IBM System z9 Business Class Technical Introduction

- Describes new server hardware and related features
- Discusses key functional elements and enhancements
- Reviews associated software support and considerations

Bill Ogden
Marian Gasparovic
Bill White

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM System z9 Business Class Technical Introduction

May 2007

SG24-7241-01

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

Second Edition (May 2007)

This edition applies to the initial announcement of the IBM System z9 Business Class server.

Contents

Notices

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

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

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

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

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

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

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

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

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

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

Trademarks

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

Redbooks (logo) ®
eServer™
z/Architecture®
z/OS®
z/VM®
z/VSE™
zSeries®
z9™
CICS®
DB2®
DFSMS™
DRDA®
ESCON®
FICON®
HiperSockets™
IBM®
OS/2®
OS/390®
Parallel Sysplex®
Power PC®
PR/SM™
Redbooks®
Resource Link™
S/360™
S/390®
Sysplex Timer®
System z™
System z9™
VSE/ESA™
VTAM®
WebSphere®

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, JVM, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. 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 introduces the IBM System z9™ Business Class (z9 BC) server, which provides the advantages of the IBM System z9 Enterprise Class (z9 EC) servers in a less expensive base. The z9 BC server, which is based on z/Architecture®, provides extensions to this architecture. At the time of writing, the z9 BC server is the latest product in this line. The z9 BC server provides the first implementation of zIIP special processors. It also provides an unprecedented range of capacity versions.

This document provides basic information about z9 BC hardware functions and features, associated software support, and migration considerations. It is intended for systems engineers, hardware planners, and anyone else needing to understand the new elements.

This document is not intended as an introduction to mainframes. Readers are expected to be generally familiar with current IBM System z™ technology and terminology. It is especially directed to readers most familiar with z890, z800, and earlier systems. Functions that are unique in the existing IBM z9 servers are briefly mentioned because they may not be familiar to this set of readers.

The team that wrote this book

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

Bill Ogden is a retired IBM Senior Technical Staff Member at the International Technical Support Organization, Poughkeepsie, who has worked since 1966 with the S/360™ and all of its descendants. He specializes in smaller entry-level and educational systems.

Marian Gasparovic is an IT Specialist from IBM Slovakia. He worked as OS/390® and z/OS® administrator at an IBM Business Partner for six years. He joined IBM in 2004 and now works as Field Technical Sales Support for z9 in the CEMA region as a member of a new workload team.

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

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

Jason Boxer, IBM Systems & Technology Group, Operations TSM Lead, IBM Poughkeepsie, who helped us with many of our questions.

Parwez Hamid, IBM UK, who helped us with the more obscure details involved in understanding this server.

Charles Shapley, IBM Poughkeepsie, who provided *very* thorough proofreading.

Become a published author

Join us for a two- to six-week residency program! Help write a 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, or customers.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll 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`

Comments welcome

Your comments are important to us!

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

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

  `ibm.com/redbooks`

- Send your comments in an e-mail to:

  `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

1

Introduction

The IBM System z9 Business Class server (also known as the *z9 BC server*) is part of the next step in the evolution of the mainframe family. Based on the IBM System z9 EC, it introduces new functions and extensions. Terminology is always important and can be confusing when several products have similar names. We use the following terminology in this book:

Full name	Short name	Product examples
IBM System z9	System z9	z9 BC, z9 EC
IBM System z9 Business Class server	z9 BC	
IBM System z9 Enterprise Class server	z9 EC	
IBM eServer zSeries	zSeries	z800, z890, z900, z990
System z9 and zSeries	System z	z9 EC, z9 BC, z990, z890, ...

The key point is that the term *System z* includes both zSeries® and System z9 products, but *System z9* includes only the z9 BC and z9 EC products. Consistent use of this terminology should reduce confusion about the product set being discussed.

This naming convention produces two convenient names for the current product line: z9 EC (for the larger server) and z9 BC (for the new, smaller server described in this book).

System z9 builds on the z/Architecture and the system structure introduced with the z990. The System z9 products have been designed using a holistic approach that includes the latest operating system, middleware, storage, and networking technologies. The synergies gained through this collaborative design and development enable the elements to support each other while helping to deliver additional value to your business.

Today, many businesses are starting their transformation to become on demand businesses, enabling them to sense and respond to change on the fly and with flexibility, to manage risk while lowering costs, and to gain greater control and transparency with less complicated management. Businesses are realizing their goals by addressing complexity in the infrastructure and by creating a flexible and resilient on demand operating environment.

The extensive selection of capacity granularity of the z9 BC system can bring these on demand benefits to help align business processes and priorities. The addition of a new *specialty processor* enables key workloads to be deployed at lower cost of ownership.

1.1 Evolution

The z9 BC server continues the evolution of IBM mainframe servers, building on the z9 EC. This evolution observes the key factors of mainframe progress:

- *Compatible evolution*. IBM recognizes the very large customer investment in existing data processing applications and associated business processes. As has been the case since the S/360 was introduced, new functions and features have been designed to accommodate existing applications whenever possible.
- *Resolving constraints*. Evolving hardware capabilities and increasing application complexity and size continue to push the limits of existing systems. Every new generation of mainframe servers reduces various constraints that may have inhibited application growth on earlier systems.

Providing such compatibility can require considerable design ingenuity to provide system growth while retaining compatibility with earlier systems. Examples of evolution include:

- Single-processor systems evolved to multiple-processor systems. The z9 EC server continued this evolution by providing a modular multi-book design that supports one to four books and up to 54 processor units (customer-usable PUs) per server. The z9 BC extends the System z9 technology to smaller, less expensive systems with up to seven customer-usable PUs in a single book.
- Real memory addressing evolved to virtual memory systems. The z9 EC and z9 BC servers expand this function by providing a more sophisticated Translation Lookaside Buffer (TLB) design that is, in effect, 16-way associative.
- Memory addressing (real and virtual) evolved from 24 bits to 31 bits to 64 bits. The z9 EC and z9 BC servers allow for full 64-bit real and virtual storage support, and any logical partition can be defined for 31-bit or 64-bit addressability.
- Real memory size has grown with every IBM mainframe generation. The z9 EC server continued this growth by providing up to 512 GB of system memory. The z9 BC uses the same memory design as the z9 EC, with up to 64 GB in its single book.
- A single system image has evolved to logical partitions providing multiple system images. The z9 EC server continued this evolution by providing up to 60 logical partitions. The smaller z9 BC server provides up to 15 or 30 logical partitions, depending on the model.
- Processor speeds have grown with every generation. There is no simple metric to characterize processor performance in today's systems because many factors interact in very complex ways. The z9 EC server clock speed increased to 1.65 GHz, providing a significant performance improvement over earlier zSeries servers. The z9 BC, to better manage costs for a smaller system, uses a clock speed of approximately 1.42 GHz.
- I/O channels evolved from a maximum of 16 channels in S/360 servers to a much larger number of channels. The z9 EC server continued such growth by increasing the number of STIs (up to 16 per book) and allowing more high-speed channel features to be used in the I/O cages. Each STI has a bandwidth up to 2.7 GBps. The z9 BC uses the same design as the z9 EC, allowing up to 16 STIs in its single book.
- The number of I/O devices in a system was initially limited by the number of channels, the number of control units on each channel, and the number of devices on each channel. The addressing structure also provided a limitation. The fixed 3-byte addresses (one byte each for channel, control unit, and device) of early systems evolved into 4-byte device numbers allowing up to almost 64 K device addresses. The z9 EC and z9 BC servers continue this growth by providing Multiple Subchannel Sets (MSS), allowing up to almost 128 K device addresses in a partition. The use of multiple channel subsystems (CSSs) can further increase the number of devices that can be addressed.

- Channel performance has grown from parallel channels to ESCON® channels to FICON® channels. The z9 EC and z9 BC server continue such growth by providing a significantly higher-performance option (MIDAWs) for channel programming and faster FICON channels.
- Specific types of workloads have been partly offloaded into segregated processors such as ICFs, IFLs, and zAAPs. The z9 EC and z9 BC servers provide a significant extension with zIIP processors designed to offload certain types of processing, with initial usage by database products. They also enhance system management by providing separate pools for PR/SM™ handling of shared CPs, ICFs, IFLs, zAAPs, and zIIPs[1].
- Basic *real* systems evolved into virtual systems, and this virtual evolution has extended to systems, processors, memory, I/O devices, LAN interfaces, and so forth. The z9 EC and z9 BC servers continue this direction with new instructions that improve the performance of virtual server QDIO operations. This is done by creating a pass-through architecture designed to reduce host programming overhead, avoiding the stopping of guest processing when adapter interruptions are present or when starting QDIO operations.
- Recent mainframe generations have extended the instruction set provided to include instructions more compatible with other platforms (such as binary floating point), instructions to better implement popular languages (such as the string-handling instructions for C/C++), instructions to improve register usage (such as the relative and immediate instructions, and the long-displacement instructions), and so forth. The z9 EC and z9 BC servers continue this expansion with new and changed instructions.
- Cryptographic hardware assistance has been available in many forms on earlier systems, and with much more emphasis in more recent servers. The z9 EC and z9 BC servers continue the evolution of cryptographic hardware processing by extending the functions of the basic cryptographic instructions and consolidating the options (secure coprocessor and accelerator) in a single feature. The two options can be individually defined to the feature. New functions provide additional remote interfaces for managing and using cryptographic functions.
- Transparent hardware recovery has been a keystone in mainframe design and has evolved in many directions. The z9 EC and z9 BC servers continue this evolution by extending such transparent recovery functions to include the paths from I/O cages to system memory.
- Concurrent maintenance is a major design goal for modern mainframes and often involves balancing a design between replicated components and more integration onto chips and MCMs. The z9 EC server allows for a single book, in a multi-book configuration, to be concurrently removed and reinstalled during an upgrade or repair. The z9 BC, being a single-book system, allows concurrent maintenance for almost everything except the book.
- Clustering of mainframes evolved from simple shared DASD to basic sysplex functions and then to Parallel Sysplex® operation. The z9 BC and z9 EC servers continue this evolution by better integrating the rigorous timer coordination[2] needed in a Parallel Sysplex.
- Virtualization has evolved greatly since the IBM S/360 model 67 was introduced in 1966, with IBM mainframes providing vastly more sophisticated virtualization than other platforms. The latest z9 extension provides more automatic virtualization of LAN adapters, operating in both layer 2 and layer 3 modes.

[1] See 2.4.1, “PU characterizations” on page 14.
[2] This function is called Server Time Protocol (STP).

1.2 z9 BC server highlights

As mentioned, the z9 BC is intended to provide an upgrade path for users of z890, z800, and older systems. The z9 BC server is offered in two hardware models and has a machine type of 2096. It has one frame the same size as the z890 frame (Figure 1-1), is air-cooled, and uses three-phase or one-phase power. An Internal Battery Feature (IBF) is also available.

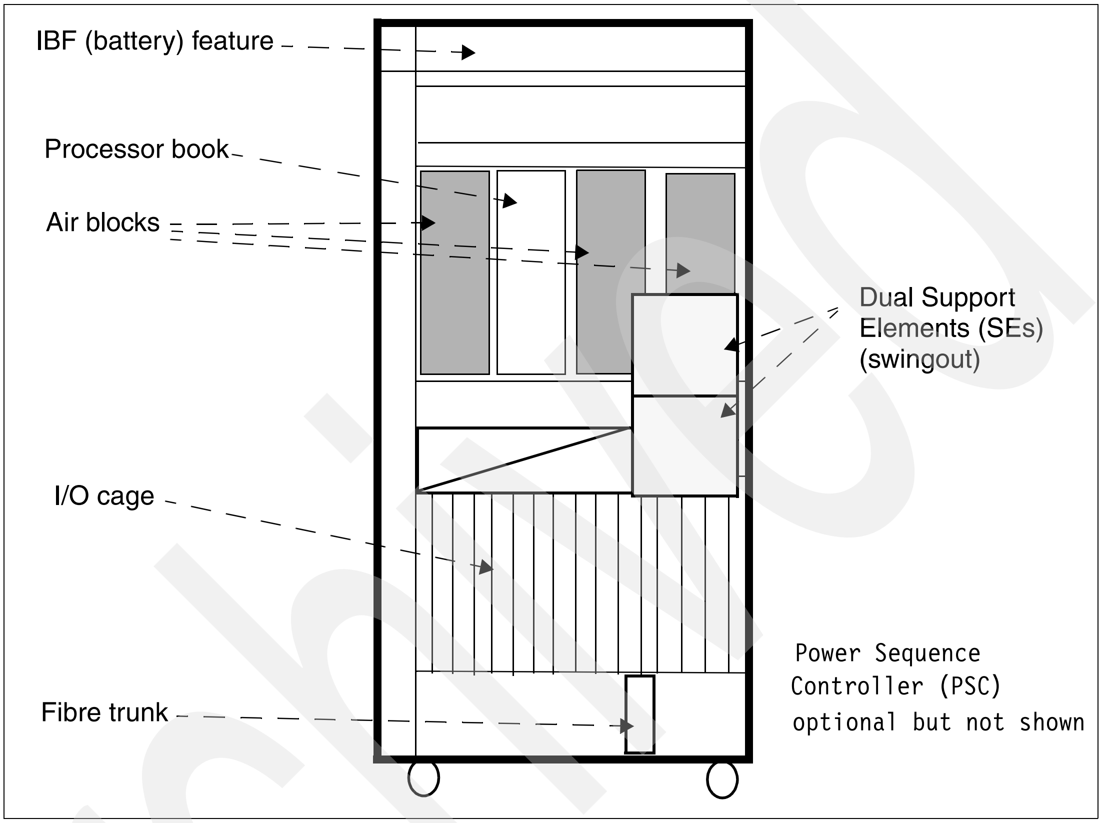

Figure 1-1 z9 BC server - under the cover - front view

The frame includes:

- One book, containing PUs, memory, and STI ports
- Two ThinkPads that are used as support elements
- One I/O cage with 16 or 28 available I/O slots (depending on model)

A more extensive description of the frames and features is provided in Chapter 2, “Hardware overview” on page 11.

Many I/O cards from a zSeries processor can be moved forward to the z9 BC server in case of an upgrade.

System z9 introduces significant new functions compared to zSeries servers, with additional functions introduced with the z9 BC server:

- zIIP specialty processors, intended to offload selected types of processing functions currently used by DB2®-related programs
- Fifteen, 30, or 60 logical partitions, depending on the system (z9 EC or z9 BC) and the model of the z9 BC server
- Multiple Subchannel Sets (MSS)

- FICON Express4
- N_Port ID virtualization
- Several new or changed instructions
- Availability enhancements
- Enhanced driver maintenance
- RoHS compliance

1.3 z9 BC models

There are two models of the z9 BC server. Both have many capacity levels. In addition, the larger model has more logical partitions, potentially more CPs, and more usable I/O slots. Key facts include:

- Both models contain eight PUs.
- Both models include an SAP® processor, leaving a maximum of seven PUs than can be characterized:
  - Model R07 can contain from one to three CPs, with the remaining PUs being specialty processors or unused.
  - Model S07 can contain zero to four CPs, with the remaining PUs being specialty processors or unused.
- There is no dedicated spare PU in the z9 BC server. However, uncharacterized PUs are used as spares.
- Model R07 provides up to 15 logical partitions. Model S07 provides up to 30 logical partitions.
- Model R07 provides up to 16 I/O slots. Model S07 provides up to 28 I/O slots.
- Both models provide up to 64 GB memory.

1.3.1 Capacity comparisons

There are 73 possible combinations of capacity levels and numbers of processors. These offer considerable overlap in absolute capacity, provided different ways. For example, a specific capacity (expressed as MSUs) might be obtained with a single faster CP or with three slower CPs. The hardware cost is approximately the same. The single-CP version might be a better choice for traditional CICS® workloads (single task), and the three-way server might be a better choice for mixed batch workloads.

The Large System Performance Reference (LSPR) should be referenced when considering performance on the z9 BC. 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 that the z9 BC has been optimized to run many workloads at high utilization rates.

The LSPR contains the Internal Throughput Rate Ratios (ITRRs) for the z9 BC 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 varies 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 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 LPAR configuration from 1-way to 54-way, based on customer profiles. This LPAR configuration table is used to establish single-number metrics.

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

The actual throughput that any user will experience varies, 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.

A key characteristic of the z9 BC server is that the processor hardware price is based on the total processor capacity purchased, not the number of processors purchased. The processor capacity purchased can be referenced as the *high-water mark*.

Basic rules for working with these capacity settings and high-water marks include:

- The configuration of the server can be changed, but it cannot exceed the purchased high-water mark. (There is a fixed service charge for supplying the necessary MES.)
- There cannot be a switch between server models (R07, S07) when making such changes (except CBU).
- The purchased capacity (high-water mark) may be upgraded.
- All CPs in a server must operate at the same capacity level. That is, there cannot be two processors with one at capacity A and the other at capacity B.
- Specialty processors always run at full capacity. (Full capacity for specialty engines on a model R07 is the same as full capacity on a model S07.)
- Server processor pricing is in terms of capacity instead of number of processors, but software pricing and hardware maintenance may be based on the number of processors. Planning may be needed to obtain the best price/capacity/performance when all cost factors are considered.

The large number of capacity levels is reflected in a large number of feature numbers used to order a system, feature numbers used to order Capacity Back Up (CBU) processing, numbers for CP capacity markers, and feature numbers for CP capacity pricing. We have omitted these feature and indicator number tables in this book.

The unusual flexibility in terms of capacity settings and variations (number of CPs) to obtain the desired capacity is in direct response to customer requests for such flexibility to help optimize software costs.

1.4 System z comparisons

Table 1-1 shows a brief comparison of several recent System z servers to the z9 BC servers.

Table 1-1 Comparison of zSeries servers

	z800	**z890**	**z9 BC model R07**	**z9 BC model S07**	**z9 EC**
Machine type	2066	2086	2096	2096	2094
Number of PUs	5	5	8	8	up to 64
Max characterized PUs (excl std SAP)	4	4	7	7	54
Standard SAPs	1	1	1	1	2 per book
Number CPs	0–4	0–4	1–3	0–4	0-54
Number specialty processors	0–4	0–4	0–6	0–7	0-54
Spare PUs	Unused PUs	Unused PUs	Unused PUs	Unusedunu sed PUs	2/server
Max memory	32 GB	32 GB	64 GB	64 GB	128 GB per book
Cycle time	1.6 ns	1.0 ns	.7 ns	.7 ns	.58 ns
Maximum logical partitions	15	30	15	30	60
Maximum I/O cages	1	1	1	1	3
I/O slots/cage	16	28	16	28	28
L1 cache (processor)	256 K/256 K	256 K/256 K	256 K/256 K	256 K/256 K	256 K/256 K
L2 cache (book)	8 MB	32 MB	40 MB	40 MB	40 MB
Multiple Channel Subsystems (CSSs)	1	2	2	2	4
Subchannel sets per CSS	1	1	2	2	2
I/O Bandwidth (GBps)[a]	6 GB	16 GB	43 GBps	43 GBps	43–172 GBps
Number STIs Speed each STI	6 (1.0 GBps)	8 (2.0 GBps)	16 (2.7 GBps)	16 (2.7 GBps)	16-64 (2.7 GBps)

a. The number shown is the number of STIs multiplied by the speed of the STI. It is unlikely that any practical system or application would stress these numbers. Perhaps a better title would be something like *memory bandwidth to I/O interfaces*.

Tables such as this can be interesting, but it is important to understand the basic details behind each of the numbers. Most of the numbers are meaningful only when used in the context of IBM System z servers.

There is special Z00 Capacity Marker for a server with no CPs. This is typically a stand-alone CF or IFL-only server.

1.5 Specialty processors

Specialty processors include IFLs, ICFs, zAAPs, zIIPs, and additional SAPs. Every z9 BC has one SAP by default, leaving seven PUs for other purposes. Briefly, the purposes are:

- CPs provide the standard processors for operating systems and applications.
- IFLs are intended for Linux® and z/VM® processing. z/OS (or z/VSE™ or z/TPF) cannot be run on an IFL.
- zAAPs are used to offload Java processing from z/OS.
- zIIPs are used to offload specialized processing elements from z/OS, where these elements are initially associated with various types of DB2 processing. zIIP processors are available for z9 BC and z9 EC, and are further described in 2.4.1, “PU characterizations” on page 14.
- ICFs are Coupling Facility processors and are automatically loaded with the appropriate internal software when started.
- SAPs provide internal I/O assistance processing. Systems working with extreme I/O loads (such as a z/TPF environment) may benefit from additional SAPs.

Specialty processors are not normally counted as *processors* for software pricing purposes. For example, a z9 BC with two CPs, two IFLs, and one zAAP is usually considered to have two processors when determining the costs of system software. In this case, there are two processors for full z/OS processing, one additional processor that provides assistance in z/OS Java processing, and two processors used for Linux (or Linux under z/VM and z/VM itself).

z/VM V5R3 and later support zAAP and zIIP processors for guest exploitation.

The standard processors (CPs) in a z9 BC have their capacity settings. Specialty processors always operate at the maximum capacity and are not affected by the capacity settings of the CPs.

The z9 BC processor configuration rules include:

- Both models use one PU as a standard SAP.
- Model R07 must contain one CP. Model S07 is not required to contain a CP.
- Model R07 has a maximum of three CPs. Model S07 has a maximum of four CPs.
- The remaining PUs (seven minus the number of CPs) can be specialty processors.
- The number of zAAPs cannot exceed the number of CPs.
- The number of zIIPs cannot exceed the number of CPs.
- Unused PUs become spare PUs. There are no dedicated spare PUs.
- The number of ICFs and IFLs (and additional SAPs) is limited only by the number of available PUs.

1.6 Upgrades

To protect investments in zSeries technology, the upgrades shown in Figure 1-2 are offered. The significance of an upgrade is that, in many cases, it is not considered a replacement for financial depreciation purposes. Note that a z9 BC model R07 server cannot be directly upgraded to a z9 EC server. Upgrades from z900 and z990 to a z9 BC server are not possible. A downgrade from a z9 BC server model S07 to a model R07 is not possible.

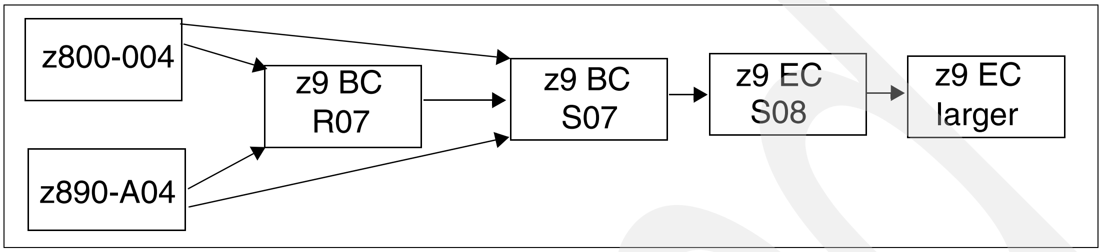

Figure 1-2 Upgrade paths

1.7 Considerations

Several other considerations are involved when migrating from earlier systems to a z9 BC. These include:

- Parallel (bus and tag) channels are not available. Converters (which convert an ESCON channel to a parallel channel) must be used. These include the converters from Optica Technologies, Inc.
- Hardware cryptographic functions are implemented differently:
  - CCF cryptographic coprocessors are not available.
  - The PCICC, PCICA, and PCIXCC cryptographic features used with earlier zSeries servers are not used with z9 EC and z9 BC servers. They have been replaced with the Crypto Express2 feature.
- Some earlier I/O features may be carried forward into z9 BC servers, but others cannot. The following cannot be used with the z9 BC:
  - OSA-Express Token Ring feature
  - OSA-2 features (FDDI and Token Ring)
  - ICB-2 feature
  - OSA-Express 155 ATM features
  - ISC-3s may not be defined in compatibility mode, they must be in peer mode (that is, CFS/CFR CHPID types)
- The SE and HMC application is completely new with the System z9 products and is no longer based on OS/2®. While the general concepts and functions remain approximately the same, the displays and detailed operational characteristics are somewhat different. Some training and time for familiarity may be needed for operators familiar with previous HMC techniques.

Migration

Migration from a z890, which supports multiple channel subsystems, should be straightforward. Migration from earlier systems requires extending existing HCD/IOCP/IOCDS definitions to include the physical channel identifiers (PCHIDs) that are associated with the channel subsystems. This may be the most complex task of a migration to a z9 BC server.

2

Hardware overview

IBM System z9 servers represent an evolution beyond the zSeries servers, with the z9 BC server generally being a smaller version of the first System z9 server. This chapter describes key hardware elements and compares them with zSeries servers where relevant. Figure 2-1 illustrates the conceptual relationships of the major elements that are described in this chapter. (This figure should not be taken as anything more than a basis for discussion.)

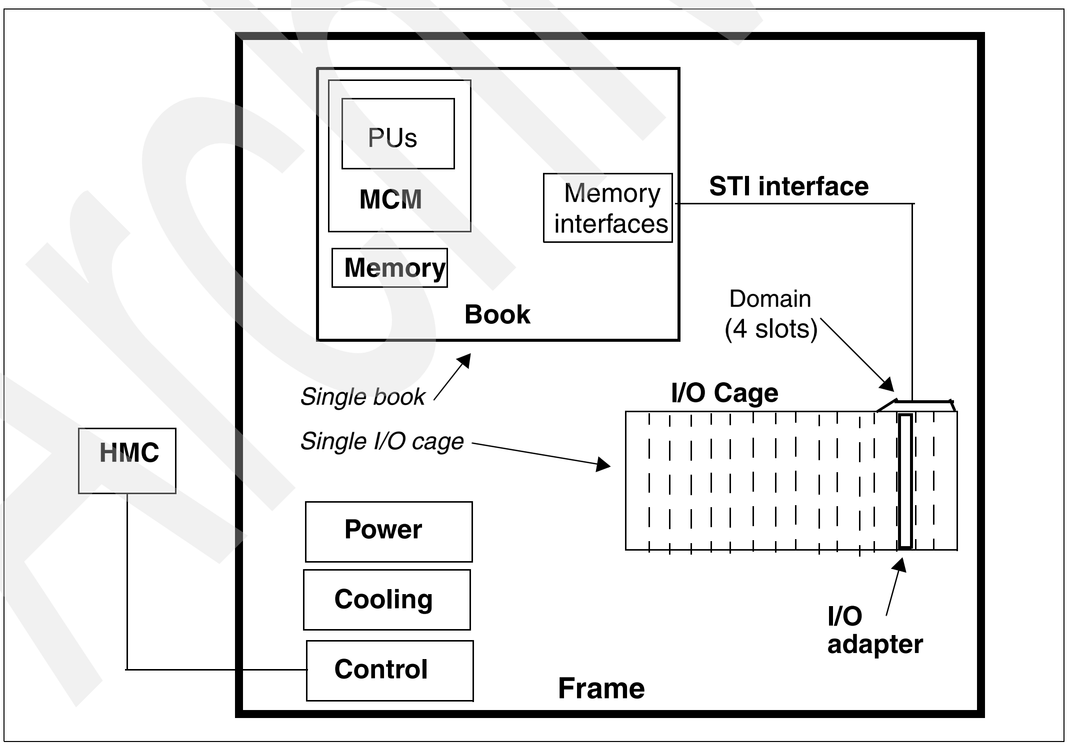

Figure 2-1 Conceptual z9 BC server component structure

2.1 System frames

The z9 BC server frame is very similar to a z890. The footprint, service clearances, and channel cabling are the same. There is only a single frame with a single I/O cage. The frame is 40 EIA units tall. The two EIA units may be removed if necessary. The approximate dimensions are:

- 158 cm deep (62 inches), including the covers
- 194 cm high (76 inches), including the casters
- 78 cm wide (31 inches)
- 699 kg (1542 pounds) for the A frame (without the internal battery feature)
- 785 kg (1730 pounds) for the A frame (with the internal battery feature)

These approximate weights assume a full load of I/O features but do not include the covers. The weights may change slightly depending on the exact mix of I/O features.

2.2 Books

A z9 BC server has one *book*. A book contains processors (on an MCM), memory, L2 cache (also on the MCM), and memory buffer interfaces for I/O connections. Figure 2-2 roughly outlines the R07 server book design.

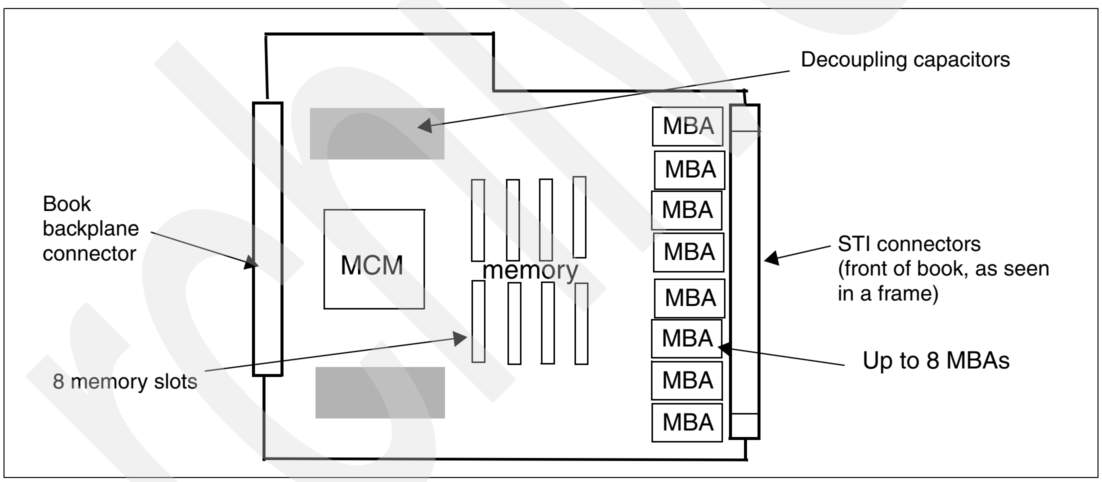

Figure 2-2 Book card conceptual layout

The primary circuit board in the book is approximately 38x47 cm. It is placed in a unit (the book) that provides support for the various wiring and cooling connectors. This book is more open than what was used in z990 and z890 servers. The *processor cage* contains four book slots. Only one slot is used for the z9 BC and the other three slots have *air blocks*.

The book contains up to eight I/O *fanout* cards. Each fanout card contains a Memory Bus Adapter (MBA) module and two STI ports for a maximum of 16 STI ports. The STI ports each operate at 2.7 GBps.[1] Figure 2-3 shows the conceptual fanout arrangement, from the memory subsystem to I/O adapters.

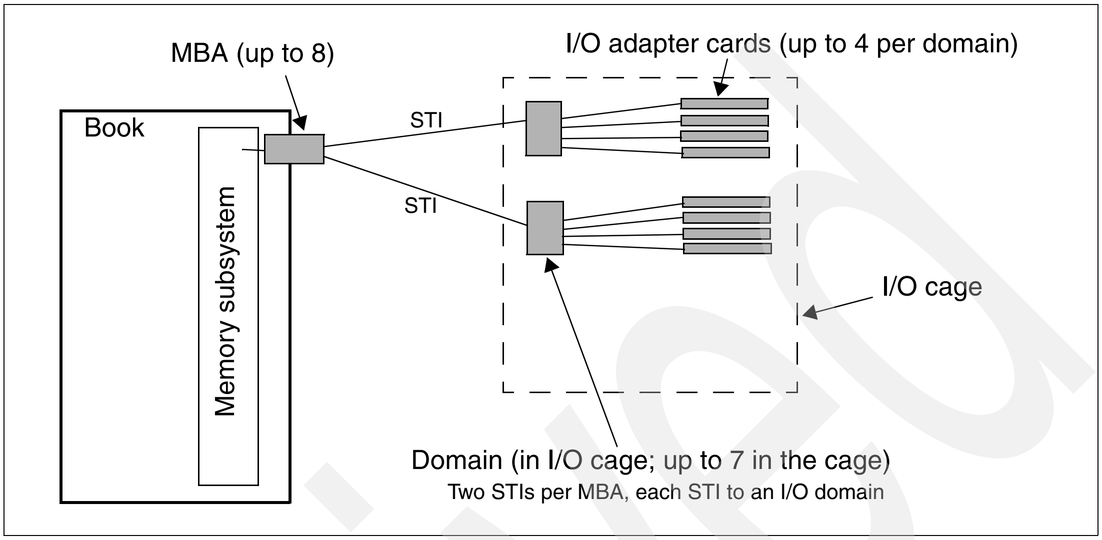

Figure 2-3 Fanout from memory to I/O adapters

The configurator determines the number of MBA elements and the arrangement of I/O adapters in domains. The I/O cage for the z9 BC model R07 supports a maximum of 16 usable slots.

2.3 MCM

The Multiple Chip Module (MCM) is the heart of the server. An MCM contains all of the processor chips and L2 cache of a book. The MCM itself is a 102-layer carrier that is the product of very sophisticated materials engineering. Figure 2-4 shows the layout of chips on an MCM.

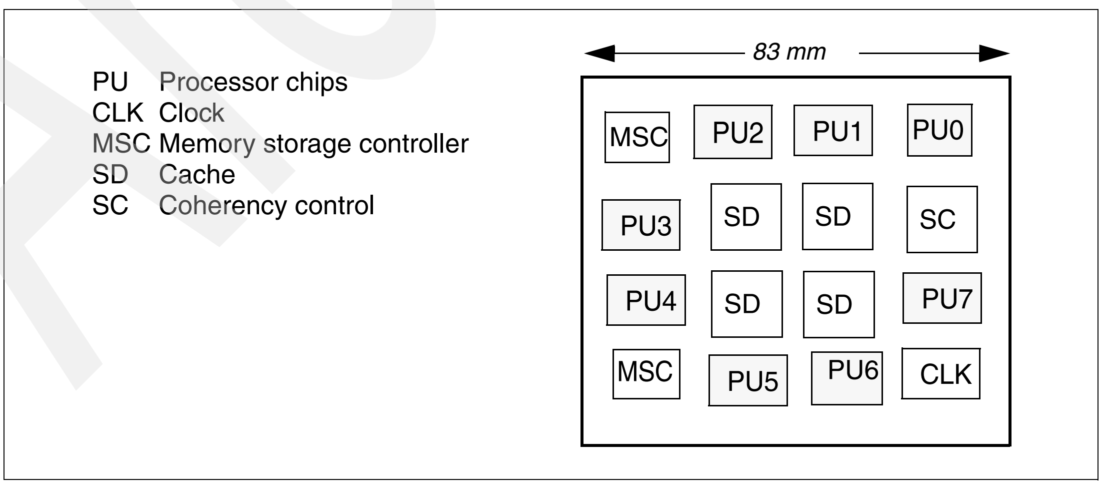

Figure 2-4 MCM layout

[1] Except when the STI is used for ICB-4 or ICB-3 coupling, in which case it operates at 2.0 GB per second or 1 GBps, respectively

The MCM contains eight processor units, each with a single active processor core. One PU is always used as an SAP. The remaining PUs may be characterized in various ways, depending on the model and features ordered with the server, or may remain uncharacterized. No PUs in the server are reserved as spares.

2.4 Processor units

Each PU has an internal mirror function in which most instructions are, in effect, executed twice for Reliability, Availability, and Serviceability (RAS) purposes. This internal mirror function in each PU should not be confused with dual-core chips, which are not used with the z9 BC server. Each PU is a superscalar processor with these characteristics:

- The basic cycle time is approximately 700 picoseconds.
- Up to two instructions may be decoded per cycle.
- Up to three instructions may be executed (finished) per cycle.
- Instructions are completed in the order in which they appeared in the instruction stream.
- Memory accesses might not be in the same instruction order.
- Most instructions flow through a pipeline with different numbers of steps for various types of instructions. Several instructions may be in progress at any instant, subject to the maximum number of decodes and completions per cycle.

Each PU has two L1 caches: a 256 KB cache for instructions and a different 256 KB cache for data. Each L1 cache is four-way set associative and has its own Translation Lookaside Buffer (TLB) of 512 entries (also four-way set associative). There is also a secondary TLB with 512 entries at the first virtual memory segment (or region) level and 4096 entries at the page level. The secondary TLB is, in effect, 16-way set associative. A branch history table, with 8 K entries, is four-way set associative and enhances the address translation function. The general address translation functions have been moved to a separate engine (driven by pico code) for more flexibility in handling address translation.

2.4.1 PU characterizations

Processor units are ordered in single increments. The book always has seven processors (plus the one used as an SAP), but they are activated only when ordered. The internal server functions, based on the configuration ordered from IBM, *characterize* processors into various types during a power-on reset (POR) function. A processor that is not characterized cannot be used. A PU may be characterized as follows:

CP (Central Processor) The standard z9 BC processors. For use with an operating system and user applications. They may be used for any z9 BC OS, including Linux.

ICF[2] Internal Coupling Facility processors, which are dedicated for this purpose.

IFL[2] Integrated Facilities for Linux. Used for Linux processing and for z/VM processing in support of Linux. (The intention is that z/VM is used to host multiple Linux guests.) It is not possible to IPL operating systems other than z/VM or Linux on an IFL.

SAP System Assistance Processors. May be used to increase I/O capacity. A z9 BC server contains one SAP processor by default. Up to three more SAPs may be added by characterizing PUs as SAPs.

zAAP[2] System z Application Assist Processors. May be dispatched by z/OS to execute JVM™ functions. They may not be used for other z/OS work (or by any other OS).

[2] ICF, IFL, zAAP, and zIIP are not counted when establishing the capacity of the server (often based on the number of CPs) for software pricing.

zIIP[2] System z9 Integrated Information Processors. May be dispatched for specialized functions. They may not be used for general z/OS work (or by any other OS).

Uncharacterized PUs may be ordered for future use through a number of different plans that allow temporary or permanent upgrading of the server.

Customers can purchase one zAAP, one zIIP, or both for each CP (assigned or unassigned) on the system. However, a logical partition definition can use more zAAPs or zIIPs than CPs. For example, in a server with two CPs you can have a maximum of two zAAPs and two zIIPs. In a logical partition definition you could have one CP, two zAAPs, and two zIIPs.

Note: It is not possible to IPL a logical partition with zAAP or zIIP only. A CP or IFL is needed to IPL.

The z9 BC server processor characterization functions are the same as for the z9 EC server and, excluding zIIPs, the same as for z890 servers. However, the ability to restructure the characterizations is more sophisticated than with earlier servers.

Concurrent PU conversions

Assigned CPs, assigned IFLs, unassigned IFLs, ICFs, zAAPs, and zIIPs may be converted to other assigned or unassigned types through a standard MES. Unlike the z9 EC, the z9 BC server supports conversion from unassigned IFLs to any other PU type (summarized in Table 2-1). Most conversions are nondisruptive. In exceptional cases the conversion may be disruptive (for example, when a z9 BC server model S07 with four CPs is converted to an all-IFL system). In addition, a logical partition may be disrupted when PUs must be freed before being converted.

Table 2-1 Concurrent PU conversions

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

Capacity indicator (CI)

The STSI instruction returns a value that can be used as an indicator to determine the number and capacity of assigned CPs. Characterization of a PU as an IFL, ICF, zAAP, or zIIP is not reflected in the output of the STSI instruction.

Granular capacity

The z9 BC server offers 73 capacity indicators in 26 capacity levels.

2.5 Memory

Each z9 BC server book contains eight slots for memory cards. Memory cards are produced in 2, 4, and 8 GB sizes. Either four or eight cards are present in the book. The memory cards in the book must be the same size. IBM may install more physical memory than a customer has purchased. The additional memory may be activated through an additional order to IBM.

Table 2-2 Memory configuration

Purchased memory	Physical memory	Memory configuration
8 GB	8 GB	4x2 GB cards
16 GB	16 GB	4x4 GB cards
24 GB	32 GB	8x4 GB cards
32 GB	32 GB	8x4 GB cards
40 GB	64 GB	8x8 GB cards
48 GB	64 GB	8x8 GB cards
56 GB	64 GB	8x8 GB cards
64 GB	64 GB	8x8 GB cards

The largest memory possible for a z9 BC server is 64 GB.

The base server is supplied with 8 GB of memory, and additional memory is ordered in 8 GB increments. IBM selects the combination of memory cards that will be used. If more physical memory is installed than purchased, a memory upgrade (within the physically installed memory) may be ordered and installed concurrently with server operation.

Depending on actual configuration, some memory upgrades may be disruptive (physical installation) and some may be done concurrently (LIC upgrade). In Table 2-3, greyed cells indicate possibilities where an upgrade can be done concurrently.

Table 2-3 Disruptive versus concurrent memory upgrades for z9 BC

From	To	To	To	To	To	To	To
8 GB	16 GB	24 GB	32 GB	40 GB	48 GB	56 GB	64 GB
16 GB	-	24 GB	32 GB	40 GB	48 GB	56 GB	64 GB
24 GB	-	-	32 GB	40 GB	48 GB	56 GB	64 GB
32 GB	-	-	-	40 GB	48 GB	56 GB	64 GB
40 GB	-	-	-	-	48 GB	56 GB	64 GB
48 GB	-	-	-	-	-	56 GB	64 GB
56 GB	-	-	-	-	-	-	64 GB
64 GB	-	-	-	-	-	-	-

2.6 I/O interfaces

The book in a z9 BC server may have up to 16 STI interfaces. These are installed in pairs and the number provided is selected by IBM to meet the needs of the installed I/O adapters, features, and ICB-4/ICB-3 coupling links. The configuration required for the STI connections differs from zSeries servers.

In general, an STI drives a *domain* in an I/O cage. A domain consists of four adapter slots in the I/O cage. The STI links for two domains are tied together for failover purposes, as shown in Figure 2-5.

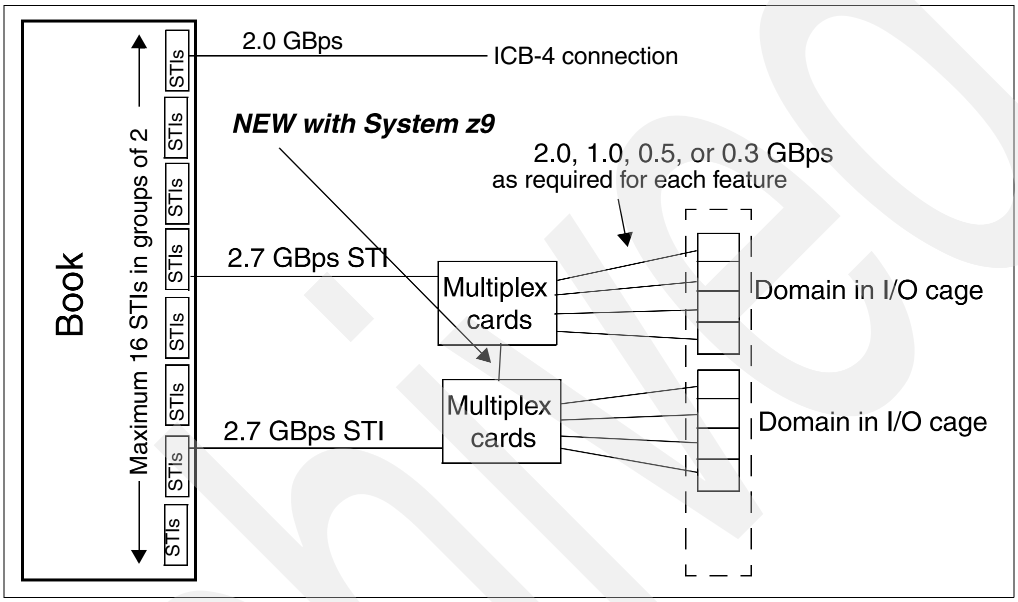

Figure 2-5 STI usage

Note the connection between the two multiplex functions in Figure 2-5. This connection provides a redundant I/O interconnect function. If one of the STI connections fails (or is disconnected for some reason), the other STI connection can service both domains in the I/O cage. In this case the data rate of a single z9 BC server STI (2.7 GBps) must be shared across both domains.

An I/O cage for the z9 BC server model R07 has four domains and the model S07 has seven domains. A fully populated I/O cage required seven STI connections in earlier servers. In order to have the redundant I/O interconnect function available to all domains, an eighth STI connection is required for a z9 BC server I/O cage in which all seven domains are used. This eighth STI connection is used only in case the primary STI for the last domain is disabled for some reason.

The z9 BC server model R07 has a maximum of 16 usable I/O slots, and these are placed in four domains. The domains are paired in order to use the redundant I/O interconnect just described. In normal cases, domains 0–3 are used for the z9 BC model R07. The following exceptions exist:

- If one of the domains in the normal 0–3 range is damaged (a bent pin, for example), then the affected domain *pair* (0/1 or 2/3) may be moved to domain pair 4/5. The pairing is needed to retain the redundant interconnect function.
- STI-3 extender cards (for ICB-3) are always placed in domain 6 (the highest numbered domain). Each of these cards reduces the number of cards that may be placed in the normal domains, such that the maximum of 16 slots for a model R07 is not exceeded.

IBM determines the number of STI connections needed and each specific connection to the I/O cages based on the number and type of I/O features ordered. A maximum of eight STIs may be used for the I/O cage. Other STIs may be used for ICB-4 connections to other servers.

2.7 I/O cages and features

The I/O cages for the z9 BC server are generally the same as for zSeries servers, except that increased power is available in the System z9 cages. This enables the maximum number of certain I/O features to be increased. Figure 2-6 shows the layout of an I/O cage.

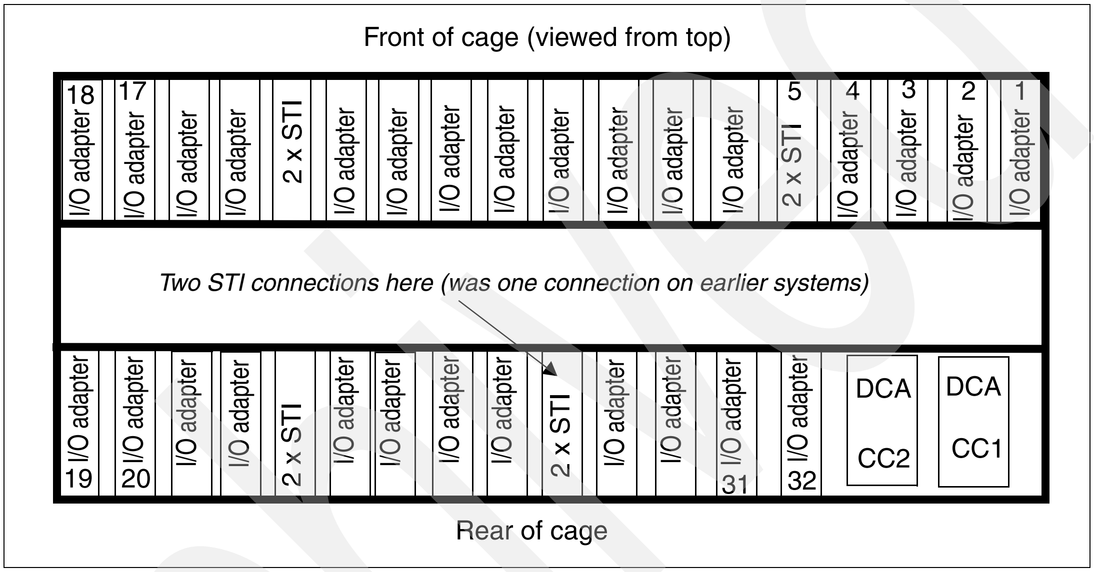

Figure 2-6 I/O cage layout

Note that slots 5, 14, 23, and 28 are used for STI connections. Assuming a fully loaded I/O cage, there are two STI connectors in each of these slots. The STI connections in slot 28 serve only one domain (slots 29–32), but two STI connections are needed to provide redundant I/O interconnect.

The full I/O cage, with seven domains, provides 28 I/O slots and is available for the z9 BC model S07. The model R07 allows use of up to 16 slots. Note that optional cryptographic adapter cards require I/O slots, although they are not I/O components.

The physical I/O cage is the same in both z9 BC models. Internal control code limits the number of slots that may be used in the model R07.

Slots are assigned to domains for each I/O cage as follows:

Domain	**I/O slots in domain**
0	1, 3, 6, 8
1	2, 4, 7, 9
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

At least one major I/O feature is required with a z9 BC server. This could be an ESCON, FICON Express, FICON Express2, FICON Express4, ICB-4, ICB-3, or ISC-3 feature. Table 2-4 shows the I/O features and interfaces used with the z9 BC server.

Table 2-4 I/O features

Feature	Max slots	Maximum connection	Max slots	Maximum connection	Ports/links per feature	Purchase increment
	Model R07[a]		Model S07			
16-port ESCON	16	240	28	420	16 ports[b]	4 ports[c]
FICON Express (carry forward)	16	32 ports	20	40 ports	2 ports	1 feature
FICON Express2 (carry forward)	16	64 ports	20	80 ports	4 ports	1 feature
FICON Express4 (4 port adapters)	16	64 ports	28	112 ports	4 ports	1 feature
FICON Express4 (2 port adapters)	16	32 ports	28	56 ports	2 ports	1 feature
OSA-Express2	16[d]	32 ports	20	40 ports	2 ports	1 feature
OSA-Express2 10Gb	16	16 ports	24	24 ports	1 port	1 feature
OSA-Express (carry forward)	12	24 ports	12	24 ports		
Crypto Express2	4	N/A	8	N/A	2 PCI-X	2 features[e]
Crypto Express2-1P	4	N/A	8	N/A	1 PCI-X	2 features[e]
ICB-4 link	N/A[f]	16 links[g]	N/A	16 links		1 link
ICB-3 link[h]	8	16 links	8	16	2 links	1 link
ISC-3	12	48 links	12	48 links	4 links	1 link
HiperSockets™	N/A	16	N/A	16	N/A	N/A
IC channel (CF link)	N/A	32	N/A	32	N/A	N/A

a. The model R07 at capacity level A01 has additional restrictions not shown in this table. It is limited to eight links for ICB-4 and to a maximum of 24 ports for any combination of OSA-Express and OSA-Express2 adapters.
b. Each ESCON feature has 16 channels, of which a maximum of 15 may be activated.
c. The purchased number of channels are activated via Licensed Machine Code Configuration Control (LCC). Channels are activated equally across all installed ESCON features.
d. In principle, 16 slots/adapter may be used. In practice, at least one adapter is needed for other I/O.
e. To support reliability, availability, and serviceability (RAS) requirements, the initial purchase increments must contain two Crypto Express2 features or two Crypto Express2-1P features, or one of each. Once the initial purchase increments are satisfied, orders for Crypto Express2-1P, in increments of one, are supported.
f. ICB-4 links do not involve an I/O cage.
g. There is a maximum of 64 CF links (ICs, ICB-3s, ICB-4s, and active ISC3s).
h. STI-3 cards (in an I/O cage) are used with ICB-3 links.

FICON Express and FICON Express2 features may be carried forward when migrating from zSeries servers. They operate at 1 or 2 GBps, depending on the link and the control unit involved. OSA-Express and OSA-Express2 features may serve various functions (CHPID types), as shown in Table 2-5.

Table 2-5 OSA-Express features

Feature	Media	Ports	Supported CHPID types
OSA-Express2 10 GbE LR	Fiber	1	OSD
OSA-Express2 GbE LX, SX	Fiber	2	OSD, OSN
OSA-Express2 1000BaseT Ethernet	Copper	2	OSD, OSE, OSC, OSN
OSA-Express GbE LX, SX	Fiber	2	OSD
OSA-Express 1000BaseT Ethernet	Copper	2	OSD, OSE, OSC

OSA-Express and OSA-Express2 features may be carried forward when migrating from zSeries servers. The external functions are the same, but the OSA-Express2 features have a more powerful internal engine. No token-ring support is provided with the z9 BC server.

The CHPID types listed in this table have the following meanings:

- OSD indicates QDIO usage by the operating system.
- OSE indicates non-QDIO usage (that is, CCW-based) by the operating system.
- OSC indicates TN3270E functions that provide support for operating system consoles and 3270 user sessions (for TSO, for example).
- OSN indicates a channel used for a link to the Communication Controller for Linux (CCL) partition on a z9 server. This function is not available on zSeries servers.

Note that LAN adapters with fiber media are used only through QDIO, while adapters with copper media may be used with either CCW or QDIO programming. The basic characteristics of the 10 Gb Ethernet adapter are as follows:

- The adapter includes a single port.
- It uses CHPID type OSD only. (This means that only QDIO operation is allowed.)
- It uses 9-micron single-mode fiber with an SC Duplex connector.

The OSA-Express2 adapters include the following enhancements:

- Each port provides up to 640 TCP/IP stacks, improving the virtualization reach of the port.
- A *large send* function enables TCP/IP segmentation to be offloaded into the adapter.
- Layer 2 support exists for protocol-independent packet forwarding.
- LIC updates may be performed concurrently with server operation.

These are I/O configuration details for the z9 BC server:

- A maximum of 64 coupling links can be used.
- Token ring and ATM features are not available.
- No ICB-2 links are available.
- FICON bridge support is not available with FICON Express2 features. It is available with the FICON Express LX feature when upgrading from zSeries servers.
- OSA-Express and OSA-Express2 features must not exceed a maximum of 24 for the server.

The OSA/SF software function is not required for OSA-Express or OSA-Express2 when operating in QDIO mode. However, we recommend its use for status and configuration information.

The following channel types are not supported on z9 BC servers:

- FICON (pre-FICON Express)
- OSA-Express Token Ring
- PCIXCC
- PCICA
- ICB-2
- ISC-3 in Compatibility Mode
- Parallel
- OSA-Express ATM 155
- OSA-2

2.7.1 FICON Express4 adapters

Two FICON adapters operating at 4 GBps are available with the z9 BC server. One version has four channels and the other version has two channels. Earlier FICON adapters had only two channels. These adapters provide faster operation (1, 2, or 4 Gbps) and more ports (four) than earlier FICON adapters.

The two-port version is roughly half the price of the four-port version and is intended to allow more redundancy in smaller systems. That is, there can be 2 two-port FICON adapters instead of a single four-port adapter, and so forth. With appropriate external switching and control unit interfaces this provides a higher-availability configuration. Two port FICON Express4 adapters are available only for z9 BC.

Speed selection is automatic, in conjunction with the other end of the connection. Adapters are available with either LX or SX interfaces. These cards introduce new optical elements that may be individually changed while the adapter is operational. These optical elements contain the fiber connector and are known as Small Form Factor Pluggable (SFP) connectors.

2.7.2 Physical I/O connections

The concept of multiple channel subsystems was introduced with z990 servers. This function is unchanged with the System z9 servers. Although it is unlikely that z9 BC system configurations will be large enough to require the use of multiple channel subsystems, compatibility with other System z servers is important for future growth and software interoperation. Note that a system with more than 16 logical partitions requires the use of multiple CSSs.

Channel subsystems add another element to the addressing of channels (CHPIDs). It introduces a physical channel identifier (PCHID) that is directly related to the physical location of the channel feature in an I/O cage. IOCDS parameters then relate PCHID numbers to CHPID numbers. CHPID numbers are defined by the user when building an IOCDS.

Each possible *physical* I/O port/connector has a PCHID number that is fixed by the physical location of the feature and connector in an I/O cage. The largest feature in an I/O cage slot can provide a maximum of 16 PCHIDs. (The ESCON feature does this, although only 15 can be used at one time.) The cage has a maximum of 28 slots. Example 2-1 shows how PCHID numbers are assigned.

Example 2-1 PCHID number assignments

```
                    PCHID numbers
I/O cage slot       I/O cage
   1                100 - 10F
   2                110 - 11F
   3                120 - 12F
   4                130 - 13F
   5                         - used for STI connections
   6                140 - 14F
   7                150 - 15F
   8                160 - 16F
   9                170 - 17F
   10               180 - 18F
   11               190 - 19F
   12               1A0 - 1AF
   13               1B0 - 1BF
   14                        - used for STI connections
   15               1C0 - 1CF
   16               1D0 - 1DF
   17               1E0 - 1EF
   18               1F0 - 1FF
   19               200 - 20F
   20               210 - 21F
   21               220 - 22F
   22               230 - 23F
   23                        - used for STI connections
   24               240 - 24F
   25               250 - 25F
   26               260 - 26F
   27               270 - 27F
   28                        - used for STI connections
   29               280 - 28F
   30               290 - 29F
   31               2A0 - 2AF
   32               2B0 - 2BF

PCHID numbers used for direct book connections: 000-0FF
```

Consider an ESCON feature in slot 15. PCHID numbers 1C0–1CF are reserved for this slot. The first ESCON connector on the feature (whether or not that particular port is enabled or used) is PCHID 1C0, the second connector is 1C1, and so forth. The PCHID numbers are fixed. The CHPID numbers are not fixed and can be assigned by user when an IOCDS is constructed. For example, an installation might decide to map PCHID 1C0 (the first ESCON channel in our example) CHPID 52 in CSS 1.

ICB-4 connections to books

Direct book connections (to STI ports) are used for ICB-4 channels. There are up to 16 STI ports on the book, and PCHID numbers 010–01F are assigned to these.

Coupling connections

The coupling connections shown in Table 2-6 are supported.

Table 2-6 Coupling connections

From/to	G5/G6	z900, z800	z990, z890	z9 EC, z9 BC
G5/G6	ISC2 ICB2 IC	ISC comp ICB2[a]	ISC comp ICB2[a]	
z900, z800	ISC comp ICB2[a]	ISC3 peer ICB-3 IC	ISC3 peer ICB-3	ISC3 peer ICB-3
z990, z890	ISC3 comp ICB2[a]	ISC3 peer ICB-3	ISC3 peer ICB-3 ICB-4 IC	ISC3 peer ICB-3 ICB-4
z9 EC **z9 BC**		ISC3-peer ICB-3	ISC3 peer ICB-3 ICB-4	ISC3 peer ICB-3 ICB-4 IC

a. ICB2 is not supported on z890 and z800.

ICB-3 usage is not recommended in situations where ICB-4 could be used because ICB-4 offers better performance. The z9s will be the last systems to support ICB-3 links.

2.7.3 Cryptographic functions

Every processor in a System z9 server has the CP Assist for Cryptographic Function (CPACF) instructions introduced with the z990. System z9 servers have extended these with the following additional functions:

- Advanced Encryption Standard operations (AES)
- Secure Hash Algorithm - 256 (SHA-256)
- Pseudo Random Number Generation
- Remote loading of initial cryptographic keys for ATMs. This can replace the on-site installation of cryptographic keys for ATMs, which often requires two people to travel to the ATM location (each with part of the key). ATMs supporting the following standards should be compatible with this function:
  - 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
- Improved key exchange with non-CCA cryptographic systems.
- ISO 16609 CBC mode Triple-DES MAC enhancement. This is a function of the Crypto Express2 adapter and involves the following services:
  - MAC Generate (CSNBMGN)
  - MAC Verify (CSNVMVR)
  - Digital Signature Verify (CSNDDSV)

Improved key exchange and CBC T-DES MAC enhancements are exclusive to System z9 servers.

To allow CPACF, no charge CPACF Enablement must be ordered. Otherwise CPACF comes as disabled.

Cryptographic adapter

Only the Crypto Express2 feature may be used with the System z9 servers. It replaces the older PCICC, PCICA, and PCIXCC features.

The Crypto Express2 feature contains either one or two PCI-X adapters (not to be confused with the two optional internal CCF coprocessors available on several earlier servers). The Crypto Express2 feature for z9 EC server always contains two PCI-X adapters. Each Crypto Express2 adapter can be configured as either:

- *Cryptographic coprocessor,* with functions similar to the previous PCICC or PCIXCC cryptographic feature, including the secure key functions. This mode is intended for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification. (This configuration can also provide accelerator functions, but not as fast as when configured as an accelerator.)
- *Accelerator,* with functions similar to those provided by the older PCICA feature. This mode is usually intended for public key handshakes, as used with Web applications.

To support reliability, availability, and serviceability (RAS) requirements, the initial purchase increments must contain two Crypto Express2 features or two Crypto Express2-1P features, or one of each. Once the initial purchase increments are satisfied, orders for Crypto Express2-1P, in increments of one, are supported.

A new Trusted Key Entry (TKE) workstation is available with the z9 BC server that supports the Crypto Express2 feature and a new graphical user interface (GUI). This new workstation provides level 5.1 of the TKE implementation.

2.8 Time functions

Time functions are used to provide an accurate time-of-day value and to ensure that the time-of-day value is properly coordinated among all of the systems in a complex. This is critical for Parallel Sysplex operation.

2.8.1 Sysplex Timer

A 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, all updates are required to be time stamped in proper sequence.

The zSeries servers can attach to either an IBM 9037 Model 1 or Model 2 Sysplex Timer Unit.[3] More information about the IBM 9037 can be found in the Redbooks publication *S/390® Timer Management and IBM 9037 Sysplex Timer*, SG24-2070.

ETR attachment

A server's External Time Reference (ETR) feature provides the interface to the IBM 9037 Sysplex Timer. The z9 BC server ETR feature is optional. When it is installed, it provides ETR connections to one or two IBM 9037 Sysplex Timer units.

[3] The IBM 9037 model 1 and model 2 were withdrawn from marketing and service support was discontinued.

The two z9 BC server ETR features (when installed) are located in the processor cage of the z9 BC server. 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 9037 Sysplex Timer unit has an optical transceiver that supports an ESCON Duplex connector.

Note: The z9 BC server ETR feature does not support a multimode fiber optic cable terminated with an ESCON Duplex connector, but 62.5 micron multimode ESCON Duplex jumper cables can be reused to connect to the ETR feature, by installing an MT-RJ/ESCON conversion kit between the ETR feature MT-RJ port and the ESCON Duplex jumper cable.

2.8.2 Server Time Protocol (STP)

Server Time Protocol is a server-wide facility that is implemented in the Licensed Internal Code (LIC) of z9 EC, z9 BC, z990, and 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, and z890 servers and Coupling Facilities (CFs). The STP design uses a new concept called Coordinated Timing Network (CTN). A Coordinated Timing Network 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.

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, 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 later.

2.9 Hardware System Area (HSA)

HSA memory is used for several internal functions, but the bulk is used by channel subsystem functions. Each subchannel in each logical partition requires 256 bytes in HSA. The HSA *could* be substantially larger on the z9 BC server than on earlier systems. This is related to the Multiple Subchannel Set function that provides approximately twice as many I/O devices as provided by earlier systems. An HSA estimation tool is available through the IBM Resource link at:

https://www.ibm.com/servers/resourcelink/hom03010.nsf/pages/tools?OpenDocument&login

2.10 System control

Figure 2-7 provides a conceptual overview of the system control design. Some details have changed, but the general structure for system control is similar to the other System z servers.

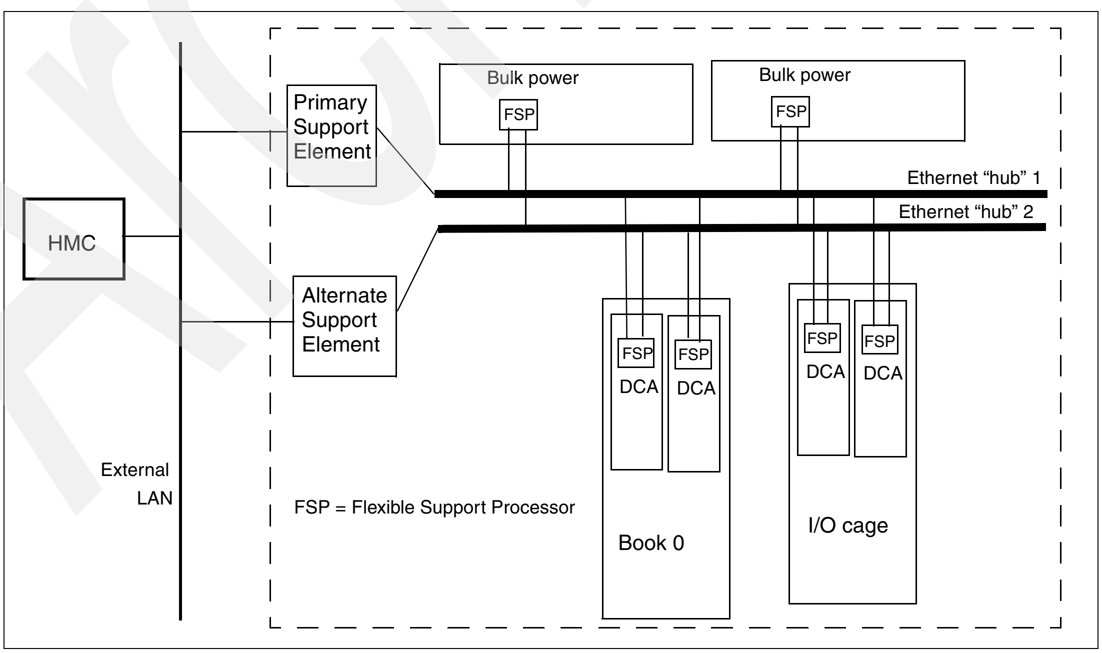

Figure 2-7 System control

Various system elements contain *Flexible Support Processors* (FSPs).[4] An FSP is based on the IBM Power PC® microprocessor. It connects to an internal Ethernet LAN (to communicate with the support elements) and provides a SubSystem Interface (SSI) for controlling components.[5]

A typical FSP operation is to control a power supply (shown as DCAs in the figure). A support element might send a command to the FSP to bring up the power supply. The FSP (using SSI connections) would cycle the various components of the power supply, monitor the success of each step, monitor the resulting voltages, and report this status to the support element.

Most system elements are duplexed (for redundancy), and each element has an FSP. There are two internal Ethernet LANs and two support elements, again for redundancy, and crossover capability between the LANs so that both support elements can operate on both LANs.

The support elements, in turn, are connected to one or two (external) LANs (Ethernet only), and the Hardware Management Consoles (HMCs) are connected to these external LANs. There can be one or more HMCs. In a production environment, the server complex is normally managed from the HMCs. If necessary, the system can be managed from either support element. Several or all HMCs can be disconnected without affecting system operation.

2.11 HMC and SE

The Hardware Management Console and support element functions of the z9 EC and z9 BC servers have a completely new implementation. OS/2 is no longer used as a base operating system, and both are *closed* systems. This means that other applications or user programs may not be added to these systems. This prohibition includes ESCON Director control programs and Parallel Sysplex Timer console application. Separate workstations are now needed for these applications. The purpose and general use of the HMCs and SEs remain unchanged, but the appearance of various panels has changed.

The HMC user interface is now implemented through a standard Web browser. Remote users may connect to the HMC using a browser, and have almost complete HMC functionality.

2.12 Concurrent maintenance and upgrades

The z9 BC server can permit a variety of concurrent maintenance and upgrades, in particular:

- Duplexed units, such as power supplies, may have one unit replaced while the other is operational.
- In many cases, new levels of driver code may be installed concurrently with server operation.
- Redundant I/O Interconnect allows for the concurrent repair of an MBA fanout card and STI cables without loss to any I/O features.

2.13 Location

The z9 BC server is suitable both for raised-floor and non-raised-floor installation. This contrasts with the larger System z servers that are designed solely for raised-floor installation.

[4] These were informally known as *cage controllers* in earlier systems.
[5] More detailed descriptions may refer to SSI-M and SSI-S notations, corresponding to master and slave functions.

Air cooling is used and care is needed in a non-raised-floor environment to ensure sufficient cooling air.

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 z frames more rugged and to tie them down to a concrete floor beneath a 9–13-inch or 12– 22-inch raised floor or in a non-raised floor installation.

- (#7995) - Bolt-Down Kit, High-Raised Floor feature provides frame stabilization and bolt-down hardware for securing a frame to a concrete floor beneath a 11.75–16.0-inch (298–405 mm) raised floor.
- (#7996) - Bolt-Down Kit, Low-Raised Floor feature provides frame stabilization and bolt-down hardware for securing a frame to a concrete floor beneath a 9.25– 11.75-inch (23–298mm) raised floor.
- (#7997) - Bolt-Down Kit, non-raised floor feature provides frame stabilization and bolt-down hardware for securing a frame to concrete non-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 BC you need one Bolt-Down kit.

2.14 Power and cooling

The z9 BC server footprint is the same as that of a z890. The power service specifications are the same, but the power consumed is slightly greater. A fully loaded z9 BC server consumes 5.4 KW and produces 18.4 kBTUs. These figures assume maximum I/O cage configurations and should be considered approximate.

A z9 BC server uses three-phase 200–480 VAC power or one-phase 200–415 VAC[6]. Two power feed cables are connected to two internal bulk power units that provide power for the server. The server can continue operation after a failure of one of these units. They are designed such that the server will continue to operate after a failure of one phase of a three-phase feed.

Internal functions adjust the power factor to almost 1.0, with little harmonic content. (This complies with increasing demands, sometimes in the form of binding requirements, for *building-friendly* power supplies.)

Incoming power is converted to high-voltage DC, which is distributed to the subsidiary power units that convert it to the various voltages needed by the controllers, cages, book, and I/O features. Very large currents are needed at low voltages and it is not practical to directly distribute the low voltage supplies throughout the server.

Units known as DCAs convert the distributed high voltage to various low voltages. Slightly different DCAs are used in the processor cage and the I/O cage. Two DCAs are used for a book, and the I/O cage contains two other DCAs. Each DCA has a Flexible Support Processor that is part of the system control function discussed in 2.10, “System control” on page 26.

[6] The voltage ranges are 200–240, 380–415, or 480.

zPower Estimation Tool

The power consumption tool for System z9 is available via IBM Resource Link™ Web site. This tool provides an estimate of the anticipated power consumption of a particular server model and its associated configuration. A user will input the server 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 IBM System z9 servers.

2.14.1 Cooling

The z9 BC server is air cooled. A sophisticated arrangement of fans (Air Movement Devices, or AMDs) and sensors provides the operation. Fan speeds are variable and are controlled to maintain required cooling. The server heat load must be managed by the external air system.

2.15 Internal Battery Feature

The Internal Battery Feature is optional on the z9 BC server. In the event of an interruption to the input power, the internal battery will provide sustained server operation for a short period of time. The duration of the battery support is highly dependent on the server model and I/O cages installed.

Many power interruptions are due to the temporary loss of a single phase on the three-phase input line. The z9 BC server can tolerate loss of a single phase even with no battery backup. This capability, coupled with the Internal Battery Feature and full operation from either of the dual independent line cords, gives a very high degree of resilience to transient power dropouts. When coupled with a UPS capability, the internal battery ensures that there is no loss of power during startup of the emergency supply.

3

Key functional elements

To a large extent, the z9 BC server is a smaller version of the currently available z9 EC server. The new functions that are unique to the z9 BC are relatively few:

- Extensive capacity granularity
- A FICON adapter with two ports, operating at 1, 2, or 4 Gbps
- Crypto Express2 card with one adapter

However, we assume that the primary interest in z9 BC servers is by those currently using G5/G6, MP3000, z800, and z890 servers. The many new functions and features previously introduced with the z9 EC may be unfamiliar to these readers. For this reason we include abbreviated descriptions of many of the z9 BC features and functions that are common with z9 EC servers. These include:

- A new specialty processor, the zIIP
- A new FICON adapter, with either two or four ports, operating at 1, 2, or 4 Gbps
- Improved FICON error recovery
- Enhancements for cryptographic functions
- Single processor Crypto Express2
- Extended environmental specifications
- RoHS compliance
- Availability enhancements
- Enhanced driver maintenance
- Redundant I/O Interconnect
- Multiple channel subsystems
- Multiple subchannel sets
- System-initiated CHPID reconfiguration
- Multipath IPL
- OSA for NCP
- GARP VLAN Registration Protocol (GVRP)
- OSA-Express2 Link Aggregation for z/VM
- OSA Layer3 VMAC support
- OSA-Express2 Network Traffic Analyzer
- QDIO Diagnostic Synchronization
- OSA Dynamic LAN Idle
- MIDAW facility
- N_Port ID virtualization

- FCP performance enhancements
- FCP performance metrics for Linux
- New and changed instructions
- Hardware Decimal Floating Point
- Last branch trace
- Program-directed re-IPL
- Open exchanges for FCP and FICON
- CHPID mapping
- More logical partitions
- LPAR Group Capacity Limit
- Hardware Management Console and support element

3.1 Capacity granularity

Software prices for mainframe applications can be very sensitive to system configurations. The capacity granularity of the z9 BC server, as shown in 1.3.1, “Capacity comparisons” on page 5, enables the user to select a capacity that provides the best balance between system capacity and software costs. The z9 BC is unusual in that the hardware price is based on capacity and not the number of processors. For a given capacity and cost, the user has choices of more (and slower) processors or fewer (but faster) processors.

3.2 zIIP

The z9 EC and z9 BC servers introduce a new specialty processor, the System z9 Integrated Information Processor (zIIP). It is designed to help improve resource optimization for eligible data workloads within the enterprise. It most closely resembles the existing specialty processor called System z Application Assist Processor (zAAP). Neither can be used to IPL an operating system or sustain system operation by itself. For example, general interrupts cannot be fielded by these processors, I/O cannot be managed, timers cannot be set, and so forth. However, other than these operating system-related functions, the full instruction set is available to zAAP and zIIP processors. The operating system is responsible for dispatching appropriate work to these processors. Only z/OS is able to dispatch work on zAAP and zIIP processors.

Similar to a zAAP, the z/OS dispatcher can direct certain work to a zIIP. In more specific terms, z/OS uses a zIIP only to handle SRB routines that are scheduled with certain parameters and are part of a defined enclave.

Users of the zIIP include:

- DRDA®: can offload up to approximately 40% of a given DRDA transaction into a zIIP.[1] (Multiple concurrent DRDA transactions can *each* offload up to approximately 40% of their workload.)
- Other elements of DB2 can also schedule work to zIIPs. Utilization information was not available when this book was being written.
- Data warehousing applications: Requests that utilize DB2 for z/OS V8 complex star schema parallel queries may have portions of these SQL requests directed to zIIP.
- DB2 for z/OS V8 Utilities: A portion of DB2 utility functions used to maintain index maintenance structures typically run during batch can be redirected to zIIP.

The amount of eligible work offloaded to a zIIP is controlled by the application (DB2 elements, for example), WLM, and the z/OS dispatcher. There is a complex mix of requirements for balancing eligible work between zIIPs and standard CPs.

z/OS includes zIIP *qualified time* and actual time in SMF type 30 records.

System z servers cannot be configured with more zIIP processors than CP processors.

Prerequisites for initial zIIP usage are:

- z/OS 1.6 or later
- DB2 for z/OS V8
- z9 EC or z9 BC with zIIP processors

[1] As with any early description, actual performance or offload percentages may be different in delivered systems. The numbers presented here are based on early experience.

3.3 New FICON adapters

FICON Express4 adapters have been introduced with the z9 BC server. The standard FICON Express4 adapter contains four channel ports. A second version, with two channel ports, is available only for the z9 BC. The smaller version provides an economical way of how to configure redundant FICON adapters for a small server. These adapters operate at 1, 2, or 4 Gbps and provide automatic speed sensing.

3.4 Improved FICON error recovery

System z9 combined with z/OS V1.7 I/O recovery processing improvements are designed to allow the system to detect fabric problems that may cause FICON links to fail and recover many times in a short period of time. The fabric problems may cause system recovery actions to be repeatedly driven while substantially reducing throughput for those FICON links.

This enhancement is designed to allow the system to detect these conditions and keep an affected path offline until an operator action is taken. This is expected to help limit the performance impacts of fabric problems.

3.5 Extended environmental specifications

The z9 BC server can be configured for either three-phase or single-phase power. In either case, two power feeds are used, and these are intended to be from separate sources.[2] The minimum nominal voltage used is 220 volts.

A raised floor is not required for the z9 BC server, although it is considered to be the normal environment. The system is air cooled, so air flow and heat loads should be considered carefully for non-raised-floor installations. (The non-raised-floor configuration is not available for initial shipments.)

3.6 Availability enhancements

The z9 EC and z9 BC servers include availability enhancements. Several of these are discussed separately throughout this book. For completeness, we list the primary availability enhancements that are relevant to the z9 BC server here:

- *Redundant I/O interconnect* groups I/O cage domains into pairs and enables the STI link to each domain to service the other half of the pair if the STI link to that domain fails or is not operational for some reason. The STI elements (also known as MBA connections) may be removed and replaced concurrently with server operation. Each element contains two STI connections.
- *Dynamic oscillator switchover* enables the backup system oscillator to transparently take over system clocking if the primary oscillator fails. Earlier servers had a backup oscillator, but a power-on reset (POR) was needed to use it.
- *Enhanced driver maintenance* enables selected driver upgrades to be installed concurrently with normal operation. This may be possible when moving from selected *sync points* of driver *n-1* to a corresponding sync point in driver *n*. This function depends on the driver sync points provided by IBM and may not be available for some upgrades.

[2] The use of separate power sources is not enforced by the system, of course, and is provided to enable users to configure a higher-availability environment.

- The *memory sparing* function used with the memory design of the z9 EC server provides more spare chips on the memory cards than were used in earlier designs. The design makes it very unlikely that DRAM failures will affect server operation.
- Improved displays for *FICON link status* provide additional information for debugging FICON connection issues.
- *FICON link recovery thresholds* (a software function) reduce the potentially severe software overhead created by multiple link failures.
- The individual small form-factor pluggable (SFP) optical element on the FICON Express4 cards can be changed without disrupting activity on the other ports on the card.

3.7 Enhanced driver maintenance

Some server driver[3] transitions can be installed concurrently with server operation. These concurrent transitions occur only at IBM-defined *sync points*, as shown in Figure 3-1.

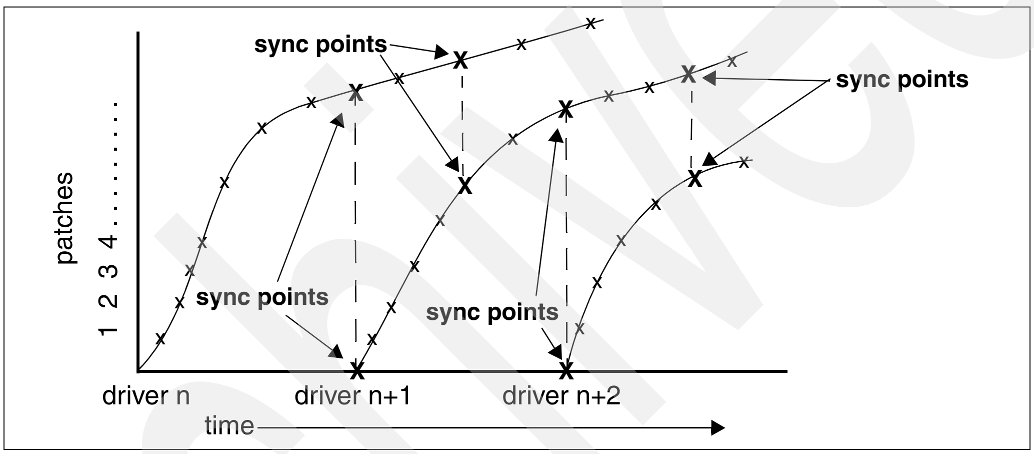

Figure 3-1 Driver sync points

The sync points are defined only at selected patch levels of various drivers. Concurrent transitions are available only for single-level jumps; that is, from driver *n* to *n+1*. A concurrent jump from driver *n* to *n+2*, for example, is not available. Also, a concurrent move to a previous driver level is not possible.

IBM generally intends to define a limited number of sync points (crossover points) for each driver. However, this does not preclude the potential occurrence of disruptive driver upgrades. These will be documented as far in advance as possible.

Certain manual actions may be needed after a concurrent driver upgrade. These include:

- Non-QDIO OSA-type CHPIDs must be varied off and varied on again to activate any new code. In particular, this applies to CHPID types OSE and OSC.
- If new basic FICON adapter LIC is present in the new driver, then FICON and FCP CHPIDs must be reset to activate the new code.

3 A *driver* is the IBM Licensed Internal Code that provides many server functions that are generally regarded as hardware functions.

3.8 Redundant I/O interconnect

STI ports (used for connections between a book and I/O cage domain connections) are installed in pairs, and pairs may be installed or removed concurrently with server operation. STI port removal assumes, of course, that there is sufficient I/O connectivity or configuration flexibility to withstand the removal of the STI connections.

Furthermore, I/O cage domains are now paired (with the last domain in a full I/O cage acting as its own pair). The two STI connections to the two domains (in a pair) provide automatic backup for each other. If one of the STI connections fails or is removed, the STI connection for the other member of the pair will handle all I/O adapters in both domains. This allows concurrent repair of the MBA fanout card or the STI cable without the loss of I/O functions. In this case the potential exists that the data rate of all of the adapters might overrun the bandwidth of the single STI connection. In this case, some I/O operations might be delayed slightly.

The redundant I/O interconnect function does not apply to ICB-4 or ICB-3 connections.

3.9 Multiple channel subsystems

Important architectural change associated with the z9 EC and z9 BC servers (and the z990 and z890 servers) is the set of extensions to the channel subsystem. Perhaps the best way to start a description is to explain *why* these extensions were introduced. The purpose is simple. It is to permit a server to have more than 256 channels. The maximum number of channels has been a concern for some time, although the move to FICON channels and devices has mitigated the limitation. Note that use of multiple CSSs is required for systems with more than 16 logical partitions.

The difficulty is to extend this number while maintaining compatibility with existing software. The 256 maximum CHPID number is reflected in various control blocks in operating systems, software performance measurement tools, and even some application code. Simply changing the control blocks is not a viable option because this would break too much existing code. The solution introduced with the z990 is in the form of multiple channel subsystems. These are implemented in a manner that has little or no impact on existing code.

It is important to note that there is no unique hardware (or feature code) associated with multiple CSSs. These are not priced features. These functions are implemented in the Licensed Internal Code of the server and are, potentially, exploited by operating systems running in the server.

Exceeding 256 channels might not be a concern for most z9 BC customers, but software compatibility with other current System z servers should be a concern. The z9 BC implementation of channel subsystem is compatible with the other current System z servers.

Existing code works with single-byte CHPID numbers, producing the limitation of 256 CHPIDs. The multiple channel subsystems provide *multiple sets* of channel definitions, each with a maximum of 256 channels. Existing operating systems would be associated with one channel subsystem (CSS) and work with a maximum of 256 CHPIDs. Different logical partitions can be associated with different CSS definitions. Thus a single operating system instance (using existing code) still has a maximum of 256 CHPIDs, but the server as a whole can have more than 256 CHPIDs. It is necessary to have multiple operating images (in multiple logical partitions) to exploit more than 256 channels, but this is a common mode of operation.

In a sense, CSS virtualizes CHPID numbers. A CHPID no longer directly corresponds to a hardware channel, and CHPID numbers may be arbitrarily assigned. A hardware channel is now identified by a PCHID, or *physical* channel identifier. A PCHID number is defined for each *potential* channel interface. An I/O adapter card has up to 16 channel interfaces[4] and 16 PCHID numbers are reserved for each I/O adapter slot in each I/O cage. Not all I/O adapters provide 16 channels, of course, but 16 PCHID numbers are allocated to each I/O slot. The PCHID numbers allocated to each I/O adapter and port on that adapter are fixed[5] and cannot be changed by the user.

The z990 and the z9 EC can use up to four channel subsystems, although the architecture permits more than four CSSs. The z9 BC and the z890 are limited to two channel subsystems.

A given logical partition is associated with a single CSS, and a single CSS has a maximum of 256 CHPIDs. Multiple logical partitions may be associated with a given CSS. Using two CSSs means that up to 512 channels can be used, and so forth. This number is reduced by *spanned* channels, and these are discussed in 3.9.1, “Spanned channels” on page 38.

Figure 3-2 illustrates the relationship between logical partitions, CSSs, CHPIDs, and PCHIDs. This is an idealized illustration and ignores such complexities as spanned channels, dynamic I/O changes, and so forth.

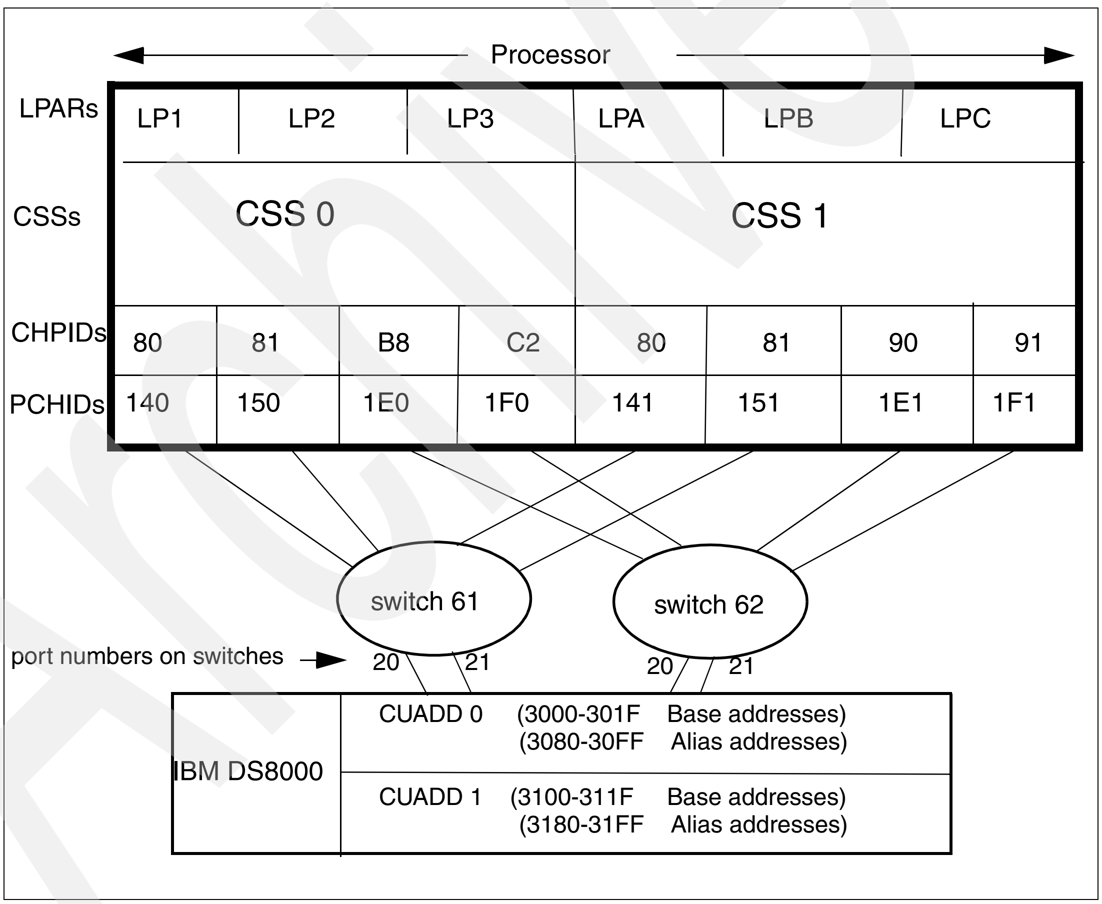

Figure 3-2 Overview of LPARs, CSSs, CHPIDs, and PCHIDs

[4] No existing I/O adapter provides 16 usable interfaces (channels). Note that an ESCON adapter has 16 interfaces, but only 15 may be used. Tthe 16th is a spare.

[5] There is a minor exception for ESCON adapters when a spare port replaces a failing port, but that is managed by the system.

Figure 3-2 on page 37 illustrates a number of important points, such as:

- Two CSSs are shown.
- Every logical partition is associated with a specific CSS. (Also note that logical partitions have unique names across the complete server. Logical partition naming and numbering has become a little more complex.)
- Multiple logical partitions (up to 15) may be associated with an CSS.
- A CHPID is associated with a specific CSS. CHPID numbers are unique within that CSS, but may be reused in other CSSs. (For example, there is a CHPID 80 in both CSSs.) A CHPID number is arbitrarily selected. For example, we could change CHPID 80 (in either or both CSSs in the illustration) to C3 simply by changing a value in the IOCDS.
- A CHPID is associated with a PCHID, and PCHID numbers are unique and fixed across the server.
- Different channels in a single I/O adapter can be used by different logical partitions. In the illustration, PCHID 0140 is the first channel on the adapter in I/O cage 1, slot 6. PCHID 0141 is the second channel on the same adapter.

3.9.1 Spanned channels

Spanning channels is the ability to configure a channel to more than one channel subsystem. The channel is transparently shared by any or all of the configured logical partitions regardless of the CSS to which the partition is configured.

Spanned channel support is an extension of the Multiple Image Facility (MIF). MIF enables sharing of channels across logical partitions, and a MIF spanned channel enables sharing across CSSs. When multiple CSSs were introduced, transparent sharing of internal channels was introduced — sharing of HiperSockets and Internal Coupling Channels (ICs) between CSSs. Support on the System z9 servers (and the z990 and z890 servers) now includes spanning of internal channels (HiperSockets and ICs), as well as spanning of external channels — FICON Express (FC and FCP), ICB-3, ICB-4, ISC-3 (CFR and CFS), and OSA-Express features. They can all now be configured as MIF spanned channels, allowing sharing of channel resources across CSSs.

Spanned channels can be shared among LPs across CSSs. A spanned channel occupies the same CHPID number in all CSSs in which it is used. For example, if a HiperSocket is to be used across two logical partitions belonging to two different CSSs, then the CHPID must be the same in both CSSs. A HiperSocket that connects logical partitions in different CSSs *must* be spanned. Table 3-1 indicates possible spanning and sharing combinations.

Table 3-1 Spanned and shared channels

Channel type		CHPID definition	MIF shared channel	MIF spanned channel
ESCON	External	CNC, CTC	Yes	No
		CVC, CBY	No	No
FICON	External	FC, FCP	Yes	Yes
		FCV	Yes	No
OSA	External	OSC,OSD,OSE,OSN	Yes	Yes
ICB-4	External	CBP	Yes	Yes
ICB-3	External	CBP	Yes	Yes
ISC-3	External	CFP	Yes	Yes

Channel type		CHPID definition	MIF shared channel	MIF spanned channel
IC	Internal	ICP	Yes	Yes
HiperSockets	Internal	IQD	Yes	Yes

3.9.2 Channel definitions in the IOCP statement

The following channel types (as defined in an IOCDS) are used in System z servers:

- FICON channel types
  - FC - Native FICON channel (both for native FICON devices and FICON CTCs)
  - FCV - FICON bridge channel (FICON Express LX only)
  - FCP - Fibre Channel Protocol (SCSI)
- ESCON channel types
  - CNC - Native ESCON channel
  - CTC - ESCON CTC channel
  - CVC - ESCON conversion channel (converter in block multiplexer mode)
  - CBY - ESCON conversion channel (converter in byte multiplexer mode)
- CF link channel types
  - CBP - Integrated Coupling Bus (ICB-3) channel, for both OS and CF partitions, to connect z900 or z800 servers
  - CBP - An ICB-4 connection to another z9 or z990/z890
  - CFP - InterSystem Coupling (ISC-3) peer mode channel, for both OS partitions and a CF partition
  - ICP - Peer mode Internal Coupling (IC-3) channel, for both OS and CF partitions (but not both at the same time), to connect among logical partitions within a z9 server internally
- OSA-Express channel types
  - OSD - (QDIO operation)
  - OSE - (non-QDIO operation)
  - OSC - (3174 emulation for consoles and VTAM® local 3270s)
  - OSN - (for emulated 3745-type functions via CCL NCP)
- HiperSockets channel type
  - IQD - HiperSockets channel, QDIO mode only

Each of these channel types requires that a CHPID be defined, even if it is an internal channel and no physical hardware (channel card) exists. Each channel, whether a *real* channel device or a virtual device (such as a HiperSocket), must be assigned a unique CHPID within the CSS. There are no default CHPID numbers, and you can arbitrarily assign whatever number you like (within the X’00’ to X’FF’ range, of course).

Most of these channel types can be shared and used concurrently among multiple logical partitions within the *same* CSS. This capability is known as the Multiple Image Facility (MIF). Exceptions are for ESCON converter channels (CVC and CBY). These channel types cannot be shared concurrently, but can be defined as reconfigurable channels by specifying the REC parameter on the channel definition. The channel can be reassigned to another logical partition after the former owning logical partition configures the channel offline.

CHPID types BY, BL, CBR/CBS, CFR/CFS, and OSA are not allowed.

3.10 Multiple subchannel sets

The *multiple subchannel sets* (MSS) functionality is new with the z9 EC and z9 BC servers and 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 hardware and software architectures. Four hexadecimal digits provides 64 K addresses, known as a *set*. IBM reserved 1024 subchannels, leaving 63 K subchannels for general use.[6] 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 Parallel Access to Volumes (PAV) has made this limitation (63 K subchannels) a problem for larger installations. A single disk drive (with PAV) often consumes at least four subchannels.[7] It was difficult to remove this constraint because the use of four hexadecimal digits for subchannels (and device numbers corresponding to subchannels) is architected in a number of places. Simply expanding the field would break too many programs.

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.

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.

An astute reader may realize that 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, 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 because subchannel set 1 is used only for alias addresses, and users (JCL, messages, documentation, programs) seldom refer directly to an alias address.

[6] The number of reserved subchannels is changed from 1024 to 256, starting with the System z9. We abbreviate this to 63 K in this discussion to easily differentiate it from the 64 K-1 subchannels available in subchannel set 1.

[7] Four appears to be a popular number for PAV. It represents the base address and three alias addresses.

Note that *each* channel subsystem (CSS) can have multiple subchannel sets, as outlined in Figure 3-3.

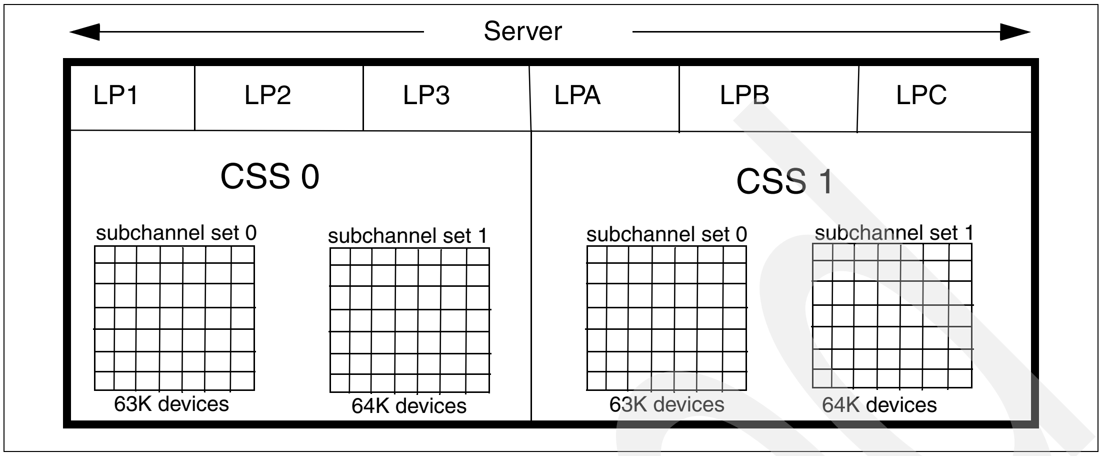

Figure 3-3 Multiple channel subsystems (CSS) and multiple subchannel sets

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.

3.10.1 IOCP example

CSS, MSS, PCHID, and logical partition information is included in an IOCDS. An IOCDS is created through the HCD utility program or by the IOCP program from IOCP file. In practical use, HCD is normally used. For illustration purposes we examine an IOCP file and assume that the reader is generally familiar with such files.

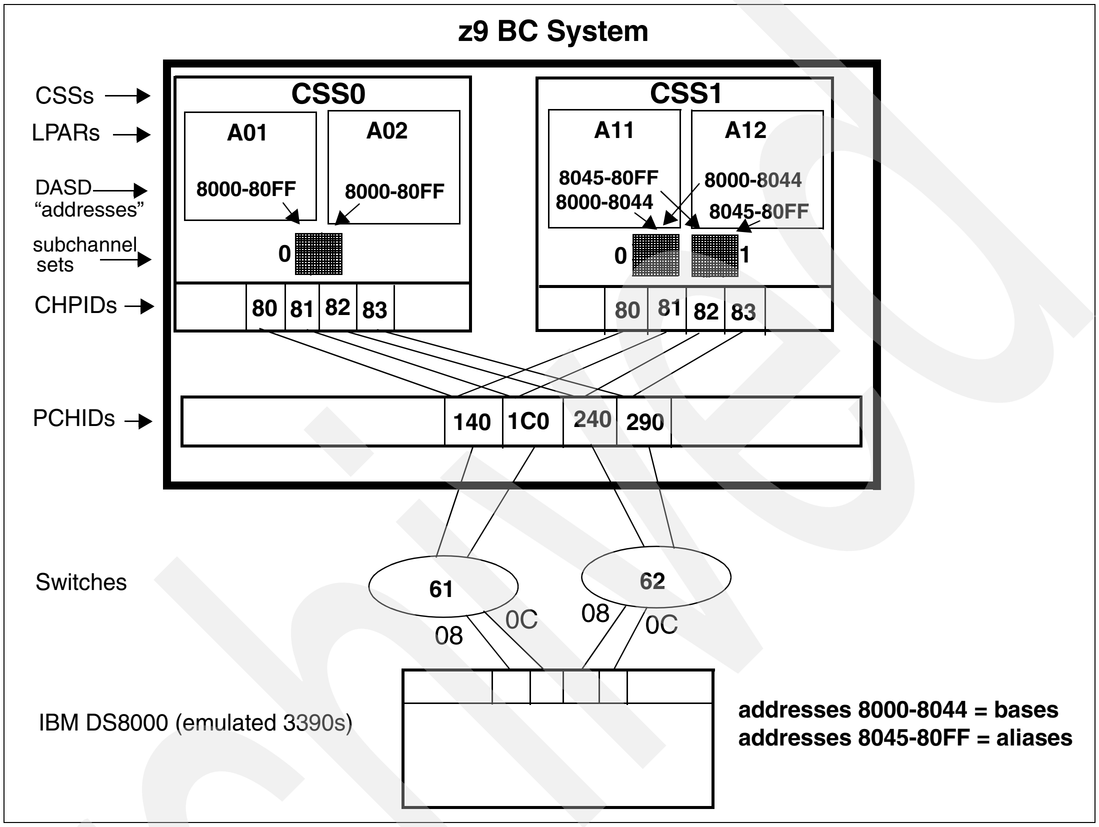

Figure 3-4 IOCP example with CSSs, MSSs, and PCHIDs

The following IOCP definition[8] describes the server shown in Figure 3-4. This is not intended to represent a practical server — it has no consoles, for example — but it illustrates the elements that are involved.

```
ID  MSG1='IODF01',MSG2='BILL.IODF01.WORK - 2006-02-23 13:59',SYSTEM=(2096,1)
    RESOURCE PARTITION=((CSS(0),(A01,1),(A02,2)),(CSS(1),(A11,1),(A12,2)))
           MAXDEV=((CSS(0),65280,0),(CSS(1),65280,65535))
    CHPID PATH=(CSS(0,1),80),SHARED,SWITCH=61,PCHID=140,TYPE=FC
    CHPID PATH=(CSS(0,1),81),SHARED,SWITCH=61,PCHID=1C0,TYPE=FC
    CHPID PATH=(CSS(0,1),82),SHARED,SWITCH=62,PCHID=240,TYPE=FC
    CHPID PATH=(CSS(0,1),83),SHARED,SWITCH=62,PCHID=290,TYPE=FC
    CNTLUNIT CUNUMBR=8000,
          PATH=((CSS(0),80,81,82,83),(CSS(1),80,81,82,83)),UNITADD=((00,256)),
          LINK=((CSS(0),08,0C,08,0C),(CSS(1),08,0C,08,0C)),CUADD=0,UNIT=2105
    IODEVICE ADDRESS=(8000,069),CUNUMBR=(8000),UNIT=3390B
    IODEVICE ADDRESS=(8045,187),CUNUMBR=(8000),SCHSET=((CSS(1),1)),UNIT=3390A
```

8 To make it more readable, this listing violates column requirements for a working IOCP and omits continuation characters.

Key elements in this IOCP include:

- Two channel subsystems (CSS0, CSS1) 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 maximum numbers of subchannels (for the relevant channel subsystem) are defined in the RESOURCE statement. CSS0 has the single, default subchannel set (with a maximum of 65280 devices that may be defined). CSS1 has two subchannel sets (with 65280 for the first and 65535 for the second as the maximum number of devices). The MAXDEV parameter is meaningful only when defining an IOCDS that can be changed dynamically.
- Four spanned channels are used (PCHIDs 140, 1C0, 240, 290). Because these are spanned (across CSSs), they must have the same CHPIDs in each CSS. They were 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 DASD system 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 uses subchannels from subchannel set 0 to access both the base and alias addresses. CSS1 uses subchannels from subchannel set 0 to access base devices, and subchannels from subchannel set 1 to access aliases. The second IODEVICE statement defaults to SCHSET 0 for CSS0. It is specified to use SCHSET 1 for CSS1.
- This example does not clearly illustrate the device address relief provided by multiple subchannel sets. However, it does exist in CSS1 in the example. Addresses 8045 through 80FF are free in subchannel set 0 and could be used for additional devices. (Addresses 8045–80FF are used for alias addresses in CSS0. The alias addresses have been moved to subchannel set 1 in CSS1.)

As can be seen from this example, the basic concepts and definitions for multiple channels subsystems are straightforward and mesh with existing IOCP parameters. Equivalent fields have been added to HCD panels for entering CSS and PCHID numbers.

3.11 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 that the CE must contact each logical partition operator and have it set the channels offline, and then after the repair, contact it 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, the 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.

3.12 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 and to reduce requirement for manual problem determination when executing an IPL for 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.

3.13 OSA for NCP

The IBM Communication Controller for Linux (CCL) on the z9 EC and z9 BC servers provides a replacement for IBM 3745 controllers working with Ethernet LANs. 3745s typically control SNA networks. A new release of CCL will be available later for the System z9 servers, and a new function of OSA-Express2 adapters will be used. Figure 3-5 illustrates the general concept.

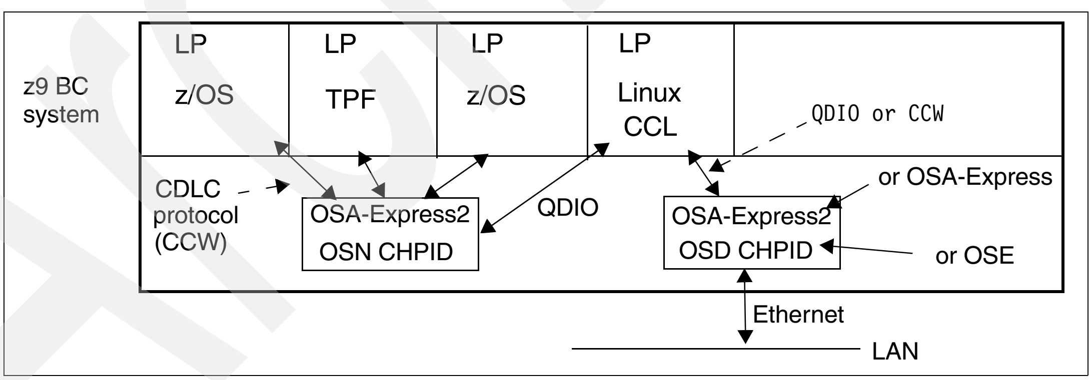

Figure 3-5 General OSA for NCP configuration with the z9 BC server

Two OSA-Express2 ports (on the same or different adapter cards) are involved. One port, operating as a CHPID type OSN, appears as an ESCON-connected 3745 interface to the associated operating systems. This OSA-Express2 adapter (or OSA-Express) has, in effect, an internal loop (logical partition-to-logical partition) that sends the data packets to the CCL product running in another logical partition under Linux. The CCL product emulates the appropriate 3745 functions, reformats the data as needed, and sends it on a LAN connected to the second OSA-Express2 port. (The figure and description are presented in terms of outgoing data. Incoming data flows in the opposite direction, of course.)

The OSA-Express2 adapter used for the OSN function can be a Gigabit Ethernet LX, Gigabit Ethernet SX, or a 1000BaseT Ethernet adapter. It *cannot* be a 10 Gigabit Ethernet adapter.

3.14 GARP VLAN Registration Protocol (GVRP)

GVRP is defined in the IEEE 802.1P standard for the control of IEEE 802.1Q VLANs. It can be used to help simplify networking administration and management of VLANs. With GVRP support, an OSA-Express2 port can register or de-register its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change (Figure 3-6).

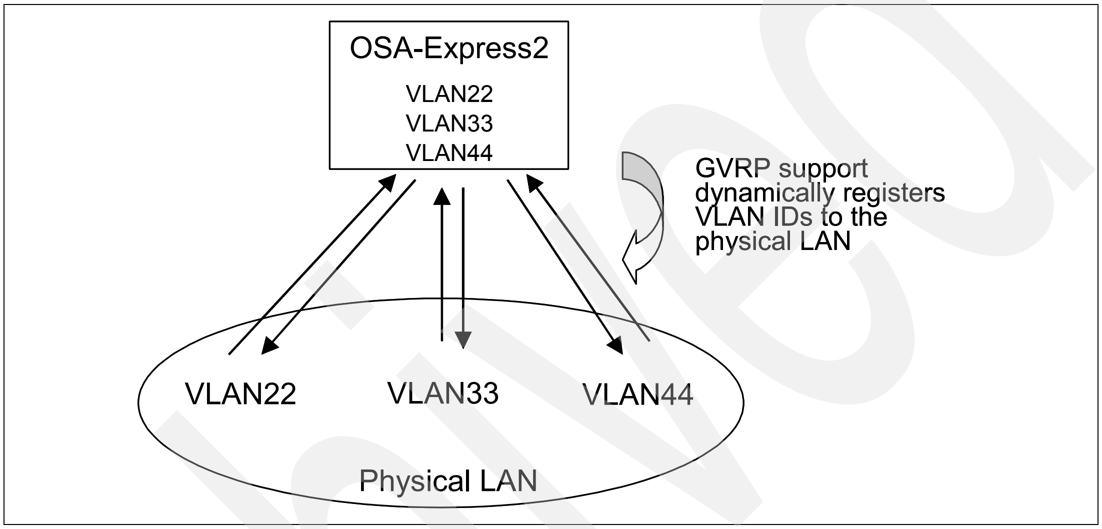

Figure 3-6 GVRP support

Support of GVRP is exclusive to System z9 and applies to all OSA-Express2 features running in QDIO mode (CHPID type OSD).

3.15 OSA Layer 2/Layer 3 Transport enhancement

This enhancement applies to z9 EC and z9 BC servers. It provides improved functionality and sharing of an OSA-Express or OSA-Express2 adapter, and applies only to QDIO operation (CHPID type OSD).

The OSA-Express adapters have provided an *IP assist* function that is used by z/OS. This assist function moves ARP processing into the OSA-Express adapter, thus offloading work from the processor. ARP processing relates to MAC addresses, and this is known as Layer 2 in TCP/IP protocols. The adapter keeps track of IP addresses being used, with their associated MAC addresses, and adds MAC headers to the Layer 3 frames (containing the target IP address) sent from z/OS TCP/IP.

The OSA-Express adapters also provide internal routing among the logical partitions sharing the adapter. Earlier versions of the OSA-Express2 adapter required the user (z/OS TCP/IP, for example) to send it only Layer 3 frames; that is, frames with the IP header but without a MAC header. Internal routing was based on the IP addresses in the frames.

Linux normally creates Layer 2 frames (with a MAC header). z/VM provides a virtual routing function that works with the MAC headers for internal routing and converts the Linux Layer 2 frames into Layer 3 frames when forwarding them to the OSA-Express adapter.

Prior to this enhancement, a given OSA adapter could not handle a mixed environment in which it received frames with and without the Layer 2 (MAC) header. The enhancement available with this announcement allows the OSA-Express adapter to handle both types of frames concurrently. This permits better sharing of adapters and also permits internal routing of traffic when appropriate. The z/VM routing function is no longer required for local (within the server) routing by Linux. This is roughly illustrated in Figure 3-7.

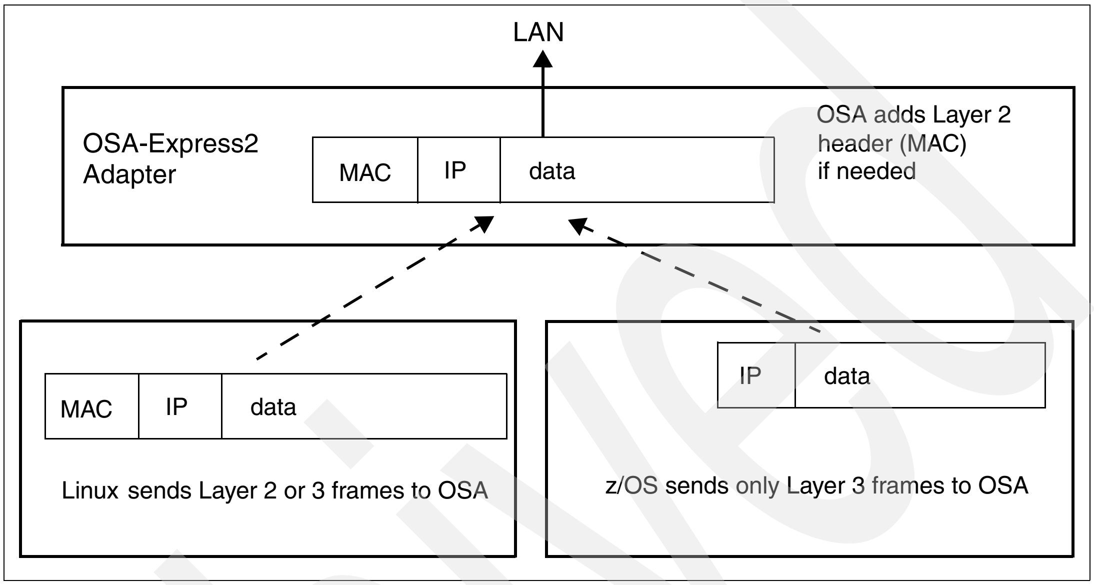

Figure 3-7 OSA-Express adapter and Layer 2/3 with enhancement

3.16 OSA-Express2 Link Aggregation for z/VM

z/VM Virtual Switch-controlled (VSWITCH-controlled) link aggregation (IEEE 802.3ad) allows the user to dedicate 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 the user to combine 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 contain 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 (for example, Gigabit Ethernet to Gigabit Ethernet).

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.

3.17 OSA Layer 3 Virtual MAC for z/OS and z/OS.e

To help simplify the infrastructure and to facilitate load balancing when a logical partition is sharing the same OSA Media Access Control (MAC) address with another logical partition, each operating system instance can now 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 the PRIROUTER/SECROUTER function in z/OS.

OSA Layer 3 VMAC is exclusive to System z9, is applicable to the OSA-Express2 and OSA-Express features when configured as CHPID type OSD (QDIO), and is supported by z/OS and z/OS.e.

3.18 OSA-Express2 QDIO Diagnostic Synchronization

QDIO Diagnostic Synchronization adds new function designed to provide system programmers and network administrators with the ability to coordinate and simultaneously capture both software and hardware traces. It allows z/OS to signal an OSA-Express2 feature (using a new 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.

3.19 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 introduction of the OSA-Express Network Traffic Analyzer and QDIO Diagnostic Synchronization on the System z9, customers will 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.

3.20 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.

3.21 MIDAW facility

The Modified Indirect Data Address Word (MIDAW) facility first introduced with the System z9 server is a new system architecture and software exploitation designed to improve FICON performance. This facility is available only on the System z9 servers and is exploited by z/OS access methods.

- 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.

  It reduces FICON channel and control unit overhead.

The MIDAW facility provides a more efficient CCW/IDAW structure for certain categories of data chaining I/O operations. MIDAW is short for Modified IDAW. An Indirect Address Word (IDAW) is used to specify data addresses for I/O operations in a virtual environment.[9] The existing IDAW design allows the first IDAW in a list to point to any address in 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 2 K or 4 K units of data. This limits the usability of IDAWs to straightforward buffering of a sequential record.

[9] There are exceptions to this statement; we skip some details in the following description. We assume that knowledgeable readers will merge this brief description with their understanding of I/O operations in a virtual memory environment.

Figure 3-8 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 in real storage.

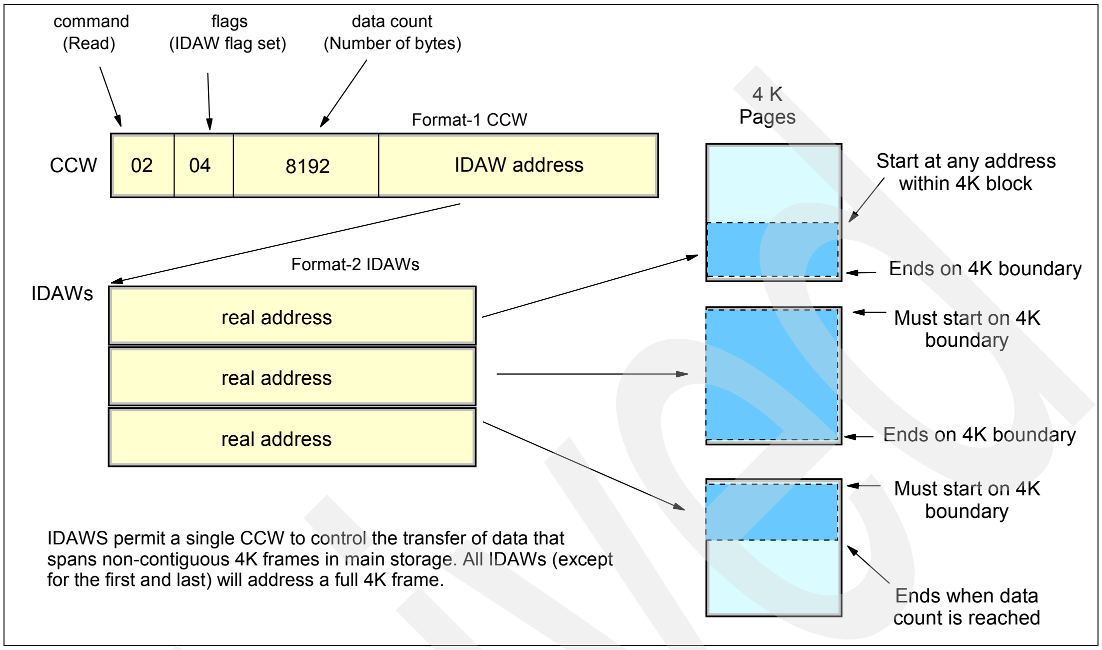

Figure 3-8 IDAW usage

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, the ORB specifies format-2 IDAWs with 4 K byte blocks, the CCW count field specifies 8 K bytes, and the first IDAW designates a location in the middle of a 4 K 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, data chaining CCWs is inefficient in modern I/O environments for a number of reasons involving switch fabrics, control unit processing and exchanges, numbers of channel frames required, and so forth.

The MIDAW facility is a new method of gathering/scattering data into/from discontinuous storage locations during an I/O operation. Figure 3-9 shows the modified IDAW (MIDAW) format. It is 16 bytes long and is aligned on a quadword.

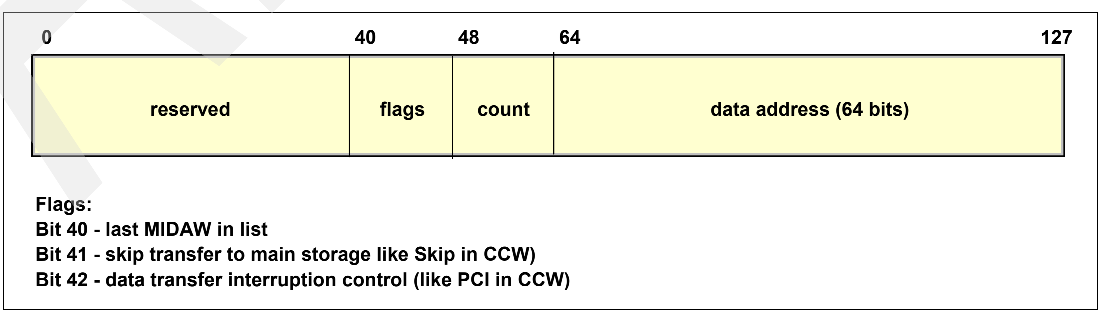

Figure 3-9 MIDAW format

Figure 3-10 shows an example of MIDAW usage.

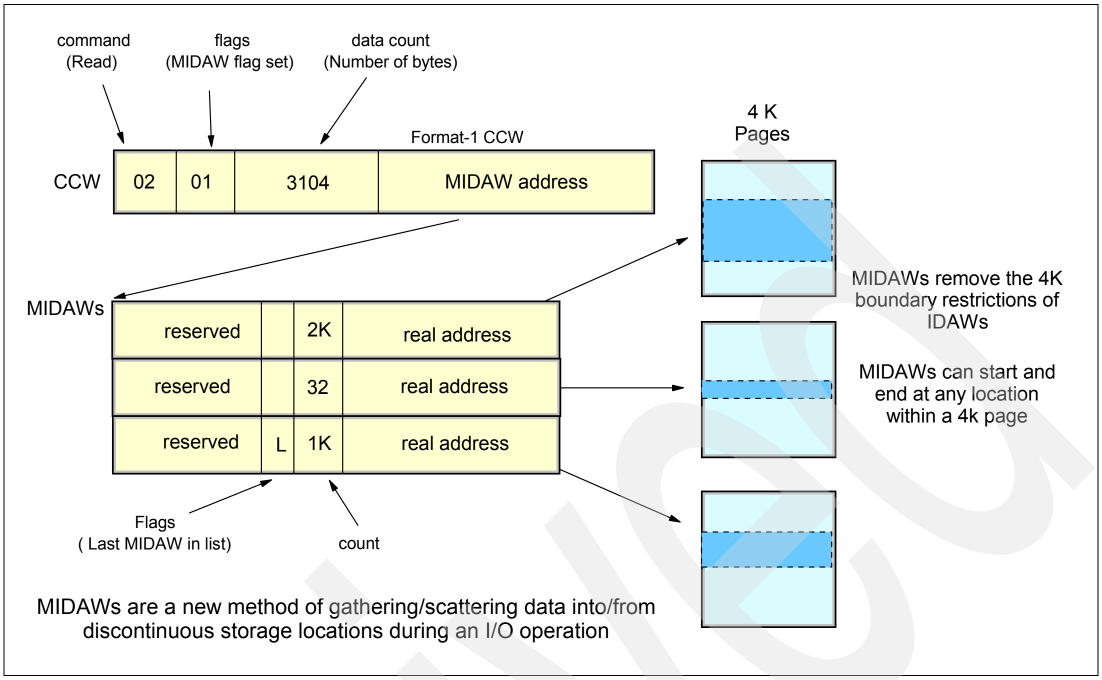

Figure 3-10 MIDAW usage

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 MIDAWs 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 that the largest possible count is 4 K. The maximum data count of all 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 buffers separate from those used for larger data areas in 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.

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. Extended format data sets were introduced in 1993 and are used widely today.

Extended format data sets

Both VSAM and non-VSAM (DSORG=PS) can be 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 on DASD. VSAM appends the suffix to the end of every Control Interval (CI), which normally corresponds to a physical record. (A 32 K CI is split into two records in order to optimize disk space usage.) The suffix is used to improve data reliability and facilitate various other functions. 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.

Besides reliability, EF data sets enable three new functions: DFSMS™ striping, access method compression, and Extended Addressability (EA). DFSMS EA is especially useful for creating large DB2 partitions (> 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.

MIDAWs are used by two IOS drivers:

- The media manager automatically exploits MIDAWs when appropriate.
- Users of the EXCPVR IOS driver may construct channel programs containing MIDAWs, provided they construct an IOBE with the new IOBEMIDA bit set.

Users of EXCP *may not* construct channel programs containing MIDAWs.

3.22 N_Port ID virtualization

FICON Express ports may be used for FCP operation instead of FICON operation. This is done by assigning the CHPID type FCP for the channel. FCP operation is used primarily by Linux. z/VM fully supports FCP for guest usage, and fully for z/VM FCP disk usage.[10] The Storage Area Network (SAN) regards the FCP channel as a single identity. In SAN terminology, it has a single World Wide Port Number (WWPN).

If the FCP channel is used by multiple independent operating systems (under z/VM or in multiple logical partitions), the SAN is unaware of this and cannot distinguish among the various independent users. This has obvious implications for security and file system integrity controls. The problem can be partly avoided by requiring separate FCP channels for each logical partition, but this does not help with z/VM.

A major improvement is provided by N_Port ID virtualization. This allows a single physical FCP channel to be assigned multiple WWPNs and appear as multiple channels to the external SAN environment. In general, a separate WWPN would be provided to each logical partition (not running z/VM) and to each guest in a z/VM logical partition (and to z/VM itself). This permits basic security controls, such as LUN masking and switch-level zoning, to be used by each independent operating system working through the FCP channel. This functionality may be used to permit shared access to the same SCSI LUN in the SAN, although additional considerations are involved in sharing LUNs.

The System z9 server N_Port ID virtualization is based on an extension to the Fibre Channel standards that allows a host bus adapter to perform multiple logins to the SAN fabric via a single physical port. The entry point to the SAN (that is, the switch to which the FCP channel is directly connected) must support *multiple N-port ID logins.* No changes are needed in any *downstream* switches, bridges, control units, or devices, although NPIV support is needed in the switch that is attached directly to the z9 server. The switch (and not the FCP channel itself) provides the multiple WWPNs used for virtualization. There are limits on the number of WWPNs that can be provided for a given FCP.

In principle, N_Port ID virtualization can be transparent to software. In practice, new z/VM commands are being provided to query the WWPN names. In addition, software changes for the SCSI driver for Linux will exploit the ability of a switch to provide multiple WWPNs.

[10] It is not supported by z/OS, which does not use FCP channels.

3.23 FCP performance enhancement

There are new enhancements designed to support more start I/Os. These enhancements are exclusive to System z9 and are available for FICON Express4 and FICON Express2 channels configured to operate in FCP mode (CHPID type FCP). Both z/VM and Linux for System z FCP environments will benefit. No service in operating system is required.

3.24 FCP performance metrics for Linux

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 will help 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.

3.25 New and changed instructions

The z9 EC server includes a number of new and changed instructions. These instructions may be grouped into several categories:

- Hexadecimal floating point instructions for various unnormalized multiply instructions and multiply and add instructions.
- Several *immediate* instructions, including various add, compare, exclusive OR, OR, subtract, load, and *insert* formats. Use of these instructions, where appropriate, improves performance and can reduce the need for a data base register.
- Various load instructions, especially those useful for handling unsigned half words (such as used for unicode).
- The cryptographic instructions (KM, KMC, KIMD, KLMD) have been extended with AES, SHA-256, and random number functions.
- The *Store facility list extended* instruction can reflect use of more than 32 facilities.
- The *Extended Translate Facility-2* instructions (TROO, TROT, TRTO, TRTT) have been improved, allowing them to bypass test-character processing when appropriate and relaxing operand-alignment requirements.
- The *Extended Translate Facility-3* instructions (CU24, CU21, CU12, CU14) have been enhanced to conform with the current Unicode 4.0 standard.

The z9 BC server extends this list with three new facilities:[11]

- The Compare and Swap and Store Facility

  This provides a new instruction that is similar to the Perform Lock Operation (PLO) instruction but does not require a separate lock. It provides an atomic compare and swap operation and stores a footprint at a separate location as part of the atomic operation. The purpose is to reduce system overhead for certain common lock management routines.
- The Extract CPU Time Facility

  This provides a new instruction that may be used to obtain a z/OS task's elapsed CPU time. Prior to this, a service routine (IETTUSD) was called for this function. The service routine had substantial overhead. It needed to disable interrupts, acquire a lock, compute

[11] The term *facility*, in this context, means one or more instructions that are not a required part of the base architecture and that might not be installed. Software can sense which facilities are present.

the time used, release the lock, enable interrupts, and return to the caller. (Problem state callers also had Program Call and Program Return overhead.) An increasing number of z/OS functions track elapsed CPU time. Moreover, the precision of the tracking has been limited by the overhead imposed by this routine. The new instruction can be used to extract elapsed CPU time in a single operation.

- The Move With Optional Specifications Facility

  This provides an instruction with variations on the existing *move long* instructions. The new facility provides a variety of specifications for controlling the move.

Figure 3-11 shows the instruction formats.

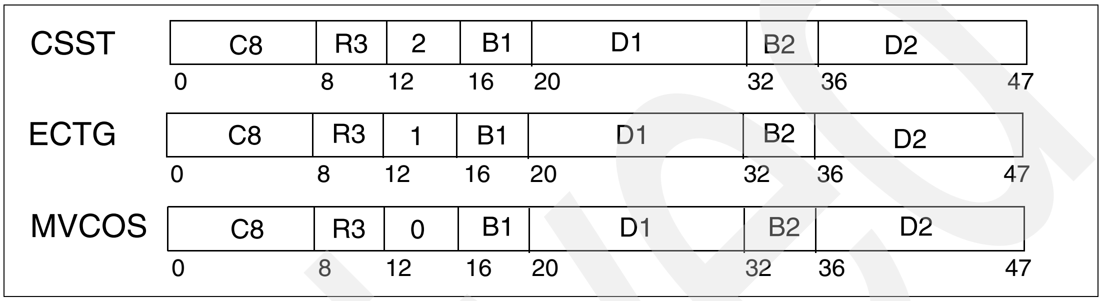

Figure 3-11 New instructions

Compare and swap and store (CSST)

With this instruction the first operand is compared with the third operand. If they are equal, the *replacement value* is stored in the first-operand location and the *store value* is placed in the second-operand location. If they are not equal, the first operand is placed in the third operand location and the second operand is unchanged. The condition code indicates whether the first and third operands were equal.

When executing the instruction, general register 0 contains a *store characteristic* and *function code* that determines the lengths of the various operands (1, 2, 4, or 8 bytes). General register 1 points to a parameter list containing the *replacement value* and the *store value*. The contents of general register 1 are used according to the current addressing mode (24, 31, or 64 bits) and must be on a quadword boundary.

This instruction causes processor serialization. It may be used in problem state. However, this instruction should be used only when an interruption between the compare-and-swap operation and the following store operation cannot be tolerated.

Extract CPU time (ECTG)

The value of the current CPU timer is subtracted from the first operand and the difference is placed in general register 0. The second operand is placed (unchanged) in general register 1. The 8 bytes at the third operand location replace the contents of general register R3. The operands are treated as unsigned 64-bit integers.

The intended use of this instruction involves standard z/OS control blocks. These are:

- PSADTSV, which contains the CPU timer value at last dispatch. (That is, when the current task was dispatched.) This is the intended first operand.
- TCBTTIME, which contains this task’s accumulated CPU time. This is the intended second operand.

- PSAPTYPE, which is a processor type indicator, can be used as the third operand. The intention is that this value could be used to scale or modify the results in some way.

Remember that the CPU timer is a signed binary counter (in the same format as the TOD clock except for the sign) that is *decremented* during normal system operation. Existing instructions to set or store the CPU timer are privileged. However, the ECTG instruction can be used in problem state assuming that the operands can be accessed in problem state. In practice the user may need to alter the address space mode to accomplish this. The contents of R3 are treated as a 24-bit, 31-bit, or 64-bit address according to the current address mode.

Move with optional specifications (MVCOS)

The first operand is replaced by the second operand. Bits in general register 0 determine the address-space-control modes and protection keys that are used to access the first and second operands. The R3 operand specifies the length and is used as a 31-bit unsigned integer when in 24-bit or 31-bit addressing modes and as a 64-bit unsigned integer when in 64-bit addressing mode.

This privileged instruction is subject to special-operation (semiprivileged) exceptions. General register 0 has controls for each of the operands. For *each operand* the following controls exist:

- Specified Access Key Validity Bit. This bit indicates whether the PSW key or a different key (in GR0) is to be used to access the operand.
- Specified Access Key. This is the alternate storage key used to access the operand if the PSW key is not to be used.
- Specified Address Space Control Validity Bit. This bit indicates whether the address space control bits in the current PSW or different address space controls (in GR0) are to be used to access the operand.
- Specified Address Space Control. These are the alternate address space control bits used to access the operand if the PSW control bits are not to be used.

3.26 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 Standard 754-1985.

3.27 Branch tracing

Debugging a *wild branch* has always been a difficult task. The System z9 servers have a register that contains the address of the last successful branch instruction. If the branch results in an ABEND (or SLIP trap), then the program issuing the branch can be immediately identified. Starting with z/OS 1.7, this *branch from* information is included in SDWA and RTM2WA dump areas.

3.28 Performance factors

Several factors are responsible for the performance characteristics of the System z9 server processor units. The most obvious factor is that the PU has been designed and optimized to meet the demands of new workloads, while at the same time maintaining excellent performance on traditional S/390 workloads.

Each PU utilizes a superscalar design that provides all of the following features:

- Decoding of two instructions per cycle
- Execution of three instructions per cycle (oldest must be a branch)
- In-order execution
- Out-of-order memory access

The z9 BC book has up to sixteen 2.7 GBps STI links that form the base for high-bandwidth I/O operations.

Other processor design features that contribute to the enhanced performance of System z9 processors are:

- Floating-point performance is optimized for IEEE arithmetic (exploited by Java applications). Floating-point operations have now been fully pipelined at five stages for both IEEE (binary) and IBM floating point (hex) arithmetic. Additional optimization for BFP is present.
- A programmable Dynamic Address Translation (DAT) translator is used. This mechanism allows for later updates to the address translation algorithm (mapping virtual addresses to physical memory locations). A high-performance DAT function is complex, and this design allows for corrections and design changes to the current algorithm, if necessary, as well as making the translator extendable for future architectural changes.
- A Second Level Translation Lookaside Buffer (TLB) is used. This mechanism provides a secondary cache for the DAT for both the instruction and data caches. The secondary buffers provide an additional 4 K of dynamic address mapping over the current 512 entries available.
- Several cryptographic instructions have been added (and extended with the z9 BC). These instructions for symmetric encryption/decryption are used to speed up SSL transactions, VPN data transfer, and data repository applications.

3.29 Program-directed re-IPL

A new function, program-directed re-IPL, enables an operating system to re-IPL itself. This function is supported for standard IPLs (based on CCWs via ESCON or FICON channels) and for SCSI IPLs (via FCP channels). A new SCLP call is provided with the System z9 servers to cause a re-IPL of an logical partition using the same IPL parameters as were used for the last manual IPL. This function could be used by any operating system, but is intended for Linux usage.

3.30 Open exchanges for FCP and FICON

An *open exchange* is part of FICON (and FCP) terminology. An open exchange represents an I/O operation in progress over the channel. Many I/O operations can be in progress over these channels at any one time. For example, a disk I/O operation might temporarily disconnect from the channel while performing a seek operation or while waiting for a disk rotation. During this disconnect time, other I/O operations can be managed. In addition, these channels can multiplex data transfer for several devices at the same time. zSeries servers could sustain up to 32 such open exchanges per FICON (or FCP channel). The System z9 servers have increased this number to 64.

This larger number contributes to system performance. If the open exchange limit is reached, new I/O operations are refused by the channel. This results in necessary queuing and retries by the operating system.

3.31 CHPID mapping

The CHPID Mapping Tool is available on the IBM Resource Link. System z9 servers do not have CHPIDs assigned to I/O ports as part of the initial IBM setup process. CHPID numbers may be assigned arbitrarily to I/O ports. The I/O ports are identified by fixed PCHID numbers. In trivial cases, this assignment is easily done manually. In most practical cases, it can be quite complex, and the use of the CHPID Mapping Tool is strongly recommended.

The CHPID Mapping Tool and an example of its use during migration from earlier servers is well described in *System z9 and zSeries CHPID Mapping Tool User's Guide* available at IBM Resource Link:

`http://www.ibm.com/servers/resourcelink`

The example is detailed, and we have not copied the material into this document. We suggest that the material in the CMT document is required reading for anyone undertaking a migration from an earlier server to a System z9 server.

In its normal usage, the CHPID Mapping Tool uses the IBM-provided configuration file for *your* z9 BC server (before it is delivered) and your IOCP file from your current (older) server. The tool may require several iterations (with manual input) to resolve assignment conflicts and produce the best configuration for availability and performance. The result is a new IOCP file.

3.32 More logical partitions

The z9 BC server can have up to 15 or 30 logical partitions, depending on the model. The key element here is that the previous *architectural limit* of 15 (which used a 4-bit field in various LIC and software control blocks) has been extended. The zArchitecture functions use a 1-byte field. A few additional changes might be important:

- A System z9 server runs only in LPAR mode. Basic mode no longer exists.
- A maximum of 15 logical partitions may be associated with a single channel subsystem.
- Partition identifiers, names, and numbers are a little more complex than with older servers, as explained in this section.

Several identifiers are associated with an logical partition:

- Logical partition identifier: A number in the range 0–X’3F’ that is assigned by the user when defining LPAR image profiles through the SE or HMC. It is unique across the System z9 server. This identifier is returned by an STIDP or STSI instruction. This identifier is also known as the user logical partition ID (UPID).
- MIF image ID[12]: Defined through HCD or IOCP, this is the partition number defined in the RESOURCE statement in an IOCP source file. It is in the range 1 to X’F’ and is unique

within an CSS. It does not have to be unique within a System z9 server. This ID is also known as the IID.

- Logical partition name: Defined through HCD or an IOCP, this is the partition name in the RESOURCE statement in an IOCP source file. A logical partition name must be unique across the System z9 server.

The logical partition identifier is used when defining activation profiles. The HCD field for the MIF Image has the label *Partition Number* and is equivalent to the MIF image ID used in IOCP RESOURCE statements. We suggest that you establish a naming convention for the logical partition identifier. For example, you could use the CSS number concatenated to the MIF ID. Logical partition ID 1A could mean CSS 1, MIF ID A. This fits within the allowed range of logical partition IDs and conveys useful information to the user.

3.33 LPAR group capacity limit

The new 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 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 later.

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.

3.34 Hardware Management Console and support element

Although not directly related to hardware itself, we would like to show here some features in the HMC Application that were not available before System z9.

The Hardware Management Console (HMC) and support element (SE) were reorganized and have a new design on System z9 servers. OS/2 is no longer used as a base operating system, but the HMC application has the same functionality as its predecessors. Older HMCs cannot be used with z9 systems.

3.34.1 HMC features

The HMC application can be accessed remotely from any standard Web browser. The major difference with the previous HMC Web interface application is that by default, most HMC functions are available to the user with the same *look and feel* characteristics of a local HMC desktop.

The user interface has two user-selectable styles: the *traditional* HMC style and the *tree Web-based* user style.

The traditional HMC style has the same format as the current HMC application, and the look and feel is similar, but there are minor differences. For example, the left mouse button is now used for drag-and-drop operations rather than the right mouse button. The tree Web-based

[12] Multiple Image Facility (MIF) was formerly known as EMIF for ESCON Multiple Image Facility because it was introduced for ESCON channels. The function is more general now, so the E portion of the name was dropped.

user style has a new design and is intended to be a standard GUI style across all IBM System platforms.

Security has been enhanced as expected for a critical Web-based application such as the HMC application. Any communication with the HMC uses certificate-based 128-bit high-grade SSL encryption. The HMC provides a self-signed certificate by default, with the option to use a customer-provided certificate instead. Additionally, a dynamic and integrated firewall controls and limits all communications and connectivity to the HMC application.

The Remote Support Facility can be done either using a modem (asynchronous communication is still supported) or through the LAN (more precisely, the Internet).

The new HMC level may be used to control System z servers. Support element updates (MCLs) may be needed on these earlier servers. Systems prior to G5 cannot be accessed by the new HMC application. The driver levels required on these systems (if they are to be accessed through the new HCM level) are as follows:

Product	Driver
System z	driver 63
zSeries z990, z890	driver 55
zSeries z900, z800	driver 3G

The new HMC Application can be installed on earlier HMC units with 1 GB memory and a 40 GB hard disk (feature codes 0075 or 0076). These units can use token-ring adapters they may already have installed. Older HMCs (earlier drivers) cannot be used to operate z9 BC systems.

3.34.2 Security

Because the HMC might connect to IBM facilities over public networks, security can be a concern. Connections may be via dial-up telephone or through an Internet link. In either case, the following safeguards apply:

- All data transferred to IBM is for service functions only. No customer data is transferred.
- All data transferred is encrypted with SSL techniques.
- The data exchange protocols are proprietary to IBM service and are not published.
- The HMC is a closed platform.
- The IBM system verifies that the connecting system is known and authorized.
- Connections are always outbound only and are initiated by the HMC. IBM never initiates a call into an HMC.

For dial-up connections, the following safeguards are present:

- The HMC will not answer an incoming call to the modem unless changed by the customer to permit such connections.
- The link protocol is PPP (with TCP/IP formats).

Internet connections are assumed to go through a customer firewall before connecting to the external Internet. The HMC has an integrated firewall as well. This enables only the specific TCP/IP ports needed by the HMC functions.

HMC Lightweight Directory Access Protocol (LDAP) Authentication

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.

Remote support facility (RSF)

There is a new 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 Network Address Translation (NAT) 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 now 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 via a phone (modem) based RSP connection or a network-based (Internet) connection.

3.34.3 HMC application enhancements

In this section we discuss HMC application enhancements.

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.

It is supported in z/VM V5R3 and later.

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 via 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.

4

Software support

This chapter describes the operating system requirements and support considerations for the z9 BC server and its new features. It covers z/OS, z/VM, z/VSE, VSE/ESA™, z/TPF, TPF, and Linux on System z. This information is subject to change. Therefore, for more current information refer to the Preventive Service Planning (PSP) bucket (2096DEVICE).

4.1 Operating system support

Table 4-1 summarizes the minimum operating system levels required by the z9 BC system.

Table 4-1 z9 BC minimum operating systems requirements

Operating systems	**ESA/390 (31-bit Mode)**	**z/Architecture (64-bit mode)**	
z/OS V1R6 and later	No	Yes	
z/VM V5R1 and later	No	Yes	
Linux for System z	Yes[a]	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 2096DEVICE Preventative Planning (PSP) bucket prior to installing an IBM System z9 BC.

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 2096DEVICE - For z9 BC.

When running in 64-bit mode (z/Architecture), Expanded Storage (ES) is supported by z/VM, but not by z/OS. However, if z/OS is used in 31-bit mode (ESA/390), then ES may be used.

The remainder of this chapter examines specific features of the z9 BC server and indicates the appropriate operating system levels needed for that feature. If an operating system is not listed for a feature or the information is blank, it means that we were unable to determine the status at the time of writing. In all cases, various PTFs (and appropriate MCL levels) may be needed with the operating system level indicated. Always check PSP buckets.

4.2 Support summary

The following summary indicates the *lowest* levels of various supported (at the time of writing) operating systems that support the indicated functions. These lower levels may require PTFs, SPEs, or other additional software to provide the function.

Table 4-2 Operating system support summary

Function	z/OS	z/VM	z/VSE	z/TPF TPF	Linux on System z
Base z9 BC support	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
30 logical partitions	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
63.75K subchannels	1.6	5.1			SLES 9 RHEL 4
Multiple subchannel sets	1.7				TBD[a]
MIDAW facility	1.6	5.3 (for z/OS guests)			
N_Port ID virtualization		5.1	3.1		SLES9 (SP3)
FICON Link incident report	1.7				
Program directed re-IPL		5.3			SLES 9 (SP3)
System-initiated CHPID reconfiguration					
Multipath IPL	1.6				
OSA Express2 1000BaseT (*CHPID OSC*)	1.6	5.1	3.1		
OSA Express2 1000BaseT (*CHPID OSD*)	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
OSA Express2 1000BaseT (*CHPID OSE*)	1.6	5.1	3.1		
OSA Express2 OSN	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
OSA Express2 GVRP	1.7	5.1	N/A	NA	TBD[b]
OSA Express2 GbE (*CHPID OSD*)	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
OSA Express2 10 GbE (*CHPID OSD*)	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
FICON Express4 (CHPID FC)	1.6	5.1	3.1	1.1 4.1	SLES 9 RHEL 4
OSA-Express2 Link Aggregation for z/VM		5.3			

Function	z/OS	z/VM	z/VSE	z/TPF TPF	Linux on System z
OSA Layer3 VMAC support	1.8	5.1[b]			
OSA-Express2 Network Traffic Analyzer	1.8	5.1[b]			
QDIO Diagnostic Synchronization	1.8	5.1[b]			
OSA Dynamic LAN Idle	1.8	5.1[b]			
FICON Express4 (CHPID FCP)		5.1	3.1		SLES 9 RHEL 4
Hipersockets IPv6	1.7	5.2			
CryptoExpress2 compatibility	1.6	5.1			SLES 9
CryptoExpress2 exploitation	1.6[c]	5.1	3.1		TBD[b]
CPACF	1.6	5.1	3.1		SLES 9 RHEL 4
CPACF Enhancements	1.6[c]	5.1			SLES 9 (SP3) RHEL 4 (U3)
Enhanced performance assist for z/VM guests		5.2			TBD[a]
OSA/SF Enhancements for IP and MAC addressing	1.6	5.2			
Server Time Protocol	1.7				
LPAR group capacity limit	1.8				

a. Statement of Direction.
b. Guest support.
c. With Cryptographic Support for z/OS V1.6/V1.7 and z/OS.e V1.6/V1.7Web deliverable.

Table 4-2 on page 63 concentrates on adapters that would normally be found in a z9 BC server. In a more general sense, the line entries for OSA-Express2 also apply to OSA-Express adapters, and the line entries for FICON Express4 also apply to FICON Express2 and FICON Express adapters.

For the most current operating system maintenance information go to the following Web sites:

- z/OS

  `http://www.ibm.com/servers/eserver/support/zseries/zos/`

- z/VM

  `http://www.ibm.com/servers/eserver/support/zseries/zvm/planning.html`

- z/TPF

  `http://www.ibm.com/software/htp/tpf/pages/maint.htm`

- z/VSE

  http://www.ibm.com/servers/eserver/zseries/zvse/support/preventive.html

- Linux on System z9

  http://www.ibm.com/servers/eserver/zseries/os/linux/support.html

5

General considerations

This chapter discusses several areas that should be considered when moving from an earlier zSeries server (z900, z800, z990, z890) to a z9 EC or z9 BC system. This includes:

- Hardware Management Console and support element
- Cabling services and cable planning
- I/O channel considerations
- I/O configuration definition and management
- Capacity upgrades
- Logical planning
- Coupling Facility planning
- Operating system migration

5.1 Hardware Management Console and support element

The Hardware Management Console (HMC) that is delivered with the z9 EC and z9 BC servers includes a completely new implementation of the HMC application code. It is no longer based on OS/2 and is now a *closed system* in which the underlying operating system is not accessible by the user. Additional applications cannot be installed. The HMC application itself is the only possible application. Older HMCs cannot be used with z9 systems.

Functionally, the HMC has not changed significantly. It is still the hardware control point with the same functions provided by earlier HMC implementations. It includes functions introduced in recent releases, such as the integrated 3270 and ASCII consoles. There are now two user interface options:

- A traditional interface that is generally similar to what is seen on previous HMC consoles
- A new *tree interface* that is oriented to more hierarchical use

The appearance of the two interfaces is quite different, but the resulting functionality is the same. Both interfaces may be used at the same time when multiple HMCs are involved.

The mechanism for viewing and using the interfaces has changed. There is no direct access to the interfaces on the HMC itself. A browser must be used. The appearance, functionality, and performance is the same whether the browser runs on the HMC or another PC. (This statement assumes reasonable LAN performance between the HMC and PC.) The HMC implementation includes an HTTP *server* as its user interface. The user connects to this interface with a browser.

This is quite different from earlier HMC functions that allowed browser connections. These were either very limited in functionality (when using a normal browser interface) or involved a complete remote desktop image (using Desktop on Call or a similar interface). The remote desktop image required much more intensive LAN usage and often had somewhat sluggish performance. The full browser interface of the new implementation requires only *normal* LAN performance.

At least one HMC is needed to run a z9 BC system. Several HMC operators, at different locations, can all use the same HMC simply by accessing it through a browser. Multiple HMCs might be desired for redundancy.

The new HMCs provided with z9 EC and z9 BC systems have only Ethernet LAN connections. Token-ring adapters are not available. However, older HMCs that are upgraded to run the new HMC code may have token-ring adapters, and these are still supported.

Mixed HMC levels

The new HMC level may be used to control System z servers. Support element updates (MCLs) may be needed on these earlier servers. Systems prior to G5 cannot be accessed by the new HMC application. The driver levels required on these systems (if they are to be accessed through the new HCM level) are as follows:

Product	Driver
System z	driver 63
zSeries z990, z890	driver 55
zSeries z900, z800	driver 3G
G5 or G6	driver 26

The new HMC Application can be installed on earlier HMC units with 1 GB memory and a 40 GB hard disk (feature codes 0075 or 0076). These units can use token-ring adapters they may already have installed. Older HMCs (earlier drivers) cannot be used to operate z9 BC systems.

5.1.1 Support elements

Every z9 BC system has two identical support element (SEs). These are IBM ThinkPads located under the system covers in the Z frame. At any given time, one SE will be operating as the primary SE while the other operates as the alternate SE to provide redundancy. The SE is intended to be used only by IBM Service Representatives. However, it may be used to manage the system in the event of an HMC failure.

Support elements on z9 BC systems have 1 GB of memory, a 60 GB hard disk, no diskette drive (internal or external), and a CD-ROM or DVD-ROM drive. Two Ethernet adapters are provided. Token-ring is not available.

5.1.2 HMC and SE connectivity

Configuring and ordering SEs and HMCs requires an understanding of the LAN interfaces that are involved. Both units normally have two LAN interfaces. The general rules are:

- An SE is connected to a given HMC by one or two LANs. In simple situations, the second LAN interfaces included with SEs and HMCs are not used. If a second LAN is connected it will be used for automatic failover.
- The two LAN interfaces on an SE may be used to connect to one or more HMCs, using two independent LANs.
- Each SE on a z9 BC system has two Ethernet adapters that may be used for HMC connectivity.
- A default HMC configuration has two Ethernet interfaces. However, token-ring interfaces can exist if the current HMC hardware was upgraded to the new HMC driver level and already had a token-ring card installed.
- The HMC desktop must have at least one Ethernet interface. There is no other way to communicate with the z9 BC support element.

A common arrangement uses at least two HMCs: one near the system and the other near the general operations control area. With the new HMC design this might no longer be necessary because a browser (based in an inexpensive PC) can be used at one location to provide a second interface to the HMC in the other location.

You can purchase any reasonable number of HMCs to use with the system.[1] You have the following elements for planning your HMC and SE installation on your system:

- Four Ethernet cables, each 50 feet (about 15m) long, are provided with the z9 BC system. Potentially, these connect to the two Ethernet adapters in each SE. In practice, only one of the Ethernet connections in each SE is normally used.
- A new HMC with two Ethernet features. An existing HMC might have one Ethernet and one token-ring adapter.
- A display for the HMC.
- An Ethernet switch. This is ordered automatically.
- A modem interface (use is optional).

1 Up to 32 HMCs can be used to control the z9 BC server.

The SE and HMC Ethernet ports can run at 10 Mbps or 100 Mbps and auto-sense the LAN speed. Figure 5-1 illustrates a basic SE/HMC configuration.

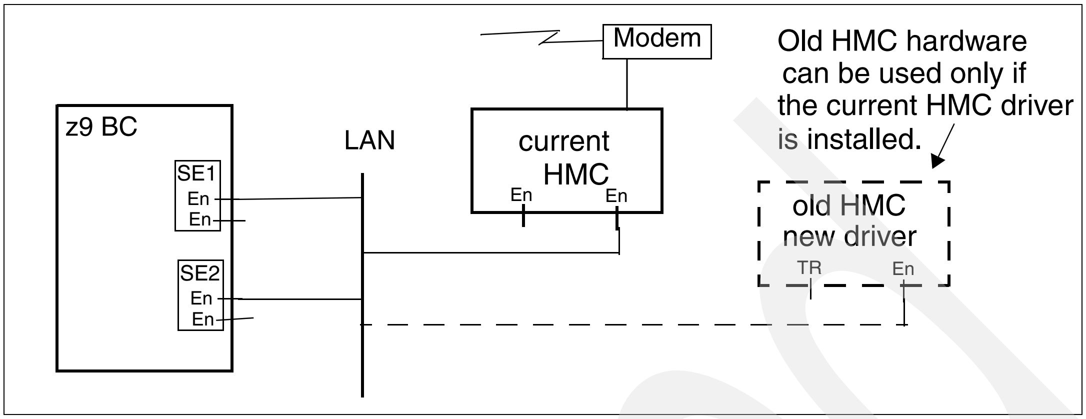

Figure 5-1 Basic SE - HMC connections

The modem shown in the figure is required if a LAN connection from the HMC to the IBM Network is not available. It is used for Remote Support Facility (RSF) connections to transfer microcode updates (MCLs) and system status information via automated scheduled transmissions. Also, if a system fault occurs, the HMC places a call to the IBM Support System, alerting IBM and transferring error information and logs to reduce the impact of unplanned outages.

Figure 5-1 contains only a single z9 BC system. SEs from other systems (System z9, zSeries, G5 or G6) could also be connected to the LAN and operated from the HMC. Other HMCs (new or at a compatible driver level) could be connected to the LAN. In this situation, all systems (assuming that they have appropriate driver levels and patches) could be operated from the current HMCs, but the z9 BC systems could not be operated from the old HMCs.

HMC Version 2.9.0 or later does not use Distributed Console Access Facility (DCAF) to access other console screens. DCAF has been replaced by a version of Desktop-on-Call (DTOC) suitable for the *closed system* HMC. Using DTOC, you can access the desktops of other consoles, such as a Sysplex Timer, an ESCON Director, or any zSeries server.

The only network protocol supported by the HMC application is TCP/IP. Several specific considerations are important:

- The auto-discovery function using NetBIOS, which automatically located SEs and HMCs, is no longer present. The HMC now uses an IP Multicast mechanism instead of NetBIOS to automatically discover SEs and HMCs. If IP Multicast is disabled in the network, then auto-discovery will not happen.
- SNA protocols are no longer supported. All SNA-based functions have been replaced by the IP equivalent. For example, tn3270 sessions are used instead of SNA 3270 sessions.
- The Remote Support Facility (RSF), also known as the *call home* function, can be done through the LAN (if the HMC desktop has access to the Internet) or via an asynchronous modem connection. SDLC modem connections are not supported.

SEs and HMCs may be connected through public LANs, and this may be the best option for the *call home* function. The other option is the use of a dial-up asynchronous line. However, the amount of data traffic may, at times, create substantial delays when depending on a dial-up connection.

A browser connection to a HMC always uses SSL encryption. A user public key certificate must be supplied with a new HMC installation. Every browser connection to an HMC requires a user ID and a password. Be sure to change the default user IDs and passwords as soon as possible.

5.1.3 Planning recommendations

A certain amount of planning and early actions are important for smooth implementation of the current HMCs. Important points to consider include:

- If possible, installation of a current HMC (or the new driver on an appropriate old HMC) should be done well before a z9 BC system is expected. Hands-on usage and training can be done while using the current HMC on existing zSeries or G5 and G6 servers. There is also a course available on IBM Resource Link.
- Hands-on usage and training for the HMC is important, whether before or after a new z9 BC system arrives. The traditional HMC GUI provided on the new HMC is fairly close to that of older HMC versions, but there are still differences. Anyone who clearly understands the functions being used should not have a problem, but operators attempting to follow existing scripts or run books may need practice on the new HMC level.
- Anyone wanting to work with the new *tree* GUI may need some practice time.
- LAN IP assignments for SEs and HMCs should be designed before a z9 BC system arrives.
- There are obviously security concerns if the SEs and HMCs will be placed on public LANs. The considerations are complex and beyond the scope of this introductory document. However, they should be addressed before the arrival of a z9 BC system.
- If the public LAN will be used, connection provisions might be needed in corporate firewalls.
- An HMC can now use DHCP to obtain its IP addresses for its Ethernet adapters.

5.2 Cabling services and cable planning

Starting with the zSeries z800 server, fiber optic cables, cable planning, labeling, and installation became customer responsibilities for new installations and upgrades. With a few exceptions, fiber cables cannot be ordered as part of your system order.[2] Several vendors, including IBM, offer cables and cable-related services through separate contracts. You must arrange a separate contract for fiber cables and their installation.

Note that ICB-3 and ICB-4 cables *are* provided as part of a system order. These are not addressed in the optional contract services discussed here. Also, remember that parallel channels are not directly supported on the z9 BC system, and parallel channel cables are not addressed by these contract services.

To better serve such cabling needs, IBM offers contract services to provide and install cables. These services take into consideration the requirements for all of the protocols and media types that are supported on the z9 BC system and zSeries servers, and can encompass a data center, a storage area network (SAN), LANs, or the end-to-end enterprise.

You can find additional information in the *IBM eServer zSeries IBM System z Connectivity Handbook*, SG24-5444.

[2] Fiber optic conversion kits and mode conditioning patch cables are not orderable as part of a z9 BC system order.

Fiber cable connectors

Some fiber cable connectors used on older servers are not used on the z9 BC system. Figure 5-2 shows the connectors being used on z9 BC systems.

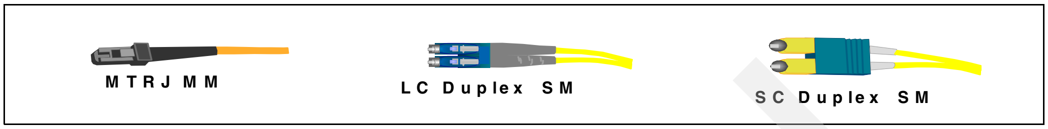

Figure 5-2 z9 BC system fiber connectors

Reuse of existing cables may require conversion kits. Figure 5-3 illustrates the most common conversion kits that are available through the IBM fiber cabling services offering.

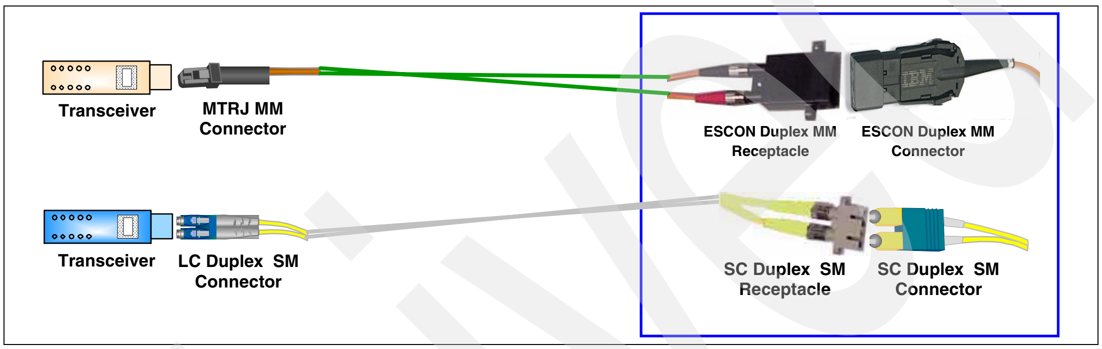

Figure 5-3 Under-floor connector conversion kits

Reuse of a multimode trunk cable is also something to consider. If you have a multimode fiber infrastructure, use of Mode Conditioning Patch (MCP) cables allows reuse of that infrastructure with LX features that normally attach to single-mode fiber connectors. Two MCP cables are required per link — one at each end. There are some restrictions when using MCP:

- Restricted to 1-Gigabit links only
- Not supported with 2-Gigabit ISC-3 or FICON Express features
- Feature-to-device distance limited to 550 meters (1804 feet)

You should consider migrating from a multimode infrastructure to a single-mode infrastructure. You can find more specific information about fiber cables in *Planning for Fiber Optic Links*, GA23-0367.

5.3 I/O channel considerations

This section discusses topics that should be considered when dealing with the I/O channel options that are available with the z9 BC server. For a list of the supported I/O features on the z9 BC server, refer to Appendix B, "Channel options" on page 91.

5.3.1 Parallel channel migration

Parallel channels[3] are not supported on z9 BC systems. If you still need parallel channel devices[4] you must use *ESCON converters* to connect the parallel devices to ESCON channels. IBM recommends using the Optica ESCON Converter.

[3] Often known as *bus and tag* or OEMI channels; named for the functions of the two cables used for these channels.
[4] We use the term *devices* here to include the associated control units. It is the control units that actually use the channels.

As a general recommendation on parallel channels:

- Keep only selected parallel devices and consolidate them onto as few channels as possible. Remember that *daisy chaining* is possible with ESCON converters.
- Consider replacing the following parallel channel devices:
  - IBM 3174 control units, used for 3270 operating system consoles and TSO/CICS 3270 terminals, can be replaced by the OSA-Express ICC function.
  - Some Communication Control Units (CCU), such as 3745-type units running NCP, can be replaced by an OSA-Express2 1000BASE-T Ethernet (CHPID type OSN) and Communication Controller software for Linux on zSeries (CCL).

5.3.2 ESCON channel planning

On z9 BC systems, ESCON channels are always delivered with 16-channel I/O feature cards. The z900 4-port cards (FC2313) cannot be used. Up to 15 channels on each card are available for use. The last channel is reserved as a spare. In practice, any unallocated ports on the card can act as a spare.

At least two ESCON channel feature cards are always installed if any ESCON channels are configured for the processor. ESCON channels are ordered in groups of four. For configurations greater than 28 ESCON channels, individual (not pairs) ESCON channel cards are added as necessary. The active channels are distributed across the physical cards to provide additional redundancy.

These ESCON cards, which are also used with z990, z890, z900, and z800 servers, use MT-RJ connectors. These are different from the traditional ESCON connections that are familiar to most S/390 owners. You can use an ESCON cable with an MT-RJ connector on one end (for the channel connection) and a traditional ESCON connector on the other end (for the control unit). Or you can use conversion cables to adapt existing ESCON cables for use with the MT-RJ connectors.

ESCON channels do not support channel spanning, and therefore cannot be shared across channel subsystems.

Consideration for ESCON conversion channels

There are considerations for ESCON conversion channels (which are connected to an ESCON convertor). When one of the conversion channel types (CHPID type CVC or CBY) is defined, the channel hardware expects that an ESCON convertor is connected to the channel. If the convertor is not connected, a permanent hardware error may be reported at POR. We recommend that you do not define a conversion channel type until the convertor is actually connected.

ESCON channels with CHPID type CVC or CBY cannot be shared across logical partitions. They can only be defined as a reconfigurable channel.

5.3.3 FICON planning

FICON Express and FICON Express2 may be carried forward from previous zSeries servers when upgrading. Only FICON Express4 can be ordered (as a new component) with a z9 BC system. The FICON Express2 LX feature and FICON Express4 feature do not support FICON Bridge (CHPID type FCV).

FICON channels can also work as Fibre Channel Protocol (FCP) channels to access SCSI devices on a Storage Area Network, for example. There is no orderable feature for FCP channels. Any FICON can work as an FCP channel by specifying the CHPID type as FCP.

FCP point-to-point attachments are now supported with FICON Express, FICON Express2, and FICON Express4 features. Point-to-point connections may be used to access data stored on devices without the use of a Fibre Channel switch. Also, an operating system or other standalone program can be IPLed via a point-to-point connection, using the SCSI IPL feature.

N_Port ID Virtualization (NPIV) provides improved sharing[5] and channel utilization of FCP channels among operating system images in logical partitions or virtual servers with the z9. To utilize NPIV, the Fibre Channel switch to which the FCP channel directly attaches is required to support NPIV. NPIV is not supported with FCP point-to-point attachments.

5.3.4 OSA-Express2 planning

z9 BC systems support OSA-Express and OSA-Express2 features. Only OSA-Express2 features may be ordered as new features. OSA-Express features can be brought forward when upgrading an older server. In the following discussion, any reference to OSA-Express2 also applies to OSA-Express unless explicitly stated otherwise.

Some OSA features are not supported on the z9 BC (and were not supported on z9 EC, z990, and z890). These include OSA-2, OSA-2 FDDI, and OSA-Express 155 ATM. Further, OSA-Express Token Ring is not supported on z9 BC systems. Possible methods of replacing these functions are:

- OSA-2: can be replaced by the equivalent OSA-Express2 feature. A new connector might be required. All LAN Channel Stations (LCS) functions are still available; that is, you can use the OSA-Express2 feature in non-QDIO mode and have TCP/IP passthru and SNA on the same port.
- OSA-2 FDDI: can be replaced by OSA-Express2 and a multiprotocol switch or router. The OSA-Express2 would have an Ethernet connection to the switch or router, and that device would perform the FDDI conversion.
- OSA-Express 155 ATM: can be replaced by OSA-Express2 1000BASE-T Ethernet or OSA Express2 Gigabit Ethernet and a multi-protocol switch or router with the appropriate network interface.
- OSA-Express Token Ring: can be replaced by OSA-Express2 1000BASE-T Ethernet or OSA Express2 Gigabit Ethernet and a multi-protocol switch or router with the appropriate network interface.

The OSA-Express Integrated Console Controller function, using an OSA-Express2 1000BASE-T Ethernet feature, should be considered as a replacement for (or instead of) 3174 or 2074 control units. The OSA-ICC is a standard option in the OSA-Express2 1000BASE-T Ethernet feature and is made available through the CHPID=OSC coding on HCD.

Another OSA-Express2 function, new with z9 systems, is the integrated Channel Data Link Control function (OSA-NCP) using OSA-Express2 Gigabit Ethernet and OSA-Express2 1000BASE-T Ethernet adapters. OSA-NCP (OSN) can be used in some cases to replace Communication Control Units such as IBM 3745 units running NCP. This feature is standard in OSA-Express2 Gigabit Ethernet and 1000BASE-T Ethernet (CHPID=OSN) adapters, and must be used with the appropriate Linux application running in a logical partition.

5.3.5 Coupling link connectivity

Coupling link connectivity should be planned carefully, especially if you have 9672 G5 or G6 participating in a Parallel Sysplex environment. For this type of configuration, the coupling link

[5] This is compared to simple FCP connections. These lack the ability to provide separate security controls for the various users connected through multiple logical partitions or multiple guests under z/VM.

connectivity (ICB-2 and ISC-3 in compatibility mode) from 9672 G5 or G6 to the z9 BC server is not supported.

The connection between the z9-EC, z9 BC, and z900/z800 servers can be established with ICB-3 or ISC-3 links in peer mode, although ICB-4 is recommended wherever possible.

The connection with z990 or z890 servers can be made with ICB-3, ICB-4, or ISC3 links in peer mode.

The ICB-3 and ICB-4 links are restricted by the distance between servers to be connected. They cannot be more than 7 meters apart.

5.4 I/O configuration definition and management

Compared to pre-z990 servers, the z9 BC system requires a new approach to initial IOCP/IOCDS setup and an awareness of the layer of I/O definitions associated with PCHIDs. Every CHPID (except internal CHPIDs) must have a PCHID associated with it.

The z9 BC system does *not* have default CHPIDs assigned to channel ports as part of the initial configuration process. CHPIDs are assigned to physical channel path identifiers (PCHIDs) in the IOCP input file. It is the customer's responsibility to perform these assignments by using the HCD/IOCP definitions or by importing the output of the CHPID Mapping Tool.

Multiple Subchannel Sets (MSS)

MSS is a new item to be addressed during the I/O configuration definition on z9 BC systems. You can specify a subchannel set during HCD definitions. By default (when no SCHSET operands are present in an IOCP or HCD definition), all subchannels are allocated on MSS 0, including DASD aliases. That is, the default produces an IOCP/IOCDS without MSS dependencies, which might be suitable for many users. If you intend to use MSS, note that there are operating system dependencies and coexistence concerns.

CHPID Mapping Tool (CMT)

The CHPID Mapping Tool (available from IBM Resource Link) provides a mechanism to map CHPIDs onto PCHIDs as required on z9 EC, z9 BC, z990, and z890 servers. Additional enhancements have been built into the CMT to cater to the new requirements of the System z9 and zSeries servers. It provides the best availability recommendations for the installed I/O features and defined configuration.

Recall that you can have up to two sets of 256 channels on a z9 BC system (that is, in two channel subsystems). Each set of channels could have, for example, two CHPIDs X'10' — one in each CSS. To identify each CHPID exclusively, associate a different PCHID to each CHPID. The CHPID Mapping Tool can help with this task.

The CHIPD Mapping Tool takes input from two sources:

- The Configuration Report file (CFreport) produced by the IBM order tool (e-Config), which can be obtained from Resource Link or from your account team
- An IOCP statement file

The following output is produced by the CMT:

- Tailored reports

  All reports should be saved for reference. The Port Report sorted by CHPID number and location should be supplied to your IBM hardware service representative for the z9 BC installation.

- An IOCP input file with PCHIDs mapped to CHPIDs

  This IOCP input file can then be migrated back into HCD, from which a production IODF can be built.

> **Important:** When an IOCP statement file is exported from a Validated Work IODF using HCD, it must be imported back to HCD for the process to be valid. The IOCP file cannot be used directly by the IOCP program.

Figure 5-4 shows the configuration management process for a new installation of a z9 BC.

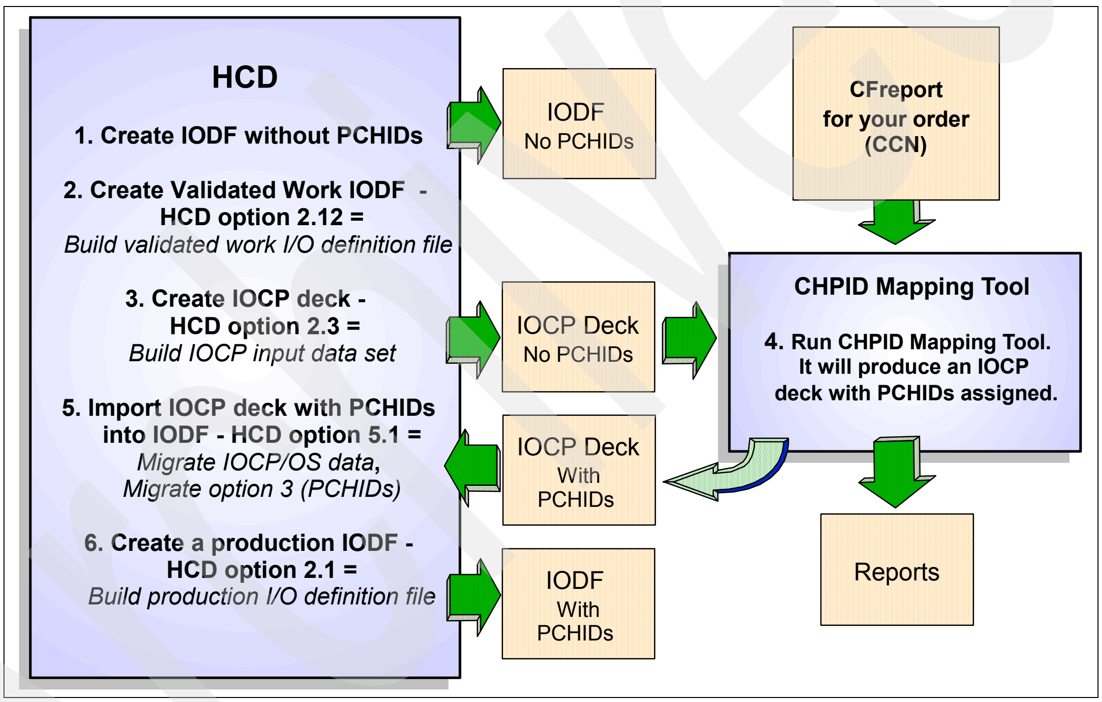

Figure 5-4 z9 BC I/O configuration definition flow

For an MES upgrade, the initial IODF and IOCP statements should contain the PCHID values for the system being upgraded.

For a comprehensive explanation of the CMT, refer to *IBM System z Connectivity Handbook*, SG24-5444.

Stand-alone IOCP

If you are moving from a z900 or previous server to a z9 BC server and you are running an earlier operating system version not supported on a z9 BC server, you may have to create a stand-alone IOCP source file to start up the new environment. To work directly with the IOCP source, you must import it from the HMC diskette drive or FTP it from another system on the LAN. (The z9 BC support elements do not have diskette drives.) If performing an MES upgrade to a z9 BC, the preferred method of creating the initial IOCDS is to use the "Write

IOCDS in anticipation of an upgrade" function in HCD or CHECKCPC=NO in IOCP to write the new IOCDS on your old server prior to the upgrade.

For more information about IOCP definitions, refer to *Input/Output Configuration Program User's Guide,* SB10-7037.

5.5 Capacity upgrades

The z9 BC system provides concurrent, on demand upgrades for the server hardware. With operating system support and appropriate planning, concurrent upgrades can also be non-disrupting to the operating system. Such upgrades provide additional capacity (memory, processors) without a server outage.

Subcapacity models (together with concurrent upgrades) provide flexibility in capacity settings and configurations. Activating or converting processor units (PUs) is used to change the capacity of the server. Existing capacity can be redistributed (for example, one faster processor instead of two slower processors) without changing overall capacity.

You can:

- Add capacity by adding central processors (CPs).
- Add capacity by increasing the capacity settings of installed CPs.
- Add capacity by adding CPs and increasing capacity settings of all CPs (to the same capacity setting).
- Add capacity by increasing the capacity settings and reducing the number of CPs.
- Add capacity by decreasing the capacity settings and increasing the number of CPs.

The same options are available for reducing the server capacity. Specialty engines do not provide subcapacity settings. Changes in their capacity are made by adding or removing specialty engine PUs.

Capacity upgrades can be either permanent or temporary:

- Permanent
  - Ordering process (MES)
  - Capacity Upgrade on Demand (CUoD)
  - Customer Initiated Upgrade (CIU)
- Temporary
  - On/Off Capacity on Demand (On/Off CoD)
  - Capacity Backup (CBU)

Although most capacity upgrades to the z9 BC server are concurrent, software may not be able to take advantage of the increased capacity without performing an Initial Programming Load (IPL) or an image activation.

5.5.1 Capacity Upgrade on Demand

CUoD allows for the nondisruptive addition of Central Processors (CPs), Internal Coupling Facilities (ICFs), Integrated Facilities for Linux (IFLs), System z9 Integrated Information Processors (zIIPs), and System z Application Assist Processors (zAAPs). CUoD can add capacity up to the maximum number of available inactive PUs. For certain configurations, 8 GB memory increments can be added by CUoD.

Specific CUoD functions are:

- Non-disruptive CP, ICF, IFL, zAAP, and zIIP upgrades
- Dynamic upgrade of all I/O cards in the I/O cage
- Dynamic upgrade of spare installed memory

5.5.2 Customer Initiated Upgrade (CIU)

CIU is designed to enable you to respond to sudden increased-capacity requirements by downloading and applying a processor unit or memory upgrade via IBM Resource Link and the Remote Support Facility. With the Express option for CIU, an upgrade may be installed within a few hours after order submission.

Orders (MESs) for processor units (PUs) and memory can be delivered by Licensed Internal Code Configuration Control (LIC CC).

5.5.3 On/Off Capacity on Demand

On/Off CoD is used when you need short-term additional capacity. On/Off CoD is designed to temporarily turn on previously uncharacterized processor units (PUs) (or any unassigned IFLs that are available within the current server) as CPs, ICFs, IFLs, zAAPs, and zIIPs.

These rules apply:

- The number of zAAPs (after the upgrade) cannot exceed the number of CPs (after the upgrade). The number of zIIPs (after the upgrade) cannot exceed the number of CPs (after the upgrade).
- The number of CPs, zAAPs, and zIIPs (after the upgrade) cannot be more than twice the number of the same element before the upgrade. If the server has one CP before the upgrade, it cannot have more than two CPs after the upgrade. If it has one zAAP before the upgrade, it cannot have more than two zAAPs after the upgrade, and so forth.
- The server (after upgrade) cannot provide more than twice the *capacity* of the server (before the upgrade).

Activation of this capability is mutually exclusive with Capacity Backup Upgrade (CBU) activation. Both On/Off CoD and CBU can reside on a z9 BC server, but only one can be activated at a time. On/Off CoD is delivered through the function of Customer Initiated Upgrade (CIU). To participate in this offering, you must have CIU Enablement (FC 9898) and On/Off CoD Enablement (FC 9896) installed. Subsequently, you may concurrently install temporary capacity by ordering:

- 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 will be billed for each On/Off CoD capacity turned on in any given 24-hour period, continuing until such On/Off CoD capacity is turned off. Each month your bill is calculated for the sum of all orders installed in the prior month. Monitoring occurs through the server call-home facility and a bill is generated if the capacity has been enabled for any portion of a calendar month. You will continue to be billed for use of temporary capacity until you return the server to the original state. You do not have to return (but you can) to original capacity if you want to activate another On/Off CoD. You can activate it on top of an existing configuration, provided that the basic upgrade rules apply. If you increase capacity multiple times during a 24-hour period, the charge applies to the highest amount of capacity activated, regardless of how long it was activated.

A 60-minute grace period is allowed during a 24-hour On/Off CoD period. It 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.

It is also possible to pre-stage a number of orders on Resource Link, and they do not have to be installed in order sequence.

When you dispose of the z9 BC server, or decide that you want to disable future temporary upgrades, you are required to remove the enablement feature. One no-charge test period per contract is available for a maximum of 24 hours, starting with the activation of an On/Off CoD order. Tests that exceed 24 hours are treated in their entirety as billable On/Off CoD upgrades.

A special *Administrative On/Off CoD Test* is available. It enables a customer to order *zero quantity* features via ResourceLink for either pre-staging an On/Off CoD order or for activating and deactivating a zero-quantity On/Off CoD upgrade. This zero-cost feature enables the customer to test the On/Off CoD acquisition process for testing and training purposes. The number of zero-quantity tests is unlimited, and no time period restrictions apply. The actual server capacity is not affected by these tests.

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.

There is also a new enhancement for automation code. Using a new API, automation code can be modified by commands based on computer-generated events instead of human intervention.

5.5.4 Capacity Backup Upgrade

The z9 BC server Capacity Backup Upgrade (CBU) capability is a non-disruptive temporary addition of CPs, IFLs, ICF, zAAPs, and zIIPs for disaster-recovery purposes. This makes it possible to concurrently increment the capacity of your system, using LIC CC, in the event of an unforeseen loss of substantial System z9 and zSeries computing capability at one or more of your eligible sites. CBU can be active on a server for up to a 90-day period.

> **Note:** CBU is for disaster/recovery purposes *only* and *cannot* be used for peak workload management.

A Special Bid Contract for CBU must be approved before the CBU features can be ordered. The upgrade (and subsequent downgrade) is executed via the HMC. The CBU features are as follows:

- FC 7870-7895 (for CPs) (Each capacity setting for the CPs has a separate feature code.)
- FC 7896 (for IFLs)
- FC 7897 (for ICFs)
- FC 7899 (for zAAPs)
- FC 7898 (for zIIPs)

The count of required CBU features is the number of CPs, IFLs, ICFs, zAAPs, and zIIPs to be added or altered. Both On/Off CoD and CBU can reside on the server, but only one can be activated at a time.

The CBU offering allows up to five tests over five years, with each test up to 10 days in length. It is not allowed to run production workload on CBU server during the test. However, it is allowed to run production copy.

Subcapacity CP CBU

Subcapacity CBU is available for z9 BC CPs. The rules are:

- The CBU configuration cannot be a configuration that does not exist as a permanent offering. For example, we cannot have CBU configuration E03 because a three-way system at capacity level E is not available as a standard system.
- We can have only one CBU configuration. It is activated in CBU mode. This configuration is used for both CBU testing and CBU disaster operation. The CBU configuration may be changed by ordering a different CBU configuration to replace the existing CBU configuration. However, at any one time, there is only one CBU configuration.
- The CBU order is based on the *active* configuration. If, for example, we order a D03 (three processors at capacity level D) but have IBM activate only two of the processors, then a CBU order is based on two processors. (In this case the third processor is for future growth. At the time IBM activates it, the CBU order should be revised.)
- A CBU order cannot reduce the number of CPs that are *active*.
- Any type of PU can be added as a CBU engine. This is true even when no permanent engines of the same type exist. For example, the CBU configuration could include a zAAP even if no zAAP exists in the permanent configuration. However, the CBU configuration cannot exceed the standard rules for the number of zAAPs and zIIPs installed.
- Permanent PUs cannot be converted to another PU type for the CBU configuration. We cannot, for example, convert an IFL to a CP. If we have a normal configuration of two IFLs and want a CBU configuration of two CPs, then the CBU configuration will be two IFLs plus two CPs.
- All CPs must operate at the same capacity setting in both normal and CBU configurations. This capacity setting might be different between normal and CBU configurations.
- The z9 BC system can change models during CBU operation. That is, it can change from a model R07 to an S07.

Figure 5-5 illustrates an example of possible CBU upgrades for a U02 server (which has two CPs at capacity setting U). Note that CBU can neither reduce the number of CPs, nor lower server capacity.

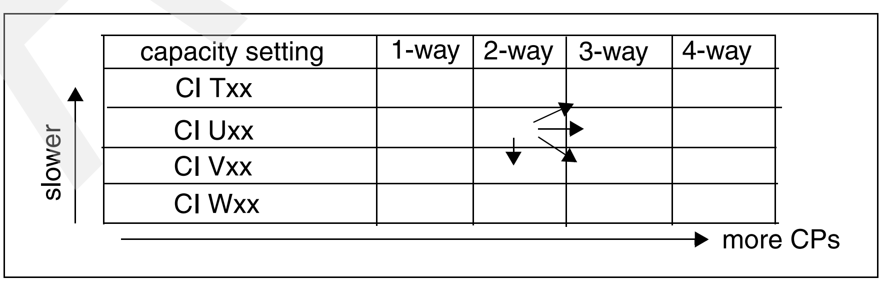

Figure 5-5 Possible CBU upgrade directions

The number of required CBU features depends on the upgrade configuration. If more CPs are added at the existing capacity setting, then the number of CBU features needed is equal to the number of CPs being added. If the upgrade involves changing the capacity setting (for all CPs), then the number of CBU features needed is equal to the number of CPs in the upgraded configuration.

Figure 5-6 shows six different scenarios:

- Going from U02 to U04 requires two “U” CBU features (feature code 7890)
- Going from U02 to U03 requires one “U” CBU feature (feature code 7890)
- Going from U02 to V02 requires two “V” CBU features (feature code 7891)
- Going from U02 to V04 requires four “V” CBU features (feature code 7891)
- Going from U02 to W04 requires four “W” CBU features (feature code 7892)
- Going from U02 to T03 requires three “T” CBU features (feature code 7889)

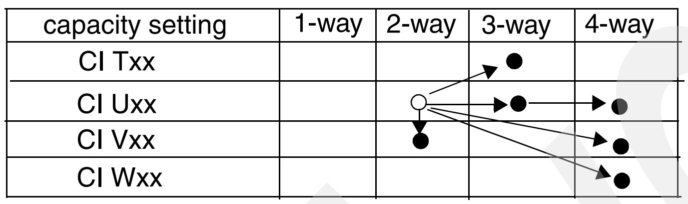

Figure 5-6 CP CBU scenarios

If the CBU configuration uses the same capacity setting as the permanent configuration, then the number of CBU features required is the number of CP increments at that capacity setting. If the CBU configuration uses a different capacity setting than the permanent configuration, then the number of CBU features required is the number of CPs wanted at the CBU capacity setting. The ordering configuration automatically produces the proper feature codes required for the CBU order.

Changes to CBU configuration

We can change our CBU configuration by ordering a different configuration. If the new CBU configuration requires additional engines to be configured or changed then there is a charge involved. If the new CBU configuration does not change the number of engines that are added or changed, or decreases the number that are added or changed, there is no charge involved (although there may be a processing fee for the change). Any future changes will use the new CBU configuration as the starting point.

Consider this example:

- Permanent system configuration is B02
- First CBU order is for D02
- If the later CBU order changes the D02:
  - If it changes to C02: There is *no* charge for the two “C” CBU feature codes needed because the CBU system capacity at C02 is less than the CBU system capacity at D02. However, future CBU changes will be based on C02 as a starting point.
  - If it changes to J01: There *will* be a charge involved.

5.5.5 Capacity upgrade summary

Table 5-1 provides a summary of the upgrade options.

Table 5-1 Capacity upgrade summary

	CUoD	CIU	OOCoD	CBU
Type	Permanent	Permanent	Temporary	Temporary
Concurrency	Concurrent	Concurrent	Concurrent	Concurrent
Purpose	Permanent capacity change	Permanent capacity upgrade	Handle peak workloads	Disaster recovery[a]
System	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890	z9 EC, z9 BC, z990, z890, z900, z800
Mutually exclusive with			CBU[b]	OOCoD[b]
On top of previous?	Yes	Yes	Yes	No
Planning	Required	Required	Required	Required
Enablement		CIU contract required	CIU and OOCoD contracts required	CBU contract required
Download and Activate/Install	IBM	Customer	Customer	Customer
Process	Ordered as a normal upgrade	Initiated via Web	Initiated via Web	Initiated via Web
Pre-staged order	Not applicable	Yes	Yes, number of orders can be pre-staged	Yes
Test allowed	Not applicable	Not applicable	Once per contract for 24 hours, 60-minute grace period[c]	Up to five times, ten days test, during five years period
Max. active period	Permanent	Permanent	No limit	Up to 90 days
Deactivation	Not applicable	Not applicable	Customer via HMC	Customer via HMC
Adding CP	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890	z9 EC, z9 BC, z990, z890, z900, z800
Changing CP capacity	z9 EC, z9 BC, z890	z9 EC, z9 BC, z890	z9 EC, z9 BC, z890	z9 EC, z9 BC, z890
Removing CP	z9 EC, z9 BC, z990, z890, z900, z800	No	No	No
Adding ICF and IFL	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890, z900, z800	z9 EC, z9 BC, z990, z890	z9

	CUoD	CIU	OOCoD	CBU
Removing ICF and IFL	z9 EC, z9 BC, z990, z890, z900, z800	No	No	No
Adding zAAP	z9 EC, z9 BC, z990, z890	z9 EC, z9 BC, z990, z890	z9 EC, z9 BC, z990, z890	z9 EC, z9 BC
Removing zAAP	z9 EC, z9 BC, z990, z890	No	No	No
Adding zIIP	z9 EC, z9 BC	z9 EC, z9 BC	z9 EC, z9 BC	z9 EC, z9 BC
Removing zIIP	z9 EC, z9 BC	No	No	No
Adding memory[d]	z9 EC, z9 BC, z990, z890, z900	z9 EC, z9 BC, z990, z890, z900	No	No
Removing memory	z9 EC, z9 BC, z990, z890, z900	No	No	No

a. CBU contract does not allow use of CBU to handle peak workloads. It is intended *only* for disaster recovery.
b. CBU and OOCoD can coexist on a server, but only one of them can be active at a time.
c. Zero capacity test for training purposes is available on z9 EC and z9 BC. No time or count restrictions apply for it.
d. Some memory upgrades can be disruptive. Careful planning is required.

Note: Table 5-1 on page 82 cannot cover all variations and exceptions. More information can be found in *System z Capacity on Demand User's Guide*, SC28-6846.

5.6 Logical planning

In this section we cover some considerations regarding the *logical planning* for the new z9 BC servers, addressing such concerns as logical partition definitions, coexistence on a Parallel Sysplex with other zSeries servers and z/OS operating systems, and migration from a previous operating system version to a new version that contemplates the new z9 BC features. We assume that:

- The oldest hardware you are migrating from, or replacing, is a z800 server.
- The oldest z/OS operating system that you are running on a single image or sysplex environment is z/OS V1 R4.
- z/OS V1 R4 is running with the exploitation feature installed.

If these assumptions are not valid for you, ask your IBM zSeries Field Technical Support to help you plan the z9 BC server migration or installation. In this case, you might have to go through intermediate steps (not covered here) before moving to the z9 BC server.

5.6.1 LPAR mode considerations

A z9 BC system can have up to 30 logical partitions (model S07) and must have at least one logical partition because basic mode is not available. The logical partition ID is a 1-byte field. Older software may work with 4-bit fields, which could be a concern when creating a system with more than 15 logical partitions. The logical partition ID is specified by the user as part of the LPAR activation profile on the HMC. The following points may be important when planning logical partition definitions:

- A maximum of 15 logical partitions may be associated with a single channel subsystem.
- Operating systems with *compatibility support*[6] can be activated only in a logical partition related to CSS 0 with a logical partition ID equal to or smaller than 15 (x’F’).
- Partition identifiers, names, and numbers associated with a logical partition:
  - *Logical partition identifier*. This is an identifier in the range 0–X’3F’ and is assigned by the user when defining logical partition image profiles through the SE or HMC. It is unique across the z9 BC system. This identifier is returned by an STIDP or STSI instruction. It is also known as the user logical partition ID (UPID).
  - *MIF image identifier*.[7] This identifier is defined through HCD or IOCP and is the partition number defined in the RESOURCE statement in the IOCP source file. It is in the range x’1’ to x’F’ and is unique *within* an CSS. It does not have to be unique within a z9 BC system. The MIF Image identifier is also known as the IID.
  - *Logical partition name*. This name is defined through HCD or an IOCP and is the partition name in the RESOURCE statement in the IOCP source file. A logical partition name must be unique across the z9 BC system.

Figure 5-7 shows where the CSS, logical partition, and MIF definitions are defined.

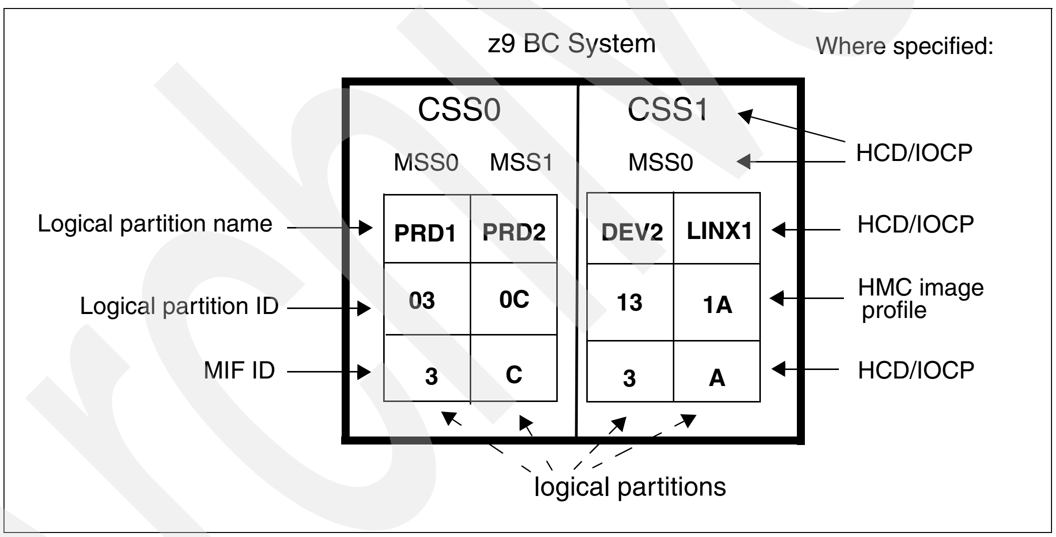

Figure 5-7 Logical partition name definitions

The logical partition ID is specified when defining activation profiles. The HCD field for the MIF Image ID has the label *Partition Number*. This is equivalent to the MIF Image ID used in the IOCP RESOURCE statement. We suggest that you establish a naming convention for the logical partition identifiers. For example, you could use the CSS number concatenated to the MIF Image ID, which means logical partition ID 1A is in CSS 1 with MIF ID A. This fits into the allowed range of logical partition IDs and conveys useful information to the user.

Dynamic add/delete of a logical partition name

The ability to add meaningful logical partition names to the configuration, without a Power-On Reset, is now available. Prior to this support, extra logical partitions were defined by adding

[6] Compatibility support is related to pre-z/OS V1 R4 operating systems. This feature enables the recognition of a logical partition with more than x’F’ for logical partition identifier and logical partition number.
[7] MIF is Multiple Image Facility. It was formerly known as EMIF for ESCON Multiple Image Facility when introduced for ESCON channels. This function now relates to all I/O, so the “E” has been dropped.

reserved names in the Input/Output Configuration Data Set (IOCDS), but it was often difficult to assign meaningful names in advance.

Dynamic add/delete of a logical partition name allows reserved logical partition *slots* to be created in an IOCDS (and reflected in HSA allocations) in the form of extra channel subsystem, Multiple Image Facility (MIF) image ID pairs. A reserved partition is defined with the partition name placeholder * (asterisk), and cannot be assigned to access channel paths or devices.

These extra channel subsystem MIF image ID pairs (CSSID/MIFID) can later be assigned a logical partition name for use (or later removed) via HCD changes and a dynamic activate. The IOCDS still must have the extra I/O slots defined in advance because many structures are built based on these major I/O control blocks in the Hardware System Area (HSA). This dynamic logical partition support is available for the z9 systems and the z990 and z890 servers, and is applicable to z/OS V1.6 and later releases.

When a logical partition is renamed, its name can be changed from NAME1 to * (with a dynamic activate) and then changed again from * to NAME2 (with another dynamic activate). The logical partition MIFID is retained across the logical partition name change. However, the master keys in Crypto Express2 card that were associated with the old logical partition NAME1 are retained. There is no explicit action taken against a cryptographic component for this.

5.7 Coupling Facility planning

The z9 BC system does not provide a special model for a CF-only processor (such as the model 100 for z900 processors). You can, however, have a z9 BC system with all enabled PUs defined as ICFs. (This applies only to the S07 model; there must be at least one CP in model R07). Such a configuration uses a Z00 capacity marker. Some precautions should be taken related to:

- The differences between z9 coupling link connectors and zSeries model connectors
- The ICB link distances and their incompatibility with earlier servers

Driver 55 (usually associated with Coupling Facility Level 13) introduced the application of patches to Coupling Facility Control Code (CFCC) concurrently, in *rolling* mode. As shown in Figure 5-8, on a single CEC you can first apply the patch on a test ICF logical partition, perform all necessary tests, and later on roll the patch to the production logical partition. To use the updated CFCC code in a CF logical partition, simply deactivate and activate the partition. When the CF comes up, it displays its version on the OPRMSG panel for that partition. With this facility there is no need to stop the entire server for CFCC patches. However, CFLevel upgrades will still be disruptive to the entire CEC.

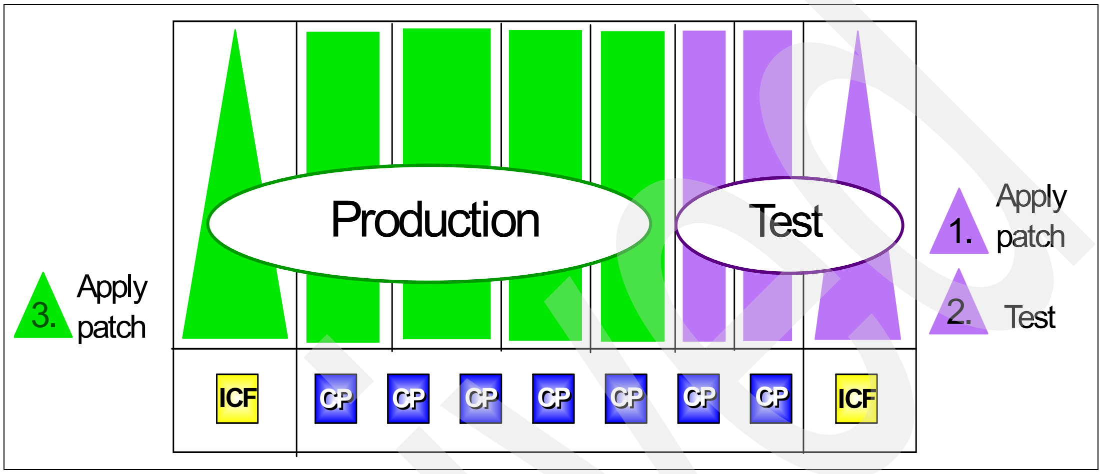

Figure 5-8 Concurrent CFCC patch

5.7.1 Sysplex Timer ETR Network ID

As part of the installation of a Sysplex Timer in either Basic or Expanded Availability configuration, each IBM 9037 Sysplex Timer Unit is assigned an ETR Network ID (0 to 31decimal) and ETR Unit ID (0 to 31decimal). The ETR network ID and ETR Unit ID values are arbitrary (within the valid range) and can be chosen by the customer to uniquely identify an ETR network and a unique ETR unit (Sysplex Timer) within the ETR Network.

For example, in a Sysplex Timer Expanded Availability configuration there are two IBM 9037 Sysplex Timer Units. The first IBM 9037 may have an ETR Network ID value of 0, and a ETR Unit ID value of 0. The second IBM 9037 will have the same ETR Network ID value of 0, but a different ETR Unit ID value of 1. The two Sysplex Timer Units in an Expanded Availability configuration must have the same ETR Network ID value defined. However, their ETR Unit ID values must be unique in the Sysplex Timer ETR Network.

A Sysplex Timer Unit's ETR Network ID, ETR Unit ID, and port number are transmitted along with timing signals to an attached server's ETR port. This information is available to sysplex systems running on the server and can be displayed using the D ETR command.

A function introduced with the z990 server, and present in the z9 BC system, requires the ETR Network ID of the attached Sysplex Timer Network to be manually set in the support element at installation time. In addition, on the same SE panel, the ETR ports must be enabled for stepping the TOD. This function checks whether the ETR Network ID being received in the timing signals via each of the server's two ETR ports matches the ETR Network ID manually set in the server's support element (SE).

This function provides greater checking capability, helping to eliminate cabling errors where a server's ETR port may be incorrectly connected to a Sysplex Timer Unit in an incorrect

Sysplex Timer ETR Network, and allows verification of cabling connectivity from the Sysplex Timer to the z9 BC system prior to IPL of z/OS.

Important: Setting the ETR Network ID to an incorrect value on a z9 BC system's SE will cause all z/OS images with ETRMODE=YES on the server to immediately enter a non-restartable disabled wait state.

Ordering ETR cables

Two ESCON MT-RJ cables for the ETR feature on the z9 BC system will automatically be added to any Fiber Cabling Service contract offered by IBM Global Services. Contact your local IBM Installation Planning Representative, IBM z9 BC system Product Specialist, or IBM Connectivity Services Specialist for details. If you choose not to use this service, you may purchase the two ESCON MT-RJ cables separately, or provide them yourself from another source. An ESCON MTRJ-to-ESCON Duplex conversion kit is also needed to connect the MT-RJ cables to the Duplex connectors on the 9037.

Sysplex Timer attachment planning

ETR ports on z9 BC systems are optional features and must be ordered separately. However, if you order any CF links, an order for ETR ports is generated automatically.

5.8 Operating system migration

Table 5-2 shows a summary of OS support for z9 BC systems. If you are running any earlier OS not listed in this table, you will need to migrate to a more recent operating system version.

Table 5-2 Operating system service dates

Operating system	z9 EC z9 BC	z990 z890	z900 z800	End of service
z/OS V1 R6	Yes	Yes	Yes	09/07
z/OS V1 R7	Yes	Yes	Yes	09/08
z/OS V1 R8	Yes	Yes	Yes	
z/OS V1 R9	Yes	Yes	Yes	
z/VM V5 R1	Yes	Yes	Yes	09/07
z/VM V5 R2	Yes	Yes	Yes	04/09
z/VM V5 R3	Yes	Yes	Yes	09/10
z/VSE V3 R1	Yes	Yes	Yes	
z/VSE V4 R1	Yes	Yes	Yes	
TFP V4 R1	Yes	Yes	Yes	
z/TFP V1 R1	Yes	Yes	Yes	

Stand-alone z9 BC systems can run any of the operating systems shown in the table with the PSP-recommended service applied. However, exploitation of the z9 BC hardware features depends on the operating system version used.

A

Frequently asked questions

Q: Can z/OS.e be used with the z9 BC server?
A: Yes.

Q: Is there a required affinity between a particular CP and a particular zIIP (or zAAP)?
A: No.

Q: Can I define more than 15 logical partitions on the R07 model?
A: Yes, but you cannot activate more than 15 at any one time.

Q: Does a PSC count against the maximum of 16 I/O slots used in the R07?
A: No, a PSC does not reduce the allowed 16 I/O slot usage.

Q: Can I use the new 2-port FICON Express4 adapter with a z9 EC server?
A: No. It is unique to the z9 BC.

Q: Can I carry forward my older OSA-Express adapters?
A: In general, yes.

Q: There are several FICON choices. Is there a preferred version?
A: The FICON Express4 4-port 4km LW version might be considered the standard version.

Q: Can an OSN CHPID be used with a z990?
A: No, it is limited to System z9 servers.

Q: What CFCC level is supplied with the z9 BC server?
A: CFCC 14 was the initial level in z9 BC. Now CFCC 15 is available.

Q: What is the Z00 model?
A: This is a z9 BC model S07 server with capacity specification Z00, meaning there are zero CPs configured. This version is typically used as a stand-alone CF system or IFL-only system.

Q: What is the performance setting for the specialty processors?
A: These always run at the maximum capacity of the processor. Note that the maximum capacity of a processor is the same on both the R07 and S07, and corresponds to the fastest (highest capacity) processor in an S07.

Q: I cannot have more zAAPs or zIIPs than CPs. Can I have more zAAPs+zIIPS than CPs?
A: Yes. For each CP you can have one zAAP, one zIIP, or one zAAP + one zIIP.

Q: Can I mix dedicated and shared processors in one logical partition ?
A: No. Dedicated and shared processors cannot be mixed in one logical partition, regardless of their type.

Q: Can I assume that the new 4 Gbps FICON channel is really twice as fast as the existing 2 Gbps FICON channel?
A: The *link speed* is twice as fast, assuming that the outboard devices operate at 4 Gbps. However, the link speed is only one of many factors that determine the *effective speed* of channel operations. A faster link speed provides the potential for faster effective speeds, but the total environment must be examined to determine how various constraints affect operation.

Q: Does MIDAW operate on multiple blocks on a track? Or is it limited to operations within a single block?
A: In traditional CKD operation MIDAW is limited to operations within a single block, in the same way that data chaining CCWs operate. However, non-traditional full-track operations treat multiple blocks on a track as though they were a single block, and MIDAW is effective in this situation.

B

Channel options

Table B-1 lists the channel options that are supported on the z9 BC server with required connector and cable types, and the maximum unrepeated distances.

Model R07 can have a maximum of 16 slots, and model S07 provides 28 slots. At least one major I/O feature (FICON, ESCON, ICB, ICS) is required.

Table B-1 z9 BC channel feature support

Channel feature	Feature codes	Connector type	Cable type	Maximum unrepeated distance
16-port ESCON	2323[a]	MT-RJ	MM 62.5 µm	3 km
FICON Express LX[b]	2319	LC Duplex	SM 9 µm	10 km/20 km[c]
FICON Express SX[b]	2320	LC Duplex	MM 62.5 µm MM 50 µm	120 m 300 m
FICON Express2 LX[b]	3319	LC Duplex	SM 9 µm	10 km/12 km[c]/20 km[c]
FICON Express2 SX[b]	3320	LC Duplex	MM 62.5 µm MM 50 µm	120 m 300 m
FICON Express4 LX (10 km)	3321	LC Duplex	SM 9 µm	10 km
FICON Express4 SX 2000MHz	3322 3318	LC Duplex	MM 50 µm	860/500/270[d] m
FICON Express4 SX 500MHz	3322 3318	LC Duplex	MM 50µm	500/300/150[d] m
FICON Express4 SX 200MHz	3322 3318	LC Duplex	MM 62.5 µm	300/150/70[d] m
FICON Express4 SX 160MHz	3322 3318	LC Duplex	MM 62.5 µm	250/120/55[d] m
FICON Express4 LX	3323	LC Duplex	SM 9 µm	4 km
FICON Express4 LX	3324	LC Duplex	SM 9 µm	4 km

Channel feature	Feature codes	Connector type	Cable type	Maximum unrepeated distance
OSA-Express GbE LX[b]	1364	LC Duplex	SM 9 µm	5 km
			MCP	550 m
OSA-Express GbE LX[b]	2364	SC Duplex	SM 9 µm	5 km
			MCP	550 m
OSA-Express GbE SX[b]	1365	LC Duplex	MM 62.5 µm	220 m 275 m
			MM 50 µm	550 m
OSA-Express GbE SX[b]	2365	SC Duplex	MM 62.5 µm	220 m 275 m
			MM 50 µm	550 m
OSA-Express 1000BASE-T Ethernet[b]	1366	RJ 45	UTP Cat 5	100 m
OSA-Express Fast Ethernet[b]	2366	RJ 45	UTP Cat 5	100 m
OSA-Express2 GbE LX	3364	LC Duplex	SM 9 µm	5 km
			MCP	550 m
OSA Express2 GbE SX	3365	LC Duplex	MM 62.5 µm	220 m 275 m
			MM 50 µm	550 m
OSA-Express2 1000BASE-T Ethernet	3366	RJ 45	UTP Cat 5	100 m
OSA-Express2 10 GbE LR	3368	SC Duplex	SM 9 µm	10 km
HiperSockets	n/a	n/a	n/a	n/a
IC[e]	n/a	n/a	n/a	n/a
ICB-3[e]	0993	FC 0227		10 m
ICB-4[e]	3393	FC 0228		10 m
ISC-3 (peer mode)[e]	0217[f] 0218 0219	LC Duplex	SM 9 µm	10 km/12 km[c]/20 km[c]
ISC-3 (RPQ 8P2197 / Peer mode at 1 Gbps)[g]				
Crypto Express2[h]	0863	n/a	n/a	n/a
Crypto Express2-1P[i]	0870	n/a	n/a	n/a
ETR link	6155	MT-RJ	MM 62.5 µm MM 50 µm	3 km/26 km[c] 2 km /24 km[c]

a. ESCON channels are ordered in increments of four using feature code 2324.
b. Only supported when carried forward on an upgrade.
c. Requires RPQ 8P2263. Maximum distances vary based on the link data rate.
d. Depends on link speed (1/2/4 Gbps).
e. The maximum number of coupling links (excluding IC) that can be ordered is 64. The total number of coupling links that can be defined is 64, including IC links.

f. 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).
Individual ISC-3 port activation must be ordered using feature code 0219.

g. RPQ 8P2197 enables the ordering of a different daughter card supporting 20 km unrepeated distance for 1 Gbps peer mode. RPQ 8P2262 is a requirement for that option, and other than the normal mode, the channel increment is two (that is ,both ports (FC 0219) at the card have to be activated).

h. The initial order is two features (4 PCI-X adapters). Each PCI-X adapter can be configured (via HMC/SE) as either a secure coprocessor or accelerator.

i. The initial order is two features (2 PCI-X adapters). Single and dual Crypto Express2 features can be mixed. The two feature minimum can be one of each or two of either. Each PCI-X adapter can be configured (via HMC/SE) as either a secure coprocessor or accelerator.

Related publications

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

IBM Redbooks

For information about ordering these publications, see "How to get IBM Redbooks" on page 95. 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
- *HiperSockets Implementation Guide*, SG24-6816
- *IBM System z9 109 Technical Introduction*, SG24-6669
- *OSA-Express Implementation Guide*, SG24-5948
- *OSA-Express Integrated Console Controller Implementation Guide*, SG24-6364
- *System z9 and zSeries Networking Solutions,* SG24-7125
- *z9-109 Crypto and TKE V5 Update*, SG24-7123
- *IBM System z9 Enterprise Class Technical Guide*, SG24-7124

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

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM System z9 Business Class Technical Introduction

IBM System z9 Business Class Technical Introduction

Describes new server hardware and related features

Discusses key functional elements and enhancements

Reviews associated software support and considerations

This IBM Redbooks publication introduces the IBM System z9 Business Class (z9 BC) server, which provides the advantages of the IBM System z9 Enterprise Class (z9 EC) servers in a less expensive base. The z9 BC server is based on z/Architecture and provides extensions to this architecture. At the time of writing the z9 BC server is the latest product in this line. The z9 BC server provides the first implementation of zIIP special processors. It also provides an unprecedented range of capacity versions.

This book provides basic information about new hardware functions and features, associated software support, and migration considerations. It is intended for systems engineers, hardware planners, and anyone else needing to understand the new elements.

Readers are expected to be generally familiar with current IBM System z technology and terminology. This book is not intended as an introduction to mainframes. It is especially directed to readers most familiar with z890, z800, and earlier systems. Functions new in the existing IBM z9 servers are briefly mentioned because they may not be familiar to this set of readers.

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

SG24-7241-01 ISBN 0738489352