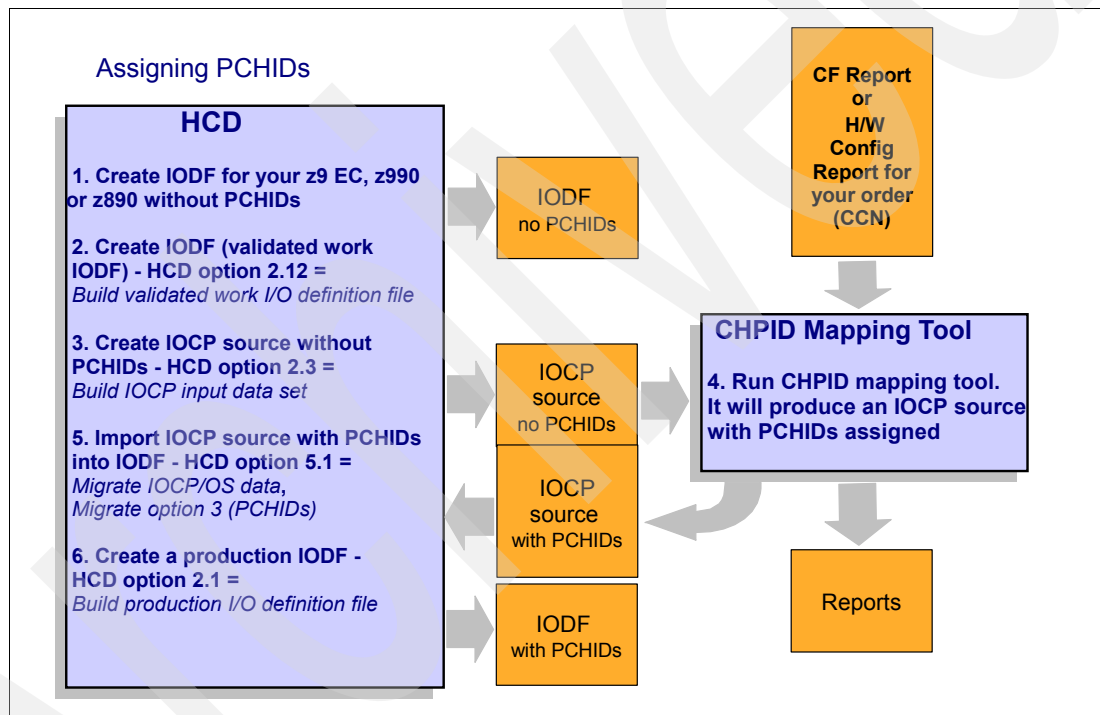


Figure B-1 PCHID assignment on the z9 EC

### Mapping function

CMT will map the CHPIDs from an IOCP source to Physical Channel Identifiers (PCHIDs), which are assigned to the I/O ports. These PCHID assignments are fixed and cannot be changed. The process used for determining the PCHID assignments is described below. A list of PCHID assignments for each hardware configuration is provided in the PCHID report that is available when z9 EC hardware is ordered. There are no default CHPID assignments. CHPIDs can be mapped by importing the IOCP source into the CHPID Mapping Tool. The IOCP source must be built for the new hardware order.

<sup>1</sup> The figure shown here covers a newly installed z9 EC. For an upgrade from a z990 to the process would be different, because we would want the PCHIDs from the existing z990 configuration to remain in the IOCP input. Only the new channels part of the z9 EC upgrade would not have PCHIDs defined.

## PCHID assignments

The purpose of the CHPID Mapping Tool is to provide a method of assigning CHPIDs to I/O ports in a way that avoids attaching critical paths to a single-point-of-failure. The first step in doing this is done by the configurator by assigning an identifier to each port that can later be associated with a CHPID. PCHIDs are assigned for you when your hardware is manufactured.

If a 16-port ESCON card were installed in card slot 1 of I/O cage 1, the first port would be assigned PCHID 100, the second port would be assigned PCHID 101, and so on. A two-port card would use the first two PCHIDs assigned to its slot and the rest of the 16 PCHID numbers for that slot would be unused.

Both manual mapping and availability mapping require the following input file to the CMT:

- ▶ Hardware Configuration report. This can be either:
  - Supplied by your IBM Representative (The output from eConfig and has the extension .CFR.)
  - Downloaded from Resource Link (This also has the extension .CFR.)
- ▶ A validated z9 EC, IOCP source

## Configuration files and PCHID assignment changes for the z9 EC

The following information aboutly applies if you are performing:

- ▶ An upgrade from a z990 to a z9 EC.
- ▶ A MES to your z9 EC that results in a card movement.

For certain channel types, see Table B-1. Information is stored on the SE that contains configuration data for these channels. The configuration data is stored on the SE under the PCHID name. The section below describes the process to preserve this data during an upgrade.

*Table B-1 CHPIDs that contain information in configuration files*

Channel or CHPID type	Information in configuration files
OSA-Express2	Any user-specified MAC addresses and OAT tables
1000BaseT channel defined as CHPID type OSC	Console session information

For an upgrade of a z990 to a z9 EC, most of the existing channels cards are moved from the old server and installed into the new server, but are rebalanced to take advantage of the availability characteristics of the z9 EC. This rebalancing also takes into consideration any new channels that may be added as part of the upgrade. The configurator creates machine configuration data and reports that contain information about which PCHIDs have been relocated.

During this upgrade, the idea is to preserve this configuration data for the card types listed in Table B-1, even though the CHPID Mapping Tool, HCD, or IOCP can all override the default PCHID assignments.



The following section explains how the system preserves the config files on an upgrade from a z990 to a z9 EC. However, it is ultimately your responsibility to have a record or backup of the customization data stored in config files. You should always make a backup record:

- ▶ For OSA-Express2 channels, record all user-assigned MAC addresses using the Display or alter MAC address function in Card specific advanced facilities or user-specified OSA Address Tables (OAT).

**Note:** A better method is to use OSA/SF.

- ▶ For CHPIDs defined as OSC (OSA-ICC), use the Export source file function in the Manual configuration options.
- ▶ For FCP ACT, ensure the access rights source is current and available to the privileged Linux image (image with access to the unit address of 0xFC or 0xFD).

The configurator creates machine configuration data and reports that contain information about which PCHIDs have been relocated. The information contained in the configurator data is used to create a CD that the service representative uses as part of the installation (or MES) activities of the new z9 EC. This CD supports the relocation for channel types that have specific files containing customization data.

It is important that the files containing the customization data be renamed to the new PCHID value. As part of the installation (or MES) process, the service representative is presented with a window (Migrate Channel Configuration Files) that shows the movement of PCHIDs based on the information contained in the manufacturing-provided CD. This function provides the capability to automatically rename the files so that they reflect the new PCHID values (so that the customization data is not lost and follows the physical movement for those cards). If you used the CHPID Mapping Tool to perform the CHPID to PCHID assignments (a process that is recommended), then the service representative need only accept the values on the window and the files are copied correctly.

For customers who choose to not follow the recommendations (those who do not use the CMT CHPID Mapping Tool or who override the CHPID Mapping Tool default assignments), the customer and service representative must team together and develop their own from-to-PCHID list; manual entry can be used to override the CD.

It is strongly recommended that you use the CHPID Mapping Tool to configure the CHPID-to-PCHID assignments because the tool ensures that logical CHPIDs are assigned to the new PCHID values. The tool has been changed to handle the previously described issues. When the appropriate files have been loaded into the tool (that is, CFReport file and IOCP file with PCHID assignments for the z990 or for the current z9 EC, if this is an MES to a z9 EC), the tool assigns new PCHID values to those affected CHPIDs based solely on the physical movement of the channel cards. If you use the manual mode, these PCHIDs have an identifier (config file). You are not recommended to change these assignments.

After you have completed working with the CHPID Mapping Tool, one of the available reports (see Figure B-2) will identify all the CHPIDs that have new PCHID values. A subset of these may have had config files associated with the old PCHIDs. The service representative can use this report to verify or change the 'TO PCHID' column during of the Migrate Channel procedure.

### List of CHPIDs having modified PCHID assignments

CHPIDs	Previous PCHID	Previous Location	F/C	Current PCHID	Current Location	F/C
0.00	017	....	....	Not Assigned		
0.01	01B	....	....	Not Assigned		
0.02	027	....	....	Not Assigned		
0.03	02B	....	....	Not Assigned		
0.06	6A0	....	....	Not Assigned		
0.07	6A1	....	....	Not Assigned		
0.0B	201	....	....	Not Assigned		
0.0E	380	Z01BLG10J.00	2319	Not Assigned		
0.15	220	....	....	190	A01BLG11J.00	1366

Figure B-2 CMT CHPID report

Windows® are available in the CMT tool that show which CHPIDs contain configuration files (see Figure B-3).

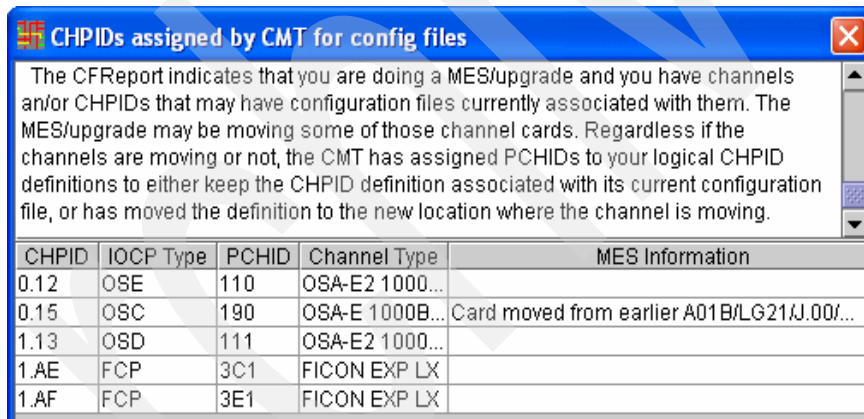


Figure B-3 CMT Configuration file screen

Information about CHPIDs that may contain configuration files is contained in the source column field in the CMT tool (Figure B-4).

Row #	Book/Fanout/Slot/Jack	Frame/Cage	Slot/Port	PCHID	ChannelType	CHPID	SOURCE
101	2/D5/1	Z01B	LG15J.00	3C0	FICON EXP LX		
102	2/D5/1	Z01B	LG15J.01	3C1	FICON EXP LX	1.AE	Config File
103	1/D5/1	Z01B	LG16J.00	3D0	FICON EXP LX		
104	1/D5/1	Z01B	LG16J.01	3D1	FICON EXP LX		
105	2/D5/1	Z01B	LG17J.00	3E0	FICON EXP LX		
106	2/D5/1	Z01B	LG17J.01	3E1	FICON EXP LX	1.AF	Config File
107	1/D5/1	Z01B	LG18J.00	3F0	ESCON	0.42	IOCP
108	1/D5/1	Z01B	LG18J.01	3F1	ESCON	1.53	IOCP

Source Column indicating CHPIDS that may contain Config Files

Figure B-4 CMT Source column information

### Manual mapping

The manual method can be used to define the relationships between each logical CHPID and physical ports on the server. There is availability checking and the accuracy of the mapping with HCD definitions is dependent on the user's knowledge of the server availability characteristics. However, after you have completed manual mapping and have assigned PCHIDs for all your CHPIDs, you can use the availability function to see if there are any intersects in your mapping.

### Availability mapping for z9 EC

Availability mapping is the recommended method for mapping. This method allows you to input the IOCP source for the proposed machine configuration and then define the order in which channel and control unit mapping should be done. This function takes into account the availability characteristics of the server and will insure that the highest levels of availability will be achieved by assigning channel paths to avoid a single-point-of-failure in that server.

While using the tool, you will have the ability to switch between manual and availability mapping. You could, for example, map your CHPIDs with the availability mapping option and then make changes manually.

You must provide a copy of the system's IOCP source and define the priorities for channels and control units in the configuration. The CHPID Mapping Tool then decides how to assign CHPIDs to the I/O ports and provides a CHPID Report for a system that will provide maximum system I/O availability. Maximum availability is achieved by distributing channel paths across different channel cards and STI links. In this way, a failure to one channel card or STI link will not affect the availability of a device.

### Setting Control Unit priority

Priorities are assigned (from 0001 to 9999) for each control unit in the IOCP source. More than one CU can be assigned the same priority. Assigning the same priority to more than one CU means that these units will be mapped together for availability. There are good reasons to do so:

- ▶ If one device serves as a backup for another device, you should assign the same priority to both of their control units, for example, CUs defined for CTC communication and OSA CUs going to the same network.

- ▶ All operator or system consoles should have the same priority.
- ▶ If multiple control units are used to provide multiple paths to devices, you should assign the same priority to these control units.
- ▶ You may want to group CHPIDs without CUs, for example, peer Coupling Links used by a CF image to a common CPC, or maybe extra channels defined in anticipation of future requirements.

When CMT maps control units with the same priority, it will assign them with the emphasis on availability.

Once you have decided which control units should be mapped as a group, you should assign the lowest number priority to the control units in the group that you want to be mapped first. The tool will map the CUs with priority 0001 first. It is advantageous to leave gaps in your group numbering (for example, 0010, 0020, 0030, and so on). This allows room for additional groups to be created in the event of a conflict. Any CUs that do not have a priority will be assigned a priority *afterwards* in path to CU order. The CU with the highest number of paths will take priority over one with less paths. Multiple CUs with the same number of paths will be further prioritized by CU number (the lowest CU number being given highest priority).

Once priorities have been entered, the availability mapping can be processed. When the tool completes processing, it will also display a list of port assignments not included in the IOCP source if there is more new hardware than was defined in the IOCP source.

Once the priorities have been entered, they can be processed by the CMT. The following options are presented:

- ▶ Reset CHPIDs assigned by Availability.
- ▶ Reset CHPIDs assigned by Manual remap.
- ▶ Reset CHPIDs assigned by IOCP.
- ▶ Reset CHPIDs assigned for config files (this is not recommended).

You can choose to reset any, none, or all of the above categories of assigned CHPIDs. Selecting none of the options will only process the unassigned CHPIDs.

**Attention:** Choosing to reset CHPIDs assigned by IOCP may require you to recable your hardware.

The priority settings are added to the IOCP source as comments. They will be reused the next time the IOCP source is loaded in the CMT.

### ***Intersects***

The tool displays intersects on the next window. Intersects are potential availability problems detected by the tool. Reason codes for intersects are provided and an explanation of the codes is displayed at the bottom of the window. Remember that these codes refer to channels in the same group, where this group is defined solely by the same priority codes to the mapping tool. Briefly, the warnings are:

- ▶ D = Assigned channels are on the same daughter card.
- ▶ C = Two or more assigned channels use the same channel card.
- ▶ S = Greater than half the assigned channels use the same STI.
- ▶ M = All the assigned channels are supported by the same MBA.
- ▶ B = Greater than half the assigned channels are connected to the same book.

These could be considered informational warnings and not errors. For example, all the channels to a CU from the same MBA may not be a concern because of Redundant I/O Interconnect and the fact that an alternate path from a different MBA is available in case the primary MBA fails. However, all the ICBs for a given CF on the same MBA should be a cause for concern.

Eventually, the tool will run out of highly available places to assign CHPIDs and will resort to plugging more than one control unit on the same channel card, STI, or MBA group. A possible cause for intersects may be that prior mapping to other groups left a small number of unassigned ports.

Intersects might be corrected by assigning the group that displays an intersect to a lower numbered priority or dividing the group into smaller groups when possible. If you find an intersect to be unacceptable and regrouping and reprioritizing does not resolve it, you may need more hardware.

### ***CMT output***

The CMT output for a z9 EC differs to that provided for the previous servers. PCHIDs are in fixed physical locations, mapped to CHPIDs by the CMT, and assigned to the IOCP. Consequently, the output of the CMT for a z9 EC is as follows:

- ▶ The tailored reports. All reports should be saved for reference. The Port Report sorted by location should be supplied to your IBM Service Representative for reference.
- ▶ An IOCP with PCHIDs mapped to CHPIDs by Channel Subsystem. This IOCP source can then be migrated back into HCD and a production IODF can be built.

**Note:** If the IOCP source was exported from HCD for CMT input, it must be migrated back into HCD. It cannot be used directly by IOCP.

### **References**

See the following sources for more information:

- ▶ *CHPID Mapping Tool*, GC28-6825, also available on Resource Link Web site.
- ▶ The Resource Link Web site:  
<http://www.ibm.com/servers/resourceLink>

Archived

## Fiber cabling services

This appendix describes the IBM Networking Services fiber cabling services options offered by IBM Global Services to customers.

The following topics are covered:

- ▶ Fiber cabling services options
  - Option 1: Fiber-optic jumper cabling package
  - Option 2: Fiber-optic jumper cable migration and reuse
  - Option 3: Fiber-optic jumper cabling and installation
  - Option 4: System z fiber-optic trunk cabling package
  - Option 5: Enterprise fiber cabling services
- ▶ “Summary” on page 278
- ▶ “References” on page 279

## Fiber cabling services options

When integrating a System z server into a data center, an IBM Installation Planning Representative (IPR) provides planning assistance to customers for equipment power, cooling, and the physical placement of the server.

However, the fiber-optic cable planning and connecting of the System z server channels to I/O equipment, Coupling Facilities, networks, and other servers, is a customer's responsibility, both for new server installations and server upgrades.

**Note:** Customers, especially those with complex system integration requirements, may request connectivity assistance from IBM. See IBM Networking Services fiber cabling services.

Customers with the resources and personnel to plan and implement their own connectivity, or those with less complex system configurations, can consult *System z9 EC Installation Manual for Physical Planning*, GC28-6844 to help them determine and order the required fiber-optic cabling. It is available on the IBM Resource Link Web site:

<http://www.ibm.com/servers/resourceLink>

## IBM Networking Services fiber cabling services

IBM Global Services offers customers the option of engaging IBM to help plan and implement their enterprise connectivity with a suite of services offerings. The IBM Networking Services fiber cabling services, being offered by IBM Global Services, provides five options under System z Fiber Cabling Services (options 1 to 3) and Enterprise Fiber Cabling Services (options 4 and 5).

### System z Fiber Cabling Services

The System z Fiber Cabling Services options are:

► Option 1: Fiber-optic jumper cabling package

This option provides planning, new fiber-optic cables (jumper cables, conversion kits, and mode conditioning patch (MCP) cables), installation, and documentation:

- a. IBM does the System z server cabling configuration and connection planning, including:
  - Analysis of the customer's server channel configuration, connecting I/O devices, and any existing fiber-optic cables, to determine the required fiber-optic cables
  - Analysis of the server e-config report with CHPID/PCHID placement report
  - Creating a bill of materials (BOM) of the required fiber-optic jumper cables, conversion kits, and MCP cables
- b. IBM does the fiber-optic cable ordering, labeling, and installation based on the customer's channel configuration.
- c. IBM documents the new server channel connections.



► Option 2: Fiber-optic jumper cable migration and reuse

This option provides planning, reuse of existing fiber-optic cables, and documentation:

- a. IBM does the System z server cabling configuration and connection planning, including:
  - Analysis of the customer's server channel configuration, connecting I/O devices, and existing fiber-optic cables
  - Analysis of the configurator CHPID/PCHID placement report
- b. IBM does the relabeling and rerouting of existing fiber-optic cables, including:
  - Sorting, organizing, and relabeling of existing fiber-optic cables
  - Rerouting of existing fiber-optic cables under the raised floor to the appropriate server frame openings
  - Reconnecting the existing fiber-optic cables to the appropriate server channel ports based on the customer's channel configuration
- c. IBM documents the new server channel connections.

► Option 3: Fiber-optic jumper cabling and installation

This option provides new fiber-optic cables (jumper cables, conversion kits, and mode conditioning patch (MCP) cables), installation, and documentation:

**Note:** Planning is a customer responsibility.

The customer is responsible for the analysis of their server channel configuration, I/O device connections, and any existing fiber-optic cables, to determine the required fiber-optic cables.

The customer provides a bill of materials (BOM) to IBM of the required fiber-optic jumper cables, conversion kits, and MCP cables. The customer provides IBM with the server plugging information, based on this analysis.

- a. IBM does the fiber-optic cable ordering, labeling, and installation.
- b. IBM documents the server channel connections.

## Enterprise Fiber Cabling Services

The Enterprise Fiber Cabling Services options are:

► Option 4: System z fiber-optic trunk cabling package

This option provides planning, fiber-optic trunking commodities (trunk cables, harnesses, panel-mount boxes), installation, and documentation:

- a. IBM does the System z server cabling configuration and connection planning, including:
  - Analysis of the customer's server channel configuration, connecting I/O devices, and any existing fiber-optic cables or trunking, to determine the required fiber-optic trunking commodities
  - Analysis of the server configurator CHPID/PCHID placement report
  - Creating a bill of materials (BOM) of the required fiber-optic trunking commodities
- b. IBM does the fiber-optic trunking commodities ordering, labeling, and installation based on the customer's channel configuration.
- c. IBM documents the new fiber-optic trunking and server channel connections.

This option also supports the Fiber Quick Connect (FQC) feature, which provides factory installed fiber-optic harnesses for ESCON channels in System z servers. See “Fiber Quick Connect (FQC)” on page 277.

This option does not provide Enterprise planning.

► Option 5: Enterprise fiber cabling services

This option provides Enterprise level (Data Center, SAN, LAN) planning, fiber-optic trunking commodities (trunk cables, central patching cabinets, harnesses, and panel-mount boxes), installation, and documentation:

- a. IBM does the Enterprise cabling configuration and connection planning, including:
  - Analysis of the customer’s Enterprise configuration, connecting I/O devices, and any existing fiber-optic cables or trunking, to determine the required fiber-optic trunking commodities.
  - Analysis of the server configurator CHPID/PCHID placement reports.
  - Create a bill of materials (BOM) of the required fiber-optic trunking commodities.
- b. IBM does the fiber-optic trunking commodities ordering, labeling, and installation based on the customer’s channel configurations.
- c. IBM documents the new fiber-optic trunking and server channel connections across the Enterprise.

This option also supports the Fiber Quick Connect (FQC) feature, which provides factory installed fiber-optic harnesses for ESCON channels in System z servers. See “Fiber Quick Connect (FQC)” on page 277.

The Enterprise Fiber Cabling Services options utilize IBM Fiber Transport System (FTS) products that aid in the migration from an unstructured fiber-optic cabling environment to a structured, flexible, and easy-to-manage fiber-optic cabling system.

With the proliferation of industry-standard fiber-optic interfaces and connector types, the management of a data center fiber-optic cable environment is becoming increasingly important.

**FTS benefits**

The most apparent benefit of the structured trunking system is the large reduction in the number of fiber-optic cables under the raised floor. The smaller number of cables makes documenting what cables go where much easier. Better documentation means tracing a fiber-optic link is much easier during problem determination and when planning for future growth.

Table C-1 shows the advantages of a structured cabling system over a non-structured cabling system.

*Table C-1 Benefits of the structured cabling system*

<b>Non-structured cabling</b>	<b>Structured cabling</b>
Unknown cabling routing.	Known cable pathways.
No cable documentation system.	Defined cable documentation.
Unpredictable impact of moves, adds, and changes.	Reliable outcome of moves, adds, and changes.
Every underfloor activity is an unknown risk.	Underfloor activity can be planned to minimize risk.

FTS is a long-term connectivity solution that provides an organized network of cabling options for future equipment reconfiguration and addition. Changes can be performed with minimal floor disruptions and are accomplished by rearranging short jumpers at the panel-mount boxes in the CPL.

Each FTS design is unique in that it is based on physical room characteristics and equipment configurations and placement preferences.

FTS provides the following benefits:

- ▶ Proven connectivity:
  - FTS components are IBM-tested, approved, and sold under the IBM logo.
  - FTS trunk-mounting kits are designed and tested with System z servers and devices to prevent any impact on equipment operation or serviceability.
  - Multimode and single mode fiber solutions are available.
- ▶ Elimination of long jumper cables, making underfloor areas less congested and easier to manage by using direct-attach trunking
- ▶ Cost reductions and productivity gains through:
  - Faster, less disruptive installation and removal of equipment
  - Easier reconfiguration of equipment
  - More efficient use of underfloor space (potential savings in air conditioning requirements)
- ▶ Improved cable management with a reduced risk of damage to fiber cables for the following environments:
  - ESCON
  - FICON
  - Parallel Sysplex
  - Open Systems Adapter
  - Open system architectures (Fibre Channel and Gigabit Ethernet)
- ▶ Growth flexibility by being:
  - Expandable
  - Relocatable
  - A long-term solution

These benefits and potential cost savings must be assessed for each data center and its particular environment.

### ***Fiber Quick Connect (FQC)***

Fiber Quick Connect is an option in the eConfig configuration tool when ordering a new build or upgrade of a System z server

The FQC features are for factory installation of IBM Fiber Transport System (FTS) fiber-optic harnesses for connection to all ESCON channels in I/O cages of the System z servers. FTS fiber-optic harnesses enable connection to FTS direct-attach fiber-optic trunk cables.

FQC, when coupled with the Fiber Transport System (FTS) products from IBM Global Services, delivers a solution to reduce the amount of time required for on-site installation and setup of cabling, to minimize disruptions, and to isolate the activity from the active system as much as possible. FQC facilitates adds, moves, and changes of ESCON multimode fiber-optic cables in the data center and reduces fiber-optic cable installation time.

Enterprise fiber cabling services provides the direct-attach trunk harnesses, patch panels, and central patching location (CPL) hardware, as well as the planning and installation required to complete the total structured connectivity solution.

CPL planning and layout is done prior to arrival of the server on-site, and documentation is provided showing the channel layout and how the direct-attach harnesses are plugged.

FQC supports all of the ESCON channels in System z servers' I/O cages. FQC cannot be ordered for selected channels and cages within the server.

FQC is based on a quick connect/disconnect trunking strategy utilizing the 12 fiber MTP connector, so it is able to transport six ESCON CHPIDs. For example, six trunks, each with 72 fiber-optic pairs (twelve MTP connectors), can displace up to 420 fiber-optic cables, the maximum quantity of ESCON channels supported in one I/O cage on a z9 EC or z990 server. This significantly reduces ESCON cable bulk.

The MTP connector enables Fiber Quick Connect trunk cables to be installed and later disconnected and relocated very quickly. The MTP connector also enables Fiber Quick Connect to bring its fiber trunk cables directly under the covers of the System z servers, and ESCON Directors.

## Summary

Enterprise fiber cabling services provides a structured FTS fiber cabling system that combines planning, installation, and service. It consists of fiber trunk cables, direct attach harnesses, and a variety of MDF panel mounts. The installation may also include a fiber conveyance system, which is an underfloor tray system that protects the fiber-optic cables.

Each Enterprise fiber cabling services design is unique in that it is based on physical room characteristics and equipment placement preferences. It is a long-term *connectivity solution* that provides an organized network of cabling options for future equipment reconfiguration and addition, such as:

- ▶ Secure configuration environment.
- ▶ Under-floor cable control (long jumper cables can be eliminated, making underfloor areas less congested and easier to manage).
- ▶ Efficient cable management:
  - Uniform cable documentation
  - Structured cable routing
  - Simplified system configuration
  - Machine ports located in one central distribution facility
  - CHPIDs and device ports in order at patch panels
  - Factory-installed harnesses and tailgates to ensure consistent routing, allowing quick connection of fiber trunks
  - Space reserved for future growth

- ▶ Changes can be performed with minimal floor disruptions and are accomplished by rearranging short jumpers at the panel-mount boxes in the MDF.

In summary, enterprise fiber cabling services solutions can provide the following benefits:

- ▶ Complements the ESCON architecture by providing easier copper-to-fiber migration.
- ▶ Cost reductions and productivity gains through:
  - Faster, less disruptive install and removal of equipment
  - Easier reconfiguration of equipment
  - More efficient use of under-floor space (potential savings in air conditioning requirements)
- ▶ Improved cable management with a reduced risk of damage to fiber-optic cables for the following environments:
  - ESCON
  - FICON
  - Parallel Sysplex
  - Open Systems Adapter
- ▶ Growth flexibility by being:
  - Expandable
  - Re-locatable
  - A long-term solution

These benefits and potential cost savings need to be assessed for each data center and its particular environment.

## References

For further information about the IBM Networking Services fiber cabling services offered by IBM Global Services, and related topics, see:

- ▶ The IBM Resource Link Web site:  
<http://www.ibm.com/servers/resourceLink>
- ▶ *Fiber Transport Services Direct Attach Planning*, GA22-7234
- ▶ *Installing the Direct Attach Trunking System in zSeries 900 Servers*, GA27-4247
- ▶ *ESCON I/O Interface Physical Layer Document*, SA23-0394
- ▶ *Coupling Facility Channel I/O Interface Physical Layer*, SA23-0395
- ▶ *Fiber Channel Connection for S/390 I/O Interface Physical Layer*, SA24-7172
- ▶ *Fiber Optic Link Planning*, GA23-0367
- ▶ *Fiber Optic Link (ESCON, FICON, Coupling Links and OSA) Maintenance Information*, SY27-2597

Archived

# Glossary

**active configuration.** In an ESCON environment, the ESCON Director configuration determined by the status of the current set of connectivity attributes. Contrast with *saved configuration*.

**ADMF.** Asynchronous Data Mover Facility.

**allowed.** In an ESCON Director, the attribute that, when set, establishes dynamic connectivity capability. Contrast with *prohibited*.

**American National Standards Institute (ANSI).** An organization consisting of producers, consumers, and general interest groups, which establishes the procedures by which accredited organizations create and maintain voluntary industry standards in the United States.

**ANSI.** See *American National Standards Institute*.

**AP.** Adjunct processor.

**APAR.** See *authorized program analysis report*.

**Application Assist Processor (AAP).** A special processor configured for running Java applications on z9 EC, z990 and z890 class machines.

**ARP.** Address Resolution Protocol.

**authorized program analysis report (APAR).** A report of a problem caused by a suspected defect in a current, unaltered release of a program.

**BBU.** Battery Back-up Unit.

**BL.** Parallel block multiplexer channel.

**blocked.** In an ESCON Director, the attribute that, when set, removes the communication capability of a specific port. Contrast with *unblocked*.

**BPA.** Bulk Power Assembly.

**bus.** (1) A facility for transferring data between several devices located between two endpoints, only one device being able to transmit at a given moment. (2) A network configuration in which nodes are interconnected through a bidirectional transmission medium. (3) One or more conductors used for transmitting signals or power.

**BY.** Parallel byte multiplexer channel.

**CAP.** Cryptographic Asynchronous Processor.

**CAW.** channel address word.

**CBA.** Concurrent Book Add.

**CBP.** Integrated cluster bus Coupling Facility peer channel.

**CBU.** Capacity BackUp.

**CBY.** Mnemonic for an ESCON channel attached to an IBM 9034 convertor. The 9034 converts from ESCON CBY signals to parallel channel interface (OEMI) communication operating in byte multiplex mode (Bus and Tag). Contrast with *CVC*.

**CCC.** Channel control check.

**CCF.** Cryptographic Coprocessor Facility.

**CCL.** Communication Controller for Linux CCL.

**CCW.** Channel command word.

**CDC.** Channel data check.

**CEC.** Central Electronic Complex.

**central processor complex.** A physical collection of hardware that consists of Central Storage, one or more central processors, timers, and channels.

**central processor.** The part of the computer that contains the sequencing and processing facilities for instruction execution, initial program load, and other machine operations.

**CFCC.** Coupling Facility Control Code.

**chained.** In an ESCON environment, pertaining to the physical attachment of two ESCON Directors (ESCDs) to each other.

**channel.** (1) A processor system element that controls one channel path, whose mode of operation depends on the type of hardware to which it is attached. In a Channel Subsystem, each channel controls an I/O interface between the channel control element and the logically attached control units. (2) In the z/Architecture, the part of a Channel Subsystem that manages a single I/O interface between a Channel Subsystem and a set of controllers (control units).

**channel address.** In S/370™ mode, the 8 left-most bits of an input/output address that identify the channel. See also device address and input/output address.

**channel-attached.** (1) Pertaining to attachment of devices directly by data channels (I/O channels) to a computer. (2) Pertaining to devices attached to a controlling unit by cables rather than by telecommunication lines.

**channel control check.** A category of I/O errors affecting channel controls and sensed by the channel to which a device is attached. See also *channel data check*.

**channel data check.** A category of I/O errors, indicating a machine error in transferring data to or from storage and sensed by the channel to which a device is attached. See also *channel control check*.

**channel Licensed Internal Code.** That part of the Channel Subsystem Licensed Internal Code used to start, maintain, and end all operations on the I/O interface. See also *IOP Licensed Internal Code*.

**channel path (CHP).** A single interface between a central processor and one or more control units along which signals and data can be sent to perform I/O requests.

**channel path configuration.** In an ESCON or FICON environment, the connection between a channel and a control unit or between a channel, an ESCON Director, and one or more control units. See also *point-to-point channel path configuration*, and *switched point-to-point channel path configuration*.

**channel path identifier (CHPID).** In a Channel Subsystem, a value assigned to each installed channel path of the system that uniquely identifies that path to the system.

**Channel Subsystem (CSS).** Relieves the processor of direct I/O communication tasks, and performs path management functions. Uses a collection of subchannels to direct a channel to control the flow of information between I/O devices and main storage.

**channel-to-channel adapter (CTCA).** An input/output device that is used by a program in one system to communicate with a program in another system.

**check stop.** The state that occurs when an error makes it impossible or undesirable to continue the operation in progress.

**CHPID.** Channel path identifier.

**CIU.** Customer Initiated Upgrade.

**cladding.** In an optical cable, the region of low refractive index surrounding the core. See also *core* and *optical fiber*.

**CMOS.** Complementary metal-oxide semiconductor.

**CMT.** CHPID Mapping Tool.

**CNC.** Mnemonic for an ESCON channel used to communicate to an ESCON-capable device.

**command chaining.** The fetching of a new channel command word (CCW) immediately following the completion of the previous CCW.

**command retry.** A channel and control unit procedure that causes a command to be retried without requiring an I/O interrupt.

**concurrent maintenance.** Hardware maintenance actions performed by a service representative while normal operations continue without interruption.

**configuration matrix.** In an ESCON environment, an array of connectivity attributes that appear as rows and columns on a display device and can be used to determine or change active and saved configurations.

**connected.** In an ESCON Director, the attribute that, when set, establishes a dedicated connection between two ESCON ports. Contrast with *disconnected*.

**connection.** In an ESCON Director, an association established between two ports that provides a physical communication path between them.

**connectivity attribute.** In an ESCON Director, the characteristic that determines a particular element of a port's status. See *allowed*, *blocked*, *connected*, *disconnected*, *prohibited*, and *unblocked*.

**control unit.** A hardware unit that controls the reading, writing, or displaying of data at one or more input/output units.

**Coordinated Server Time (CST).** Represents the true time of a CTN.

**Coordinated Timing Network (CTN).** A collection of servers, each of which exchanges STP timekeeping messages with attached servers so that all servers in the CTN may synchronize to Coordinated Server Time (CST). The servers that make up a CTN are all configured with a common identifier, referred to as a CTN ID.

**Coordinated Timing Network ID.** Specifies the ID for the Coordinated Timing Network (CTN) that the server is participating in. The form is STP ID-ETR ID.

**core.** (1) In an optical cable, the central region of an optical fiber through which light is transmitted. (2) In an optical cable, the central region of an optical fiber that has an index of refraction greater than the surrounding cladding material. See also *cladding* and *optical fiber*.

**coupler.** In an ESCON environment, link hardware used to join optical fiber connectors of the same type.

**Coupling Facility.** A special logical partition that provides high-speed caching, list processing, and locking functions in a sysplex.



**Coupling Facility control code.** The Licensed Internal Code (LIC) that runs in a Coupling Facility logical partition to provide shared storage management functions for a sysplex.

**CP.** Central Processor.

**CPACF.** CP Assist for Cryptographic Function.

**CPC.** Central Processor Complex.

**CPU.** Central Processing Unit.

**CST Offset Dispersion:** Represents the accuracy of the CST offset such that the true CST offset falls within the bounds specified by this value. That is,  $\text{CST offset} - \text{CST-offset dispersion} \leq \text{true CST offset} \leq \text{CST offset} + \text{CST-offset dispersion}$ .

**CST Offset.** Represents the estimated difference between the TOD clock at a server and CST. At a server, the TOD clock plus the CST offset equals the estimate of CST at a server.

**CST Dispersion:** Represents the accuracy of the CST determined by a server such that the server's estimate of CST falls within the bounds specified by this value. That is,  $\text{server CST} - \text{CST dispersion} \leq \text{CST} \leq \text{server CST} + \text{CST dispersion}$ .

**CST.** Coordinated Server Time.

**CTC.** (1) Channel-to-channel. (2) Mnemonic for an ESCON channel attached to another ESCON channel.

**CTCA.** See *channel-to-channel adapter*.

**CU.** Control unit.

**CUA@.** Control unit address.

**CUADD.** Control unit logical address.

**CUoD.** Capacity Upgrade on Demand.

**Current Time Server.** Specifies if the Preferred Time Server or Backup Time Server is the current time server. The Current Time Server identifies where the time reference is currently coming from. The Current Time server cannot be the Backup Time Server unless there is a Preferred/Backup or Preferred/Backup/Arbiter configuration.

**CVC.** Mnemonic for an ESCON channel attached to an IBM 9034 convertor. The 9034 converts from ESCON CVC signals to parallel channel interface (OEMI) communication operating in block multiplex mode (Bus and Tag).

**DAT.** Dynamic address translation.

**data sharing.** The ability of concurrent subsystems (such as DB2 or IMS DB) or application programs to directly access and change the same data while maintaining data integrity.

**data streaming.** In an I/O interface, a mode of operation that provides a method of data transfer at up to 4.5 MB per second. Data streaming is not interlocked between the sender and the receiver. Once data transfer begins, the sender does not wait for acknowledgment from the receiver before sending the next byte. The control unit determines the data transfer rate.

**DCA.** Distributed Converter Assembly.

**DCAF.** Distributed Console Access Facility.

**DCM.** Dynamic CHPID Management.

**DDM.** See *disk drive module*.

**dedicated connection.** In an ESCON Director, a connection between two ports that is not affected by information contained in the transmission frames. This connection, which restricts those ports from communicating with any other port, can be established or removed only as a result of actions performed by a host control program or at the ESCD console. Contrast with *dynamic connection*. **Note:** The two links having a dedicated connection appear as one continuous link.

**default.** Pertaining to an attribute, value, or option that is assumed when none is explicitly specified.

**DES.** Data Encryption Standard.

**destination.** Any point or location, such as a node, station, or a particular terminal, to which information is to be sent.

**device.** A mechanical, electrical, or electronic contrivance with a specific purpose.

**device address.** In the z/Architecture, the field of an ESCON or FICON (FC mode) device-level frame that selects a specific device on a control-unit image.

**device number.** (1) In the z/Architecture, a four-hexidecimal-character identifier (for example, 19A0) that you associate with a device to facilitate communication between the program and the host operator. (2) The device number that you associate with a subchannel that uniquely identifies an I/O device.

**DH.** Diffie Hellman.

**direct access storage device (DASD).** A mass storage medium on which a computer stores data.

**disconnected.** In an ESCON Director, the attribute that, when set, removes a dedicated connection. Contrast with *connected*.

**disk.** A physical or logical storage media on which a computer stores data (is also sometimes referred to as a magnetic disk).

**disk drive module (DDM).** A disk storage medium that you use for any host data that is stored within a disk subsystem.

**Dispersion.** A value that contains the CST dispersion reported by the attached server. The value represents the dispersion of the TOD clock at the attached server relative to the Preferred Time Server in the synchronization path selected by the attached server.

**distribution panel.** (1) In an ESCON or FICON environment, a panel that provides a central location for the attachment of trunk and jumper cables and can be mounted in a rack, wiring closet, or on a wall.

**duplex.** Pertaining to communication in which data or control information can be sent and received at the same time. Contrast with *half duplex*.

**duplex connector.** In an ESCON environment, an optical fiber component that terminates both jumper cable fibers in one housing and provides physical keying for attachment to a duplex receptacle.

**duplex receptacle.** In an ESCON environment, a fixed or stationary optical fiber component that provides a keyed attachment method for a duplex connector.

**dynamic connection.** In an ESCON Director, a connection between two ports, established or removed by the ESCD and that, when active, appears as one continuous link. The duration of the connection depends on the protocol defined for the frames transmitted through the ports and on the state of the ports. Contrast with *dedicated connection*.

**dynamic connectivity.** In an ESCON Director, the capability that allows connections to be established and removed at any time.

**dynamic I/O reconfiguration.** A function that allows I/O configuration changes to be made nondisruptively to the current operating I/O configuration.

**dynamic storage reconfiguration.** A PR/SM LPAR function that allows central or Expanded Storage to be added or removed from a logical partition without disrupting the system control program operating in the logical partition.

**EAF.** See ETR Attachment Facility.

**EBA.** Enhanced Book Availability.

**ECC.** Error checking and correction.

**ECKD.** Extended count key data.

**EEPROM.** Electrically erasable programmable read only memory.

**EIA.** Electronics Industries Association. One EIA unit is 1.75 inches or 44.45 mm.

**EMIF.** See *ESCON Multiple Image Facility*.

**Enterprise System Connection (ESCON).** (1) A z/Architecture computer peripheral interface. The I/O interface uses z/Architecture logical protocols over a serial interface that configures attached units to a communication fabric. (2) A set of IBM products and services that provide a dynamically connected environment within an enterprise.

**Enterprise Systems Architecture/390® (ESA/390).** An IBM architecture for mainframe computers and peripherals.

**environmental error record editing and printing program (EREP).** The program that makes the data contained in the system recorder file available for further analysis.

**EPO.** Emergency power off.

**ESA/390.** See *Enterprise Systems Architecture/390*.

**ESCD.** Enterprise Systems Connection (ESCON) Director.

**ESCD console.** The ESCON Director display and keyboard device used to perform operator and service tasks at the ESCD.

**ESCM.** Enterprise Systems Connection Manager.

**ESCON.** See *Enterprise System Connection*.

**ESCON channel.** A channel having an Enterprise Systems Connection channel-to-control-unit I/O interface that uses optical cables as a transmission medium. May operate in CBY, CNC, CTC, or CVC mode. Contrast with *parallel channel*.

**ESCON Director.** An I/O interface switch that provides the interconnection capability of multiple ESCON interfaces (or FICON FCV (9032-5) in a distributed-star topology.

**ESCON Multiple Image Facility (MIF).** A function that allows logical partitions to share an ESCON channel path (and other channel types) by providing each logical partition with its own channel-subsystem image.

**ETR.** External time reference.

**ETR Attachment Facility (EAF).** A feature on a server where a Sysplex Timer is attached. For example, FC6154, FC6150, or FC6152 are EAF feature codes. Sometimes these features or facilities are standard with no feature code and sometimes they must be specifically ordered, depending on the server and the specific configuration.

**ETR ID.** Identifies the Sysplex Timer that sends ETR signals.

**ETR network ID.** Identifies the network ID of the Sysplex Timer that the ETR Attachment Facility ports are connected to.

**ETR port number.** Identifies the port number of the Sysplex Timer output port to which it sends ETR signals.

**ETR Timing Mode.** When the configuration is in ETR-timing mode, the TOD clock has been initialized to the ETR and is being stepped by stepping signals from ETR. To be in ETR-timing mode, the configuration must be part of an ETR network.

**FC-AL.** Fibre Channel Arbitrated Loop.

**FCS.** See *Fibre Channel standard*.

**FCTC.** FICON Channel-to-Channel.

**fiber.** See *optical fiber*.

**fiber optic cable.** See *optical cable*.

**fiber optics.** The branch of optical technology concerned with the transmission of radiant power through fibers made of transparent materials, such as glass, fused silica, and plastic.

**Note:** Telecommunication applications of fiber optics use optical fibers. Either a single discrete fiber or a non-spatially aligned fiber bundle can be used for each information channel. Such fibers are often called “optical fibers” to differentiate them from fibers used in non-communication applications.

**Fibre Channel standard.** An ANSI standard for a computer peripheral interface. The I/O interface defines a protocol for communication over a serial interface that configures attached units to a communication fabric. The protocol has four layers. The lower of the four layers defines the physical media and interface, the upper of the four layers defines one or more logical protocols (for example, FCP for SCSI command protocols and FC-SB-2 for FICON). Refer to ANSI X3.230.1999x.

**FICON channel.** A channel having a Fibre Channel channel-to-control-unit I/O interface that uses optical cables as a transmission medium. The FICON channel may operate in (1) FC mode (FICON native mode - FC-SB-2/3), (2) FCV mode (FICON conversion mode to a IBM 9032-5), or (3) FCP mode (FICON channel operating in “open mode”, which is FC-FCP).

**FICON.** (1) A z/Architecture computer peripheral interface. The I/O interface uses z/Architecture logical protocols over a FICON serial interface that configures attached units to a FICON communication fabric. (2) An FC4 adopted standard that defines an effective mechanism for the export of the SBCON command protocol through Fibre Channels.

**field replaceable unit (FRU).** An assembly that is replaced in its entirety when any one of its required components fails.

**FRU.** Field-replaceable unit.

**GARP.** Generic Attribute Registration Protocol.

**Gb.** Gigabit.

**GB.** Gigabyte.

**GbE.** Gigabit Ethernet.

**gigabit (Gb).** A unit of measure for storage size. One gigabit equals one billion bits.

**Gigabit (Gb).** Usually used to refer to a data rate, the number of gigabits being transferred in one second.

**Gigabit Ethernet.** An OSA channel (type OSD).

**gigabyte (GB).** (1) A unit of measure for storage size. One gigabyte equals 1,073,741,824 bytes. (2) Loosely, one billion bytes.

**GVRP.** GARP VLAN Registration Protocol.

**half duplex.** In data communication, pertaining to transmission in only one direction at a time. Contrast with *duplex*.

**hard disk drive.** (1) A storage media within a storage server used to maintain information that the storage server requires. (2) A mass storage medium for computers that is typically available as a fixed disk or a removable cartridge.

**Hardware Management Console.** A console used to monitor and control hardware such as the System z processors.

**hardware system area (HSA).** A logical area of Central Storage, not addressable by application programs, used to store Licensed Internal Code and control information.

**HCD.** Hardware configuration definition.

**HDA.** Head and disk assembly.

**HDD.** See *hard disk drive*.

**head and disk assembly.** The portion of an HDD associated with the medium and the read/write head.

**IBB.** Internal Bus Buffer

**IBF.** Internal Battery Feature.

**ICB.** Integrated Cluster Bus link.

**ICF.** Internal Coupling Facility.

**ICP.** Internal Coupling Facility peer channel.

**ICSF.** Integrated Cryptographic Service Facility.

**ID.** See *identifier*.

**IDAW.** Indirect data address word.

**Identifier.** A unique name or address that identifies things such as programs, devices, or systems.

**IFCC.** Interface control check.

**IFL.** Integrated Facility for Linux.

**IML.** Initial machine load. A procedure that prepares a device for use.

**IMS.** Information Management System.

**initial machine load (IML).** A procedure that prepares a device for use.

**initial program load (IPL).** (1) The initialization procedure that causes an operating system to commence operation. (2) The process by which a configuration image is loaded into storage at the beginning of a work day or after a system malfunction. (3) The process of loading system programs and preparing a system to run jobs.

**input/output (I/O).** (1) Pertaining to a device whose parts can perform an input process and an output process at the same time. (2) Pertaining to a functional unit or channel involved in an input process, output process, or both, concurrently or not, and to the data involved in such a process. (3) Pertaining to input, output, or both.

**input/output configuration data set (IOCDs).** The data set in the System z processor (in the support element) that contains an I/O configuration definition built by the input/output configuration program (IOCP).

**input/output configuration program (IOCP).** A System z program that defines the channels, I/O devices, paths to the I/O devices, and the addresses of the I/O devices to a system. The output is normally written to a System z IOCDs.

**Integrated Facility for Applications (IFA).** A general purpose assist processor for running specific types of applications. See *Application Assist Processor (AAP)*.

**interface.** (1) A shared boundary between two functional units, defined by functional characteristics, signal characteristics, or other characteristics as appropriate. The concept includes the specification of the connection of two devices having different functions. (2) Hardware, software, or both, that links systems, programs, or devices.

**I/O configuration.** The collection of channel paths, control units, and I/O devices that attaches to the processor. This may also include channel switches (for example, an ESCON Director).

**IOCDs.** See *Input/Output configuration data set*.

**I/O.** See *input/output*.

**IOCP.** See *Input/Output configuration control program*.

**IODF.** The data set that contains the System z I/O configuration definition file produced during the defining of the System z I/O configuration by HCD. Used as a source for IPL, IOCP, and Dynamic I/O Reconfiguration.

**IPL.** See *initial program load*.

**IRD.** Intelligent Resource Director.

**ISC-3.** Inter System Channel-3.

**ISDN.** Integrated-Services Digital Network.

**ITR.** Internal throughput rate.

**ITRR.** Internal Throughput rate ratio.

**jumper cable.** In an ESCON and FICON environment, an optical cable having two conductors that provides physical attachment between a channel and a distribution panel or an ESCON Director port or a control unit/devices, or between an ESCON Director port and a distribution panel or a control unit/device, or between a control unit/device and a distribution panel. Contrast with *trunk cable*.

**LAN.** See *local area network*.

**laser.** A device that produces optical radiation using a population inversion to provide *light amplification by stimulated emission of radiation* and (generally) an optical resonant cavity to provide positive feedback. Laser radiation can be highly coherent temporally, or spatially, or both.

**LC connector.** An optical fibre cable duplex connector that terminates both jumper cable fibres into one housing and provides physical keying for attachment to an LC duplex receptacle. For technical details, see the NCITS - American National Standard for Information Technology - Fibre Channel Standards document FC-PI.

**LCU.** See *Logical Control Unit*.

**LED.** See *light emitting diode*.

**LIC.** See *Licensed Internal Code*.

**LIC-CC.** Licensed Internal Code Configuration Control.

**Licensed Internal Code (LIC).** Software provided for use on specific IBM machines and licensed to customers under the terms of the IBM Customer Agreement. Microcode can be Licensed Internal Code and licensed as such.

**light-emitting diode (LED).** A semiconductor chip that gives off visible or infrared light when activated. Contrast *Laser*.

**link.** (1) In an ESCON or FICON environment, the physical connection and transmission medium used between an optical transmitter and an optical receiver. A link consists of two conductors, one used for sending and the other for receiving, thereby providing a duplex communication path. (2) In an ESCON or FICON I/O interface, the physical connection and transmission medium used between a channel and a control unit, a channel and an ESCON or FICON Director, a control unit and an ESCON or FICON Director, or, at times, between two ESCON Directors or two FICON Directors.

**link address.** On an ESCON or a FICON interface, the portion of a source or destination address in a frame that ESCON or FICON uses to route a frame through an ESCON or FICON director. ESCON and FICON associates the link address with a specific switch port that is on the ESCON or FICON director. **Note:** For ESCON, there is a one-byte link address. For FICON, there can be a one-byte or two-byte link address specified. One-byte link address for a FICON non-cascade topology and two-byte link address supports a FICON cascade switch topology. See also *port address*.

**local area network (LAN).** A computer network located in a user's premises within a limited geographic area.

**Local Timing Mode.** When the configuration is in Local Timing Mode, the TOD clock has been initialized to a local time and is being stepped at the rate of the local hardware oscillator. The configuration is not part of a synchronized timing network.

**logical address.** The address found in the instruction address portion of the program status word (PSW). If translation is off, the logical address is the real address. If translation is on, the logical address is the virtual address.

**logical control unit (LCU).** A separately addressable control unit function within a physical control unit. Usually a physical control unit that supports several LCUs. For ESCON, the maximum number of LCUs that can be in a control unit (and addressed from the same ESCON fiber link) is 16; they are addressed from x'0' to x'F'.

**logical partition (LPAR).** A set of functions that create a programming environment that is defined by the z/Architecture. A logical partition is conceptually similar to a virtual machine environment, except that LPAR is a function of the processor and does not depend on an operating system to create the virtual machine environment.

**logical processor.** In LPAR mode, a central processor in a logical partition.

**logical switch number (LSN).** A two-digit number used by the I/O Configuration Program (IOCP) to identify a specific ESCON Director.

**logically partitioned (LPAR) mode.** A central processor mode, available on the Configuration frame when using the PR/SM facility, that allows an operator to allocate processor hardware resources among logical partitions.

**LPAR.** See *logical partition*.

**LUPS.** Local Uninterruptible Power Supply.

**MAC.** Message Authentication Code.

**machine check.** An error condition that is caused by an equipment malfunction.

**maintenance change level (MCL).** A change to correct a single licensed internal code design defect. Higher quality than a patch, and intended for broad distribution. Considered functionally equivalent to a software PTF.

**MAU.** Multistation access unit.

**Mb.** Megabit.

**MB.** Megabyte.

**MBA.** Memory bus adapter.

**MCCU.** Multisystem channel communication unit.

**MCL.** See *maintenance change level*.

**MCM.** Multi Chip Module.

**MDA.** Motor Drive Assembly.

**megabit (Mb).** A unit of measure for storage size. One megabit equals 1,000,000 bits.

**megabyte (MB).** (1) A unit of measure for storage size. One megabyte equals 1,048,576 bytes. (2) Loosely, one million bytes.

**Message Time Ordering (MTO).** The capability on a I/O data path to ensure that information about the path is not delivered to a system that represents an event in the future relative to the TOD clock at the receiving system.

**Message Time Ordering Facility (MTOF).** A server that has MTOF is capable of exploiting MTO.

**MIDAW.** Modified Indirect Data Address Word.

**MIF.** Multiple Image Facility.

**Mixed CTN.** Timing network that contains a collection of servers, and has at least one STP-configured server stepping to timing signals provided by the Sysplex Timer. STP-configured servers in the Mixed CTN not stepping to the Sysplex Timer achieve synchronization by exchanging STP messages.

**MRU.** Modular Refrigeration Unit.

**MSA.** Motor Scroll Assembly.

**MSC chip.** Memory Storage Control chip.

**MT-RJ.** An optical fibre cable duplex connector that terminates both jumper cable fibres into one housing and provides physical keying for attachment to an MT-RJ duplex receptacle. For technical details, see the NCITS - American National Standard for Information Technology - Fibre Channel Standards document FC-PI.

**multidrop topology.** A network topology that allows multiple control units to share a common channel path, reducing the number of paths between channels and control units.

**multi-mode optical fiber.** A graded-index or step-index optical fiber that allows more than one bound mode to propagate.

**Multiple Image Facility (MIF).** In the z/Architecture, a function that allows logical partitions to share a channel path by providing each logical partition with its own set of subchannels for accessing a common device.

**National Committee for Information Technology Standards.** NCITS develops national standards and its technical experts participate on behalf of the United States in the international standards activities of ISO/IEC JTC 1, information technology.

**NCITS.** See *National Committee for Information Technology Standards*.

**ND.** See *node descriptor*.

**NED.** See *node-element descriptor*.

**node descriptor.** In an ESCON and FICON environment, a node descriptor (ND) is a 32-byte field that describes a node, channel, ESCON Director port, FICON Director port, or a control unit.

**node-element descriptor.** In an ESCON and FICON environment, a node-element descriptor (NED) is a 32-byte field that describes a node element, such as a disk device.

**NPIV.** N\_Port ID Virtualization.

**OEMI.** See *original equipment manufacturers information*.

**open system.** A system whose characteristics comply with standards made available throughout the industry and that therefore can be connected to other systems complying with the same standards.

**optical cable assembly.** An optical cable that is connector-terminated. Generally, an optical cable that has been terminated by a manufacturer and is ready for installation. See also *jumper cable* and *optical cable*.

**optical cable.** A fiber, multiple fibers, or a fiber bundle in a structure built to meet optical, mechanical, and environmental specifications. See also *jumper cable*, *optical cable assembly*, and *trunk cable*.

**optical fiber connector.** A hardware component that transfers optical power between two optical fibers or bundles and is designed to be repeatedly connected and disconnected.

**optical fiber.** Any filament made of dielectric materials that guides light, regardless of its ability to send signals. See also *fiber optics* and *optical waveguide*.

**optical waveguide.** (1) A structure capable of guiding optical power. (2) In optical communications, generally a fiber designed to transmit optical signals. See *optical fiber*.

**original equipment manufacturers information (OEM).** A reference to an IBM guideline for a computer peripheral interface. More specifically, refers to IBM S/360™ and S/370 Channel to Control Unit Original Equipment Manufacture's Information. The interfaces use z/Architecture logical protocols over an I/O interface that configures attached units in a multi-drop bus environment.

**parallel channel.** A channel having a System/360™ and System/370™ channel-to-control-unit I/O interface that uses bus and tag cables as a transmission medium. Contrast with *ESCON channel*.

**path.** In a channel or communication network, any route between any two nodes. For ESCON or FICON, this would be the route between the channel and the control unit/device, or sometimes from the operating system control block for the device and the device itself.

**path group.** The z/Architecture term for a set of channel paths that are defined to a controller as being associated with a single System z image. The channel paths are in a group state and are online to the host.

**path-group identifier.** The z/Architecture term for the identifier that uniquely identifies a given logical partition. The path-group identifier is used in communication between the system image and a device. The identifier associates the path-group with one or more channel paths, thereby defining these paths to the control unit as being associated with the same system image.

**PCHID.** Physical Channel Identifier.

**PCI.** Peripheral Component Interconnect.

**PCICC.** PCI Cryptographic Coprocessor.

**PCI-X.** Peripheral Component Interconnect eXtended.

**physical channel identifier (PCHID).** A value assigned to each physically installed and enabled channel in the server that uniquely identifies that channel. For the System z9, the assigned PCHID values are between 000 and 6FF.

**physical TOD clock.** The hardware clock at a server that is stepped by a hardware oscillator or, when a 9037 Sysplex Timer port is enabled, that is stepped by synchronization signals from the 9037 Sysplex Timer.

**PIN.** Personal Identification Number.

**PKA.** Public-Key-Algorithm.

**PKSC.** Public-Key Secure Cable.

**POR.** Power-on Reset.

**port.** (1) An access point for data entry or exit. (2) A receptacle on a device to which a cable for another device is attached. See also *duplex receptacle*.

**port address.** In an ESCON Director or a FICON Director, an address used to specify port connectivity parameters and to assign link addresses for attached channels and control units. See also *link address*.

**port card.** In an ESCON or FICON environment, a field-replaceable hardware component that provides the optomechanical attachment method for jumper cables and performs specific device-dependent logic functions.

**port name.** In an ESCON Director or a FICON Director, a user-defined symbolic name of 24 characters or less that identifies a particular port.

**Power-on Reset.** A function that re-initializes all the hardware in the system and loads the internal code that enables the machine to load and run an operating system. This function is intended as a recovery function.

**Power-on Reset state.** The condition after a machine power-on sequence and before an IPL of the control program.

**PR/SM.** Processor Resource/Systems Manager.

**Preferred Time Server.** The server assigned as the most likely choice as the current time server (active stratum-1).

**Primary Reference Time (PRT).** The primary-reference time (PRT) is maintained at each primary-time server in a CTN and is the reference time for the CTN. PRT may be provided to a server through the console or by a 9037 Sysplex Timer.

**Primary Reference Time Offset.** When PRT is provided through the console, the PRT offset represents the difference between the TOD-clock at the active-stratum-1 server and PRT at the most recent update.

**Primary Time Server.** This is a development term for the Preferred Time Server. A server that provides a primary-reference time for the CTN. A primary-time server operates at stratum-level 1.

**processor complex.** A system configuration that consists of all the machines required for operation, for example, a Processor Unit, a processor controller, a system display, a service support display, and a power and coolant distribution unit.

**program status word (PSW).** An area in storage used to indicate the sequence in which instructions are executed, and to hold and indicate the status of the computer system.

**program temporary fix (PTF).** A temporary solution or bypass of a problem diagnosed by IBM in a current unaltered release of a program.

**prohibited.** In an ESCON Director or FICON Director, the attribute that, when set, removes dynamic connectivity capability. Contrast with *allowed*.

**protocol.** (1) A set of semantic and syntactic rules that determines the behavior of functional units in achieving communication. (2) In SNA, the meanings of and the sequencing rules for requests and responses used for managing the network, transferring data, and synchronizing the states of network components. (3) A specification for the format and relative timing of information exchanged between communicating parties.

**PSC.** Power Sequence Controller.

**PSCN.** Power Service Control Network.

**PSP.** Preventive Service Planning.

**PTF.** See *program temporary fix*.

**QDIO.** Queued Direct Input/Output.

**RAS.** Reliability, Availability, Serviceability.

**remote service facility (RSF).** (1) A control program plus associated communication equipment that allows local personnel to connect to an IBM service center, and allows remote personnel to operate the remote system or send new internal code fixes to it, if properly authorized. (2) A system facility invoked by Licensed Internal Code that provides procedures for problem determination and error detection.

**RETAIN.** Remote Technical Assistance and Information Network.

**RII.** Redundant I/O Interconnect.

**RMF.** Resource Measurement Facility.

**route.** The path that an ESCON frame or FICON frame (Fibre Channel frame) takes from a channel through an ESCON Director or FICON Director to a control unit/device.

**RSA.** Rivest-Shamir-Adelman.

**saved configuration.** In an ESCON or FICON environment, a stored set of connectivity attributes whose values determine a configuration that can be used to replace all or part of the ESCON Director's or FICON Director's active configuration. Contrast with *active configuration*.

**SC chip.** Storage Controller chip.

**SC Connector.** An optical fibre cable duplex connector that terminates both jumper cable fibres into one housing and provides physical keying for attachment to an LC duplex receptacle. For technical details, see the NCITS - American National Standard for Information Technology - Fibre Channel Standards document FC-PI.

**SCP.** System control program.

**SCSW.** Subchannel status word.

**SD chip.** System Data cache chip.

**SE.** See *Support Element*.

**SEC.** System Engineering Change.

**Secondary Time Server.** A server operating at a stratum level equal to 2 or above

**Self-Timed Interconnect (STI).** An interconnect path cable that has one or more conductors that transit information serially between two interconnected units without requiring any clock signals to recover that data. The interface performs clock recovery independently on each serial data stream and uses information in the data stream to determine character boundaries and inter-conductor synchronization.

**Server Timing Mode.** The timing mode specifies the method by which the TOD clock is maintained for purposes of synchronization within a timing network. A TOD clock operates in one of the following timing modes: Local Timing Mode, ETR Timing Mode or STP Timing Mode. To be in STP-timing mode, the server must be part of an STP network.

**Small Computer System Interface (SCSI).** (1) An ANSI standard for a logical interface to a computer peripherals and for a computer peripheral interface. The interface uses a SCSI logical protocol over an I/O interface that configures attached targets and initiators in a multi-drop bus topology. (2) A standard hardware interface that enables a variety of peripheral devices to communicate with one another.

**SNMP.** Simple network management protocol.

**spanning channels.** Spanning channels have the ability to be configured to multiple Channel SubSystems, and be transparently shared by any or all of the configured logical partitions without regard to the Channel SubSystem to which the logical partition is configured.

**STI.** See *Self-Timed Interconnect*.

**STI-MP.** Self-Timed Interconnect Multiplexor.



**storage director.** In a logical entity consisting of one or more physical storage paths in the same storage cluster.

**STP.** Server Time Protocol. A time synchronization feature designed to enable multiple servers to maintain time synchronization with each other.

**STP Link.** A physical connection between two servers that is capable of supporting STP message transmission and reception. An STP link may include DWDMs and repeaters as part of the link.

**STP Path.** The combination of an STP link and STP facility resources required to perform STP message functions.

**STP Timing Mode.** When the server is in STP-timing mode, the TOD clock has been initialized to coordinated server time (CST) and is being stepped at the rate of the local hardware oscillator. In STP timing mode, the TOD clock is steered so as to maintain, or attain, synchronization with CST. To be in STP-timing mode, the server must be part of an STP network.

**stratum level.** A stratum level is associated with each server and indicates the number of servers in the timing path between the server and a primary-time server. A primary-time server is assigned a stratum level equal to one. A stratum level of zero is used to indicate that the stratum level is undefined. A stratum level of  $n$  indicates the server is  $n-1$  hops away from the current time server.

**Stratum-1 configuration.** Defined for STP-only CTNs and specifies the server that is to act as the active stratum-1 server. The configuration may also define an alternate and Arbiter server. All servers are identified by node descriptors.

**subchannel.** (1) A logical function of a Channel Subsystem associated with the management of a single device. (2) The facility that provides all of the information necessary to start, control, and complete an I/O operation. subchannel number.

**subsystem.** (1) A secondary or subordinate system, or programming support, usually capable of operating independently of or asynchronously with a controlling system.

**Support Element (SE).** (1) An internal control element of a processor that assists in many of the processor operational functions. (2) A hardware unit that provides communications, monitoring, and diagnostic functions to a central processor complex (CPC).

**SWCH.** In ESCON Manager, the mnemonic used to represent an ESCON Director.

**switch.** In ESCON Manager, synonym for ESCON Director.

**switched point-to-point channel path configuration.** In an ESCON or FICON I/O interface, a configuration that consists of a link between a channel and an ESCON Director and one or more links from the ESCD, each of which attaches to a control unit. This configuration depends on the capabilities of the ESCD for establishing and removing connections between channels and control units. Contrast with point-to-point channel path configuration.

**switched point-to-point topology.** A network topology that uses switching facilities to provide multiple communication paths between channels and control units. See also *multidrop topology*.

**Synchronization Threshold.** The amount of time that TOD clocks in a CTN are allowed to differ from CST and still be considered synchronized. The value is set so that it is less than one half of the best-case communication time between two servers over communication paths that do not support message-time ordering (MTO).

**synchronized state.** When a server is in the synchronized timing state, the TOD clock is in synchronization with the timing-network reference time as defined here:

- ▶ If the configuration is in ETR-timing mode, the server is synchronized with the ETR.
- ▶ If the server is in STP timing mode, the server is synchronized with coordinated server time (CST).

A server that is in the local-timing or uninitialized-timing mode is never in the synchronized state

**synchronized.** A timing state that indicates that the TOD clock for a server is within the synchronization threshold value relative to CST.

**Sysplex Timer.** An IBM table-top unit that synchronizes the time-of-day (TOD) clocks in as many as 16 processors or processor sides.

**sysplex.** A set of systems communicating and cooperating with each other through certain multisystem hardware components and software services to process customer workloads.

**system reset.** To reinitialize the execution of a program by repeating the load operation.

**TDES.** Triple Data Encryption Standard.

**time-of-day (TOD) clock.** A system hardware feature that is incremented once every microsecond, and provides a consistent measure of elapsed time suitable for indicating date and time. The TOD clock runs regardless of whether the processor is in a running, wait, or stopped state.

**Time server.** A computing system that has the STP facility installed and enabled.

**Time zone.** (1) Any of the 24 longitudinal divisions of Earth's surface in which a standard time is kept, the primary division being that bisected by the Greenwich meridian. Each zone is 15° of longitude in width, with local variations, and observes a clock time one hour earlier than the zone immediately to the east. (2) In some countries, standard time zones are not an integral number of hours different from UTC. STP supports time-zone offsets in integral number of minutes.

**TKE.** Trusted Key Entry.

**TOD.** See *Time of day*.

**TOD clock state.** The following states are distinguished for the TOD clock: set, not set, stopped, error, and not operational. The state determines the condition code set by execution of STORE CLOCK, STORE CLOCK EXTENDED, and STORE CLOCK FAST. The clock is incremented, and is said to be running, when it is in either the set state or the not-set state

**TOD Error State.** The clock enters the error state when a malfunction is detected that is likely to have affected the validity of the clock value. It depends on the model whether the clock can be placed in this state. A timing-facility-damage machine-check-interruption condition is generated on each CPU in the configuration whenever the clock enters the error state. When the TOD-clock-steering facility is installed, the TOD clock is never reported to be in the error state. Errors in the TOD clock cause a system check stop.

**TOD Not-Operational State.** The clock is in the not-operational state when its power is off or when it is disabled for maintenance. It depends on the model whether the clock can be placed in this state. Whenever the clock enters the not-operational state, a timing-facility-damage machine-check-interruption condition is generated on each CPU in the configuration. When the TOD-clock-steering facility is installed, the TOD clock is never reported to be in the not-operational state.

**TOD Not-Set State.** When the power for the clock is turned on, the clock is set to zero, and the clock enters the not-set state. The clock is incremented when in the not-set state. When the TOD-clock-steering facility is installed, the TOD clock is never reported to be in the not-set state, as the TOD clock is placed in the set state as part of the initial machine-loading (IML) process

**TOD Programmable Register.** Each CPU has a TOD programmable register. Bits 16-31 of the register contain the programmable field that is appended on the right to the TOD-clock value by STORE CLOCK EXTENDED. The register is loaded by SET CLOCK PROGRAMMABLE FIELD. The contents of the register are reset to a value of all zeros by initial CPU reset.

**TOD Set State.** The clock enters the set state only from the stopped state. The change of state is under control of the TOD-clock-sync-control bit (bit 34 of control register 0) of the CPU that most recently caused the clock to enter the stopped state. If the bit is zero, the clock enters the set state at the completion of execution of SET CLOCK. If the bit is one, the clock remains in the stopped state until the bit is set to zero on that CPU or until another CPU executes a SET CLOCK instruction affecting the clock. If an external time reference (ETR) is installed, a signal from the ETR may be used to set the set state from the stopped state. Incrementing of the clock begins with the first stepping pulse after the clock enters the set state.

**TOD Stopped State.** The clock enters the stopped state when SET CLOCK is executed and the execution results in the clock being set. This occurs when SET CLOCK is executed without encountering any exceptions and either any manual TOD-clock control in the configuration is set to the enable-set position or the TOD-clock-control-override control (bit 42 of control register 14) is one. The clock can be placed in the stopped state from the set, not-set, and error states. The clock is not incremented while in the stopped state.

**TOD Timing State.** The timing state indicates the synchronization state of the TOD clock with respect to the timing network reference time.

**TOD clock offset.** The TOD-clock offset is a value that, when added to the physical-TOD clock, produces the system-TOD clock. The TOD-clock-offset may be gradually steered or stepped to keep the system-TOD clock synchronized to CST.

**TOD clock steering.** Provides a means to change the apparent stepping rate of the TOD clock without changing the physical hardware oscillator that steps the physical clock. This is accomplished by means of a TOD-offset register that is added to the physical clock to produce a logical-TOD-clock value.

**TPF.** See *Transaction Processing Facility*.

**Transaction Processing Facility.** Transaction Processing Facility is a specialized high availability operating system designed to provide quick response times to very high volumes of messages from large networks of terminals and workstations.

**trunk cable.** In an ESCON environment, a cable consisting of multiple fiber pairs that do not directly attach to an active device. This cable usually exists between distribution panels and can be located within, or external to, a building. Contrast with *jumper cable*.

**TSO.** Time sharing option.

**UCW.** Unit control word.

**unblocked.** In an ESCON Director, the attribute that, when set, establishes communication capability for a specific port. Contrast with *blocked*.

**unit address.** The z/Architecture term for the address associated with a device on a given controller. On ESCON or FICON interfaces, the unit address is the same as the device address. On OEMI interfaces, the unit address specifies a controller and device pair on the interface.

**Unsynchronized State.** When a server is in the unsynchronized timing state, the TOD clock is not in synchronization with the timing network reference time as defined here:

- ▶ If the server is in ETR-timing mode, the server has lost synchronization with the ETR.
- ▶ If the server is in STP timing mode, the server has lost or has not been able to attain synchronization with coordinated server time (CST). The server is out of synchronization with CST when the TOD clock differs from CST by an amount that exceeds a model dependent STP-sync-check-threshold value.

**UPC.** Universal Power Controller.

**UPS.** Uninterruptible Power Supply.

**VLAN.** Virtual Local Area Network.

**VPD.** Vital Product Data.

**WLM.** Workload Manager.

**z/Architecture.** An IBM architecture for mainframe computers and peripherals. Processors that follow this architecture include the System z9 and zSeries servers.

**zAAP.** System z9 and zSeries Application Assist Processor. See *Application Assist Processor (AAP)*.

Archived

# Related publications

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

## IBM Redbooks

For information about ordering these publications, see “How to get IBM Redbooks” on page 296. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *FICON Implementation Guide*, SG24-6497
- ▶ *IBM System z Connectivity Handbook*, SG24-5444
- ▶ *IBM System z9 109 Technical Introduction*, SG24-6669
- ▶ *OSA-Express Implementation Guide*, SG24-5948
- ▶ *OSA-Express Integrated Console Controller Implementation Guide*, SG24-6364
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *Server Time Protocol Planning Guide*, SG24-7280

## Other publications

These publications are also relevant as further information sources:

- ▶ *Hardware Management Console Operations Guide Version 2.9.0*, SC28-6821
- ▶ *IBM System z9, IBM eServer zSeries, and S/390 Functional Matrix*, GM13-0623
- ▶ *IOCP User's Guide*, SB10-7037
- ▶ *Planning for Fiber Optic Links*, GA23-0367
- ▶ *Standalone IOCP User's Guide*, SB10-7040
- ▶ *Support Element Operations Guide (V2.9.1)*, SC28-6858
- ▶ *System z9 EC Installation Manual for Physical Planning*, GC28-6844
- ▶ *System z9 EC Processor Resource/Systems Manager Planning Guide*, SB10-7041
- ▶ *System z9 EC System Overview*, SA22-6833
- ▶ *System z9 Stand-Alone Input/Output Configuration Program User's Guide*, SB10-7152
- ▶ *System z Hardware Management Console Operations Guide Version 2.9.1*, SC28-6857
- ▶ *z/Architecture Principles of Operation*, SA22-7832
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Administrator's Guide*, SA22-7521
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Application Programmer's Guide*, SA22-7522
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Messages*, SA22-7523

- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Overview, SA22-7519*
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility System Programmer's Guide, SA22-7520*

## Online resources

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

- ▶ IBM Crypto cards Web site  
<http://www.ibm.com/security/cryptocards>
- ▶ GDPS  
<http://www.ibm.com/server/eserver/zseries/gdps.html>
- ▶ Resource Link  
<http://www.ibm.com/servers/resourceLink>

## How to get IBM Redbooks

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

[ibm.com/redbooks](http://ibm.com/redbooks)

## Help from IBM

IBM Support and downloads

[ibm.com/support](http://ibm.com/support)

IBM Global Services

[ibm.com/services](http://ibm.com/services)

# Index

## Numerics

16-port ESCON feature 105  
3270 console 259  
60 logical partitions support 172  
63.75K subchannels 9, 135, 173  
9672 server 188

## A

A Frame 39  
Advanced Encryption Standard (AES) 12, 51, 150, 161  
ASCII console 260  
Asynchronous Transfer Mode (ATM) 115  
availability mapping 269

## B

Bill of materials (BOM) 274–276  
book 37, 39, 132  
    book 0 30, 70, 99  
    book-to-book communication 36  
    configuration 36  
    jumper book 36  
    logical structure 47  
    ring structure 36–37  
    ring topology 36  
    slots 36  
Branch History Table (BHT) 51–52  
Bulk Power Assembly 41

## C

cage  
    CEC cage 10, 28–29, 39–40, 91  
    I/O cage 3, 31, 90–91, 155, 207, 215–216, 250, 277–278  
Capacity BackUp (CBU) 19, 24, 54, 57, 68, 75, 209, 220, 228–229  
    activation 229  
    contract 209  
    deactivation 230  
    enablement 231  
    example 231  
    testing 230  
capacity marker 58, 71  
capacity planning 8  
capacity setting 58  
Capacity Upgrade on Demand (CUoD) 19, 54, 57, 59–60, 69, 180–181, 205, 208–211, 213–214  
    for I/O 215  
    for memory 213  
    for processors 211  
    processor identification 180  
CCN Number 264  
Central Processor (CP) 3, 46, 51, 54, 57, 130, 149–150,

154, 160, 198, 202  
    pool 57  
Central Processor Complex 28, 79  
Central storage (CS) 69–70, 84  
CFCC 4, 57, 59, 81, 177, 188–189  
    enhanced patch apply 189  
    level 177, 189, 243  
CFLEVEL 86, 189  
CFRFile 266  
CFRM policy 185  
CFSIZER tool 178  
channel  
    spanning 138  
    sparing 107  
Channel Data Link Control (CDLC) 176  
Channel Subsystem (CSS) 9, 18, 65, 83, 130–132, 197–199, 201–202  
Checksum Offload 116–117, 119–120  
Chinese Remainder Theorem (CRT) 157  
chip lithography 42  
CHPID 17, 81, 96, 100, 102–103, 126, 131, 135, 137–138, 190, 265–268  
CHPID Mapping Tool 9, 81, 96–97, 100–101, 138, 141, 143, 263–267, 269, 271  
    reports 271  
coexistence code 142  
Commercial Batch Short (CB-S) 15  
Common Cryptographic Architecture 13, 152, 161  
Compression Unit 51  
concurrent book  
    add 20, 206  
    replacement 235, 240  
concurrent hardware upgrade 207  
    book 207, 232  
concurrent I/O upgrades 90  
configuration report 39  
Configurator for e-business 143  
configuring for availability 38  
connector 103  
Control unit (CU)  
    priority 269  
cooling 30  
Coupling Facility (CF) 3–4, 8, 17, 55, 57, 59, 70, 79, 86, 177, 184–185, 188, 193, 229, 274  
    connectivity 188  
    copy structures 189  
Coupling Facility image 193  
Coupling Link 4, 41, 47, 90, 103, 124, 190  
    peer mode 4  
coupling link 9  
CP 5, 7, 31–32, 39, 44, 57, 185, 192–193, 208  
    assigned 73  
    CP pool 55–57  
    CP4 feature 58  
    CP5 feature 58

- CP6 feature 58
- CP7 feature 58
- logical processors 66
- CPACF 90
  - cryptographic capabilities 12
  - description 51
  - design highlights 46
  - feature code 154
  - instructions 54
  - PU design 50–51
- CPC Node-Descriptor 181
- CPU
  - resources 198
- Crypto Express2 4, 10, 12–13, 41, 48, 51, 83, 90, 95, 97–98, 101–103, 127, 151, 154–158, 169–170, 243, 245, 247
  - accelerator 13, 156–157, 160
  - compatibility support 170–171
  - coprocessor 13, 90, 127, 151, 153–156, 159–160, 169–171, 246
  - exploitation support 171
  - support 162
- Cryptographic
  - synchronous function 150
- Cryptographic Accelerator (CA) 90, 154–155
- cryptographic coprocessor 247
- cryptographic domain 156–158
- CSS 135
  - configuration management 143
  - ID 83
  - Image ID 131
  - priority 198, 201–202
  - structure 131
- CTC 181
- CUoD 69, 208, 210
  - activation 226
  - for I/O 210, 215
  - for memory 210, 213
  - for processors 210–211
  - initiation 225
  - ordering 225
  - termination 227
- Customer Initiated Upgrade (CIU) 19, 54, 57, 59–60, 64, 180–181, 208, 210, 216–217, 220
  - activation 217, 220
  - enablement 216, 222
  - Ordering 219
  - Registration 217
  - registration 217

## D

- data chaining 145
- Data Encryption Standard (DES) 12, 51, 150, 159
- DCA 29, 92
- DCA-CC 92
- DCM 200
- Decimal Floating Point 53
- Dense Wave Division Multiplexing (DWDM) 190
- DES 150
- DFSMS striping 147

- Digital Signature Verify (CSFNDFV) 152
- director port cards 200
- Display ios,config 142
- disruptive upgrades 246
- Distributed Converter Assembly (DCA) 29, 92
- dual processor design 50
- Dynamic CF dispatching 59–60, 193
- Dynamic Channel Path Management 197, 199–200, 203
- Dynamic ICF expansion 59, 193–194
- dynamic oscillator switchover 21, 206
- Dynamic storage reconfiguration (DSR) 70, 87, 244
- dynamic workload balancing 185
- DYNDISP 60

## E

- Electronic Industry Association (EIA) 39
- ELIGIBLE DEVICE Table (EDT) 142
- Enhanced Book Availability
  - prepare 235
- Enhanced Book Availability (EBA) 7, 20, 35, 38–39, 68, 206, 232, 235
- Enhanced driver maintenance (EDM) 7, 20, 206, 242–243
- Enhanced ETR Attachment Facility (EEAF) 48
- Enterprise Extender (EE) 118
- Enterprise Fiber Cabling Services 274
- Error Correction Code (ECC) 34, 68
- ESA/390 Architecture mode 85
- ESA/390 TPF mode 86
- ESCON 200
  - 16-port ESCON (2323) 105
  - director 255
- ESCON channel 11, 41, 47, 81, 90, 92, 103, 105, 139, 176, 207, 215, 276–278
  - feature codes 97
- Ethernet switch 116, 118
- ETR
  - Network ID 187
- Europay Mastercard VISA (EMV) 2000 151
- EXCP 147
- EXCPVR 147
- Expanded storage 69, 84
- Extended Addressability 147
- Extended Format Data Set 147
- Extended Translation Facility 53–54
- External Time Reference (ETR) 90
  - cards 29, 103
  - Enhanced ETR Attachment Facility 48
  - ETR ID 186
  - feature 98, 126–127
  - Network ID 186
  - ports 29
  - receiver 42

## F

- FCP 10, 113
  - concurrent patch 114
- feature code 97, 206, 216
  - CBU 228



- FC 1995 216
- FC 2824 206, 213
- FC 28xx 241
- FC 9904 114
- FICON FCP 114
- flexible memory option 232
- I/O and cryptographic feature 98
- ISC-3 102
- PCHID report 99
- purchased PU 73
- STI Rebalance 211
- zAAP 60
- zIIP 64
- Federal Information Processing Standard (FIPS) 13
- fiber
  - Single mode (SM) 104, 190
- Fiber Cabling Services 274
- Fiber Distributed Data Interface (FDDI) 115
- fiber optic
  - cable 275
  - harness 276
  - trunking 275–276
- fiber optic cable
  - 50.0 micron 116–118
  - 62.5 micron 111, 116–118
- Fiber Quick Connect (FQC) 104, 107, 276–278
- Fiber Transport System (FTS) 104, 107, 276–278
- fiber-optic
  - cable 274
- Fibre Channel
  - switch 10, 108–111
- Fibre Channel Protocol 10, 46, 90, 108–111, 113
- FICON 41, 243
  - cascaded directors 9, 11
- FICON channel 7–10, 41, 103, 108, 110–111, 113, 174
- FICON Express 4, 10–11, 47, 95, 97–98, 101
  - 4 LX (10Km) 97
  - channel 41, 110, 113, 139
  - LX 103–104, 108, 110–111
  - SX 104, 109–111
- FICON Express2 4, 7, 10–11, 41, 47, 95, 97, 101, 167, 169–171
  - LX 108, 110
- FICON Express4 47, 95
  - performance improvement 112
- FICON link incident reporting 175
- Five-Model structure 4
- FlashCopy 133
- flexible memory option 20, 35, 39, 68, 206, 213, 232, 235
- frame 39

## G

- GARP VLAN Registration Protocol (GVRP) 122, 176
- GDPS 183, 231
  - Hyperswap 133
- Glass ceramic substrate 7, 42
- Global Mirror 133
- Graphical User Interface (GUI) 257

## H

- hardware compression 51
- Hardware Configuration Dialog 132
- Hardware Configuration Dialog (HCD) 136
- Hardware Management Console (HMC) 19, 65, 70, 79, 116, 118, 198, 200–201, 203, 216, 220, 224, 230, 257–260
  - APIs 259
  - Application 255
  - Integrated 3270 Console 259
  - Integrated ASCII Console support 260
  - local 257–258
  - remote operation 257–258
  - User interface 258
  - Web browser control 257
- Hardware System Area (HSA) 9, 39, 69–70, 83, 132, 142
- HCD 70, 81, 83, 132, 134, 136, 143, 246
- high bandwidth 90
- HiperSockets 124
  - IPv6 12, 177
- hybrid cooling system 31

## I

- I/O
  - connectivity 10, 46, 48
  - path 200
  - performance 201
- I/O cage
  - ESCON channels 277
  - I/O slot 10, 91–92, 95, 97–99, 103, 215, 245–246
  - planar board 94
- I/O card 90, 92, 94, 207, 210, 215–216, 245
- I/O Configuration Program (IOCP) 81, 134, 136–139
- I/O device 38, 48, 90, 95, 101, 105, 108, 130, 274–276
- I/O domain 39, 91–92, 94, 96–97
- I/O feature
  - card 97–98
  - support 90
- I/O operation 65, 130, 144–145, 202
- I/O request 201–202
- IBM Fiber Transport System 277
- IC 190
- IC3 17
- ICB-3 link 16, 41, 94–96, 126, 190
- ICB-4 link 17, 28, 37, 48, 94, 97–98, 100, 125, 190, 244
  - cable 125
  - PCHID 100, 125
  - STI rebalance 211
- ICF 31–32, 44, 55–56, 59, 71, 73, 222
  - backup capacity 59
- ICKDSF 178
- ICSF 152, 156, 159, 162
- IEEE Floating Point 52
- IFA 61
- IFC 59
- IFL 31, 44, 55–56, 58, 71, 207, 209–211
  - assigned 73
  - backup capacity 59
- Indirect Address Word (IDAW) 144

Input/Output Configuration Dataset (IOCDs) 9, 17, 83, 136, 143  
 Installation Planning Representative (IPR) 274  
 Instruction fetching 53  
 Instruction grouping 53  
 Instruction set extensions 54  
 Integrated Cluster Bus (ICB) 90, 94, 101, 124  
 Integrated Facility 4, 12, 55, 61, 71, 79  
     for Applications 61  
     for Linux 3–4, 8, 12, 32, 54–56, 58–59, 66, 71, 208, 220, 222, 228  
 Internal Battery Feature (IBF) 40, 45, 251  
     estimated power time 40  
 Internal Coupling (IC) 55, 59, 71, 73, 102, 139, 190, 193  
 Internal Queued Direct I/O (IQDIO) 102  
 Internet connection 259  
 Intersects 270  
 IOCDs 83, 136, 139, 143  
 IOCP 81  
     definition 125–126, 140–141  
     source 265–266, 269–270  
 IODF 142–143, 203  
     V5 142  
 IRD 18, 46, 80–81  
     Channel Subsystem priority queuing 18  
     Dynamic Channel Path Management 18  
     LPAR CPU Management 18, 80  
     overview 197  
 ISC-3 16, 98, 190  
     Daughter Card 124  
     link 41, 47, 91, 102, 104, 124, 190, 207, 215  
     Mother Card 124  
 ISO 16609 CBC Mode 152  
 ITRR 15

## J

Java Runtime Environment (JRE) 264  
 Java Virtual Machine (JVM) 49, 60–62

## K

Key Exchange 152

## L

L1 cache 46, 54  
 L2 cache 37, 43, 46, 54  
 Land Grid Arrays 42  
 Large System Performance Reference 14  
 LC Duplex  
     connector 108–111, 115, 117  
 LC Duplex connector 108–111, 116–117  
 LDAP 261  
 Level 1 cache 43  
 Level 2 (L2) cache 37  
 LIC level 20, 242  
 LIC-based upgrades 207  
 LIC-CC 8, 207  
     memory 207  
     processors 207

LICCC 287  
 Licensed Internal Code (LIC) 3, 8, 33, 54, 57, 190, 205–207, 210  
     processors 207  
 Licensed Internal Code Configuration Control (LIC-CC) 8  
 Lightweight Directory Access Protocol 261  
 Linux 1, 3–4, 7, 57–58, 82, 113–114, 119–120, 151, 161, 166, 171, 173, 192  
     mode 86  
     storage 86  
 Linux for System z 58, 165–166, 171  
 Loading of Initial ATM Keys 151  
 Local area network  
     Open Systems Adapter family 11  
 Local area network (LAN) 11, 14, 103, 119, 159  
 logical partition 8, 28, 79, 103, 135, 156, 167–168, 170–172, 213–214, 260  
     alias devices 70  
     central storage 79  
     CFCC 82  
     channel resource utilization 18  
     CPU weights 18  
     Dynamic Add/Delete 83, 136  
     I/O operations 65  
     I/O priorities 201  
     identifier 135  
     logical processors 18, 80  
     MIF ID 136  
     name 136  
     number 135  
     OSA ports 8  
     PCI Cryptographic Candidate List 246  
     processing weights 198  
     processor upgrade 66  
     real storage 79  
     relative priority 18  
     required number 246  
     Reserved Processors 247  
     Reserved Storage 247  
     weights 199  
 logical processor 65–66, 79  
 LPAR 201–202  
     cluster 197–198  
     CPU management 197  
     mode 57, 83  
     single storage pool 70  
 LPAR mode 16, 57, 66, 70, 79, 82, 197, 200, 246  
     Dynamic Channel Path Management 200  
 LSPR 14

## M

machine type 4  
 managed channels 200  
 manual mapping 269  
 master key entry 156  
 MBA 29, 37, 47, 93, 206  
     fan out card 8, 21, 30, 37–38, 41, 93, 99  
     fan out plugging sequence 93  
 MCM 2, 4, 7, 29, 42–43  
     chip layout 43

- MCP 116–118
- Media Manager 148
- memory 7
  - allocation 69
  - card 8, 29, 32–34, 46, 207, 210, 213–214
  - size 28, 32, 44, 68
  - upgrade 32
- Memory Bus Adapter (MBA) 8, 37, 92, 94, 206
- Memory Coherent Controller 46
- Message authentication code (MAC) 150, 156
- Message Time Ordering 184–185
- Metro mirror 133
- MIDAW facility 7, 11, 24, 112, 129, 144–146, 167, 174
- MIF 136
  - ID 136
- MIF ID 83, 136
- Miscellaneous Equipment Specification (MES) 211
- Mixed-CTN 187
- Mod\_Raised\_to Power (MRP) 151
- Mode conditioning patch (MCP) 103, 105, 111, 116–118, 127, 274–275
- model capacity identifier 58, 74, 180–181, 209, 211
  - 701 to 754 73
  - ICF 59
  - IFL 58
  - model upgrade 208
  - subcapacity settings 73, 75
  - zAAP 62
- Model S08 31–33, 44, 90, 102, 106, 212, 214
- Model S54 2, 32, 42–43, 132
- model upgrade 5, 208
- modes of operation 81
- Modular Refrigeration Unit 30
- Modulus Exponent (ME) 157
- Motor Drive Assembly (MDA) 30–31
- Motor Scroll Assembly 30
- MSC chip 44
- MSU value 15, 60, 63, 71, 74
- Multiple CSSs 135, 138
- Multiple Image Facility (MIF) 103, 133, 135, 199, 244, 246
- Multiple Subchannel Sets 3, 7, 9, 129, 133, 141, 167, 171, 174, 245
- Multi-Terminated Push-On (MTP) 107

## N

- N+1 power supply 29
- N\_Port ID Virtualization (NPIV) 11, 24, 114, 169–171, 174
- Native FICON 10–11, 24, 174
- NED 181, 211
- Network Control Program 12
- network protocols 260
- nondisruptive CUoD 230
- nondisruptive upgrades 245

## O

- On/Off CoD 19, 24, 54–55, 57, 59–60, 64, 66, 75, 180–181, 208–211, 217, 222

- active CP 222
- active ICF 222
- active IFL 222
- active zAAP 222
- contractual terms 224
- enablement feature 222
- granular capacity 78
- Repair capability 227
- rules 78
- Upgrade Capability 227
- Open Fiber Control (OFC) 125
- Open Systems Adapter (OSA) 4, 7, 119, 121, 277, 279
- operating system 4, 58, 66, 103, 113–114, 120, 165–167, 185, 198, 202, 207–208, 260
  - channel connectivity 121
  - logical partitions 212, 217, 224
  - requirements 165
  - Support 167
  - Support Web page 182
- order and fulfillment process 217
- OSA Address Tables (OAT) 243, 266–267
- OSA Express-2
  - 1000BASE-T Ethernet 175
- OSA for NCP 12
- OSA-2
  - FENET 118
- OSA-Express 4, 11, 41, 47
  - 1000BASE-T 97, 118
    - Ethernet 98, 115, 118
    - Ethernet feature 118–119
  - 1000BASE-T Ethernet 41, 47
  - Fast Ethernet 118
  - GbE 97
  - GbE LX 118
  - GbE SX 117
  - Integrated Console Controller 119
- OSA-Express2 41, 47
  - 10 Gb Ethernet LR 41, 47, 98
  - 1000BASE-T
    - Ethernet 117
    - Ethernet feature 116
    - feature 119
    - 1000BASE-T Ethernet 7, 24, 41, 47, 115–116, 119, 121, 170–172
    - Gigabit Ethernet 12, 97, 115, 167, 170–172
    - Gigabit Ethernet 10 GbE LR 115
    - OSN 7, 12, 24, 121, 167, 176
- OSA-ICC 119
- oscillator 48, 206
- oscillator card 29

## P

- Parallel Access Volume 7, 9, 133
- parallel channel 12
- Parallel Sysplex 2, 16–17, 124, 179, 183–185, 277, 279
  - Application consideration 184
  - architecture 185
  - cluster 18, 184–185, 244
  - configuration 184–185, 188, 190
  - environment 16, 18, 46, 124, 184

- License Charge 179
- system images 185
- technology 18, 184
- third CF 196
- Web site 190
- XCF link configurations 124
- PCHID 9, 81, 96–100, 125, 137–140, 155, 161, 211, 244, 247, 265–267
  - assignment 9, 138, 265–267
- PCI Cryptographic Accelerator (PCI CA) 46, 157
- PCICA 41
- PCI-X cryptographic adapter 41, 98–99, 127, 150–151, 153–158, 245
- PCI-X cryptographic coprocessor 46, 127, 154
- peer mode 90, 103, 124–126, 188, 190–191
- Performance 14
- performance 201
- performance improvement
  - CP 6, 12
  - Crypto Express2 157
  - FICON/FCP 112
  - native FICON 112
  - QDIO 120
- Peripheral Component Interconnect 90, 101, 103
- permanent capacity 208
- Personal Identification Number (PIN) 14, 156
- physical memory 29, 32–34, 68, 232, 241
- PKA Encrypt 157
- PKA Key Import (CSNDPKI) 152
- PKA Key Token Change (CSNDKTC) 152
- Plan 216
- Plan Ahead 211, 216, 247
  - concurrent conditioning 216
- planned upgrades 208
- Point-to-point attachment 114, 169, 171
- pool
  - ICF pool 59
  - IFL pool 58
- Power On Reset (POR) 181, 211, 229, 244–245
  - expanded storage 70
  - Hardware System Area 132
- power supply 29
- PR/SM 69, 79, 198
- Preventive Service Planning (PSP) 165
- Processor Unit (PU) 2–7, 28–29, 31–32, 37, 42–44, 54–55, 57, 66, 69, 71, 90, 185, 192–194, 208, 216, 218, 228
  - characterizable PU 241
  - characterization 83
  - chip 43
  - concurrent conversion 5, 207–208
  - configuration 3
  - conversion 73, 208
  - cross-point switch 46
  - cycle time 43
  - dual-core 42, 44
  - feature code 4
  - Maximum number 4
  - pool 55, 173
  - separate management 8

- single-core 42
- spare 55, 67, 229
- sparing 55
- type 81, 83, 208, 241
- processor weighting 198
- program directed re-IPL 175
- Pseudo Random Number Generation (PRNG) 4, 12, 51, 150–151, 161–162
- PSP buckets 166
- Public Key
  - algorithm 151, 156, 159
  - Decrypt 151, 161
  - Encrypt 151, 161
- Public Key Algorithm (PKA) Facility 151

## Q

- QDIO Enhanced Buffer-State Management 114
- Queued Direct Input/Output (QDIO) 9, 113, 116–117, 119, 175

## R

- Reconfigurable Storage Unit (RSU) 87
- Red Hat RHEL 21, 166, 171–173, 176
- Redbooks Web site 296
  - Contact us xi
- Redundant I/O interconnect (RII) 7–8, 20, 24, 94, 96, 206, 235
- refrigeration 30
- Reliability, Availability, Serviceability (RAS) 20, 96
- Remote support facility (RSF) 216, 219–220, 224, 259
- Request Node Identification Data (RNID) 113, 167, 174
- reserved
  - logical partition 246
  - processor 65, 217, 247
  - PUs 228–229, 231
  - storage 86
- Resource Access Control Facility (RACF) 156
- Resource Link 70, 209, 216–217, 219, 264, 266, 271
- Resource sharing 185
- RESOURCE statement 141
- Rivest-Shamir-Adelman (RSA) 151, 156–157, 161
- RPQ 8P2197 98, 102, 125, 190
- RSF 259

## S

- SALC 179
- SAP
  - additional 71, 244
  - concurrent book replacement 241
  - definition 65
  - number of 31, 44, 71, 220
  - standard 3
- SC chip 42–43, 46
- SCHSET operand 141
- SCSI disk 169, 175
- SD chip 37, 43, 46
- Secure Hash Algorithm (SHA-1) 12
- Secure Sockets Layer (SSL) 13, 46, 51, 150–151, 154,

- 157, 160, 260
- Select Application License Charges 179
- Self-Timed Interconnect (STI) 8, 37, 41, 46, 48, 93–94, 96, 100, 200, 211, 244
  - granularity 8
- Self-Timed Interconnect (STI) 6, 8, 17, 92–94, 206, 235, 241
- Server Time Protocol (STP) 48
- Service Representative 107, 267, 271
- SET CPUID 180
- SHA-1 150
- SHA-256 12, 150
- simplified I/O definition 200
- single storage pool 70
- single system image 173, 185
- Small Computer System Interface (SCSI) 46, 90, 169–171, 175
- SNMP 259
- soft capping 179
- software licensing 178
- software support 21, 62, 70, 173, 177
- spanned channels 9, 246
- Standard SAP 44
- STI 4, 6, 30, 46, 92–93, 96, 106, 132, 200
  - connector 37, 94
  - extender card 41, 48
  - rebalance feature 96, 100, 211, 244
- STI link 94–95, 97, 269
  - balancing 96
  - distribution 96
- STI rebalance feature 39
- STI-3 card 41, 48, 92, 95, 126
- STI-A4 card 94
- STI-A8 card 94
- STI-MP card 38, 47, 92, 94–95
- storage
  - CF mode 86
  - ESA/390 mode 85
  - expanded 69
  - Linux only mode 86
  - operations 84
  - reserved 86
  - TPF mode 86
  - z/Architecture mode 85
- Storage Area Network (SAN) 103
- Storage Control (SC) 42, 44, 48
- Store System Information (STSI) 73, 180–181, 227
- structure size 177
- subcapacity models 73, 75, 222
- subchannel 133, 173
- Superscalar 28, 49
- Support Element (SE) 6, 41, 127, 135, 180, 186, 221, 234, 266
- support requirement
  - Linux on System z 171
  - z/OS 167
  - z/VM 168, 170
- SUSE SLES 21, 166, 171–173, 176
- Symmetric Multi-Processor (SMP) 32, 42
- Symmetric Multi-Processor (SMP) 32

- Sysplex Timer 126, 184–187, 255
- System Assist Processor (SAP) 3–4, 31, 65
- system image 69, 79, 84, 103, 126, 173, 175, 184–185, 194, 197, 244
- systems management 185

## T

- Time-of-Day (TOD) 126
- Timing mode 187
- TKE workstation 14, 153, 159
- Token Ring 12
- TPF 12, 172
- Translation Look aside Buffer (TLB) 50, 53
- Transmission Control Protocol (TCP) 119
- Transport Layer Security 150
- Triple Data Encryption Standard (TDES) 12, 51, 150
- Trusted Key Entry (TKE) 14, 153, 156

## U

- UDX 152
- unassigned
  - CP 71, 73
  - IFL 71, 73
- uniprocessor speed 198
- unplanned upgrades 209
- Unshielded Twisted Pair (UTP) 104, 116, 118
- upgrades
  - disruptive 246
  - non disruptive 245
- usage domain index 247
- User Datagram Protocol (UDP) 119
- User Defined Extensions (UDX) 13, 152, 157, 160
- user ID 218, 264

## V

- version code 180–181
- VLAN ID 176

## W

- WebSphere MQ 179
- wide connectivity 90
- wild branch 52
- WLM
  - Goal mode 18, 197–198, 200–201
- Workload License Charge (WLC) 80, 178–179, 217, 230
  - CIU 213
  - Flat WLC (FWLC) 178
  - sub-capacity 179
  - Variable WLC (VWLC) 178
- Workload Manager (WLM) 18, 199, 202

## Z

- Z Frame 39, 41
- z/Architecture 1, 4, 16, 53, 81, 83, 85–86, 154, 166, 200–201
- z/OS 79–81, 142–143, 162–163, 168, 172, 246
- z/TPF 21, 172, 177

z/VM 58, 82, 168–170, 175, 215  
z/VSE 170, 176–177  
z9 BC 48, 159, 185  
z990 4, 48, 56, 72, 98, 250  
zAAP 19, 31–32, 44, 54–56, 60, 71, 222, 228–229, 244  
    zAAP pool 55, 60  
zIIP 19, 31–32, 44, 54–56, 63, 71  
    zIIP pool 55–56, 64



**IBM System z9 Enterprise Class Technical Guide**

Archived









# IBM System z9 Enterprise Class Technical Guide



## Structure and design a total systems approach

## Processor unit, memory, channel subsystems, and subchannel sets

## Concurrent upgrade and maintenance options

This IBM Redbooks publication discusses the IBM System z9 Enterprise Class (z9 EC), which offers a continuation of the IBM scalable mainframe servers. Based on z/Architecture, the System z9 Enterprise Class server provides major extensions by:

- ▶ Increasing the maximum number of Processor Units and logical partitions.
- ▶ Supporting larger configurations with the concept of Channel Subsystems and Multiple Subchannels Sets.
- ▶ Providing a base for major server consolidation by further removing memory, processor, and channel constraints.

In addition to increased performance and expansion options, improved facilities for nondisruptive maintenance and growth provide better operational support and availability.

This book provides an overview of the z9 EC and its functions, features, and associated software support. More details are offered in selected areas relevant to technical planning.

This book is intended for systems engineers, consultants, planners, and anyone wanting to understand the new IBM System z9 Enterprise Class functions and plan for their usage. It is not intended as an introduction to mainframes. Readers are expected to be generally familiar with existing System z technology and terminology.

This publication is part of a series. For a more complete understanding of System z9 capabilities, also refer to our companion Redbooks:

- ▶ *IBM System z9 Business Class Technical Introduction*, SG24-7241
- ▶ *IBM System z Connectivity Handbook*, SG24-5444

## INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

## BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:  
[ibm.com/redbooks](http://ibm.com/redbooks)

SG24-7124-02

ISBN 0738490301