Note: Before using this information and the product it supports, read the information in “Notices” on page 9.

Twentieth Edition (October 2018)

This edition applies to connectivity options available on the IBM z14 (M/T 3906), IBM z14 Model ZR1 (M/T 3907), IBM z13, IBM z13s, IBM zEnterprise EC12 (zEC12), and IBM zEnterprise BC12 (zBC12).
# Contents

<table>
<thead>
<tr>
<th>Notices</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trademarks</td>
<td>10</td>
</tr>
<tr>
<td><strong>Preface</strong></td>
<td>11</td>
</tr>
<tr>
<td>Authors</td>
<td>11</td>
</tr>
<tr>
<td>Now you can become a published author, too!</td>
<td>12</td>
</tr>
<tr>
<td>Comments welcome</td>
<td>13</td>
</tr>
<tr>
<td>Stay connected to IBM Redbooks</td>
<td>13</td>
</tr>
</tbody>
</table>

## Chapter 1. Introduction

<table>
<thead>
<tr>
<th>1.1 I/O channel overview</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1 I/O hardware infrastructure</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 I/O connectivity features</td>
<td>3</td>
</tr>
<tr>
<td>1.2 FICON Express</td>
<td>4</td>
</tr>
<tr>
<td>1.3 zHyperLink Express</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Open Systems Adapter-Express</td>
<td>6</td>
</tr>
<tr>
<td>1.5 HiperSockets</td>
<td>7</td>
</tr>
<tr>
<td>1.6 Parallel Sysplex and coupling links</td>
<td>8</td>
</tr>
<tr>
<td>1.7 Shared Memory Communications</td>
<td>9</td>
</tr>
<tr>
<td>1.8 I/O feature support</td>
<td>10</td>
</tr>
<tr>
<td>1.9 Special-purpose feature support</td>
<td>12</td>
</tr>
<tr>
<td>1.9.1 Crypto Express features</td>
<td>12</td>
</tr>
<tr>
<td>1.9.2 Flash Express feature</td>
<td>13</td>
</tr>
<tr>
<td>1.9.3 zEDC Express feature</td>
<td>13</td>
</tr>
</tbody>
</table>

## Chapter 2. Channel subsystem overview

<table>
<thead>
<tr>
<th>2.1 CSS description</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1 CSS elements</td>
<td>16</td>
</tr>
<tr>
<td>2.1.2 Multiple CSS concept</td>
<td>17</td>
</tr>
<tr>
<td>2.1.3 Multiple CSS structure</td>
<td>17</td>
</tr>
<tr>
<td>2.1.4 Multiple image facility</td>
<td>18</td>
</tr>
<tr>
<td>2.1.5 Physical channel ID</td>
<td>19</td>
</tr>
<tr>
<td>2.1.6 Adapter ID</td>
<td>21</td>
</tr>
<tr>
<td>2.1.7 Multiple CSS construct example</td>
<td>24</td>
</tr>
<tr>
<td>2.1.8 Channel spanning</td>
<td>24</td>
</tr>
<tr>
<td>2.1.9 Multiple subchannel sets</td>
<td>26</td>
</tr>
<tr>
<td>2.1.10 Summary</td>
<td>28</td>
</tr>
<tr>
<td>2.2 I/O configuration management</td>
<td>28</td>
</tr>
<tr>
<td>2.2.1 Hardware configuration definition</td>
<td>29</td>
</tr>
<tr>
<td>2.2.2 CHPID Mapping Tool</td>
<td>29</td>
</tr>
<tr>
<td>2.3 I/O configuration planning</td>
<td>29</td>
</tr>
<tr>
<td>2.3.1 I/O configuration rules</td>
<td>30</td>
</tr>
<tr>
<td>2.4 References</td>
<td>31</td>
</tr>
</tbody>
</table>

## Chapter 3. zHyperLink Express

<table>
<thead>
<tr>
<th>3.1 Description</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 zHyperLink elements</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Connectivity</td>
<td>35</td>
</tr>
<tr>
<td>3.4 References</td>
<td>36</td>
</tr>
</tbody>
</table>
Notices

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

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

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

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

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

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

<table>
<thead>
<tr>
<th>AIX®</th>
<th>IBM-z13®</th>
<th>System Storage®</th>
</tr>
</thead>
<tbody>
<tr>
<td>CICS®</td>
<td><strong>IBM-z13s™</strong></td>
<td>Tivoli®</td>
</tr>
<tr>
<td>Cognos®</td>
<td>IMS™</td>
<td>VTAM®</td>
</tr>
<tr>
<td>Db2®</td>
<td>Tivoli NetView for z/OS®</td>
<td>WebSphere®</td>
</tr>
<tr>
<td>developerWorks®</td>
<td>Parallel Sysplex®</td>
<td>z/Architecture®</td>
</tr>
<tr>
<td>DS8000®</td>
<td>PartnerWorld®</td>
<td>z/OS®</td>
</tr>
<tr>
<td>ECKD™</td>
<td>PR/SM™</td>
<td>z/VM®</td>
</tr>
<tr>
<td>FICON®</td>
<td>Processor Resource/Systems Manager™</td>
<td>z/VSE®</td>
</tr>
<tr>
<td>GDPS®</td>
<td></td>
<td>z10™</td>
</tr>
<tr>
<td>Geographically Dispersed Parallel Sysplex™</td>
<td>RACF®</td>
<td>z13®</td>
</tr>
<tr>
<td>Global Technology Services®</td>
<td>Redbooks®</td>
<td>z13s™</td>
</tr>
<tr>
<td>HiperSockets™</td>
<td>Redbooks (logo)®</td>
<td>z14®</td>
</tr>
<tr>
<td>IBM®</td>
<td>Resource Link®</td>
<td>zEnterprise®</td>
</tr>
<tr>
<td>IBM Z®</td>
<td>Resource Measurement Facility™</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RMF™</td>
<td></td>
</tr>
</tbody>
</table>

The following terms are trademarks of other companies:

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

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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

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

This IBM® Redbooks® publication describes the connectivity options that are available for use within and beyond the data center for the IBM Z family of mainframes, which includes these systems:

- IBM z14®
- IBM z14 Model ZR1
- IBM z13®
- IBM z13s™
- IBM zEnterprise® EC12 (zEC12)
- IBM zEnterprise BC12 (zBC12)

This book highlights the hardware and software components, functions, typical uses, coexistence, and relative merits of these connectivity features. It helps readers understand the connectivity alternatives that are available when planning and designing their data center infrastructures.

The changes to this edition are based on the IBM Z hardware announcement dated April 10, 2018.

This book is intended for data center planners, IT professionals, systems engineers, and network planners who are involved in the planning of connectivity solutions for IBM mainframes.

Authors

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

Bill White is an IBM Redbooks Project Leader and Senior Networking and Connectivity Specialist at IBM Redbooks, Poughkeepsie Center.

Hervey Kamga is an IBM Z Product Engineer with the EMEA I/O Connectivity Team in Montpellier, France. Before that, he was a Support Engineer and Engineer On Site for 13 years with Sun MicroSystems and Oracle in EMEA. Hervey’s areas of expertise include Oracle Solaris Operating System and hardware products, virtualization (VMware, virtualBox), Linux (Ubuntu), and IBM Z I/O features and protocols (IBM FICON® and OSA).

Michal Kordyzon is an IBM Z Client Technical Specialist at IBM Poland with 12 years of experience with the IBM Z platform. His areas of expertise include mainframe architecture, Linux systems, Hyperledger fabric, Node.js, and Machine Learning algorithms. He also has expertise in LinuxONE.

Frank Packheiser is a Senior zIT Specialist at the Field Technical Sales Support office in Germany. He has 26 years of experience in IBM Z. Frank has worked for 10 years in the IBM Education Center in Germany, developing and providing professional training. He also provides professional services to IBM Z and mainframe clients. In 2008 and 2009, Frank supported clients in Middle East/North Africa (MENA) as a zIT Architect. Besides co-authoring several IBM Redbooks publications since 1999, he has been an official presenter at ITSO workshops for the last four years.
John Troy is an IBM Z and Storage Hardware National Top Gun in the northeast area of the United States. He has 36 years of experience in the service field. His areas of expertise include IBM Z server and high-end storage systems technical and customer support. John has been an IBM Z hardware technical support course designer, developer, and instructor for the last seven generations of IBM high-end servers.

Esra Ufacik is an IBM Z Lead Architect in IBM Systems covering Asia Pacific. Throughout her career, she has held various roles in different companies with Z “flavors,” including IBM z/OS®, IBM Parallel Sysplex®, implementation, migration, SAN, storage, software problem support, solution designing, Linux on IBM Z®, IT Optimization, TCO assessments, technical authoring, and guest lecturing. In March 2016, she was assigned as IOT lead for Blockchain in the IBM Systems division.

Bo Xu is a Senior System Service Representative in China. He has more than 20 years of experience with IBM Z products maintenance support. He has been working in the GTS department. He provides IBM Z support to clients as a local Top Gun, and provides second-level support for IBM DS8000® as a Top Gun and Skill Owner. His areas of expertise include IBM Z hardware, channel connectivity, and IBM DS8000.

Octavian Lascu is an IBM Redbooks Project Leader and a Senior IT Consultant for IBM Romania with over 20 years of experience. He specializes in designing, implementing, and supporting complex IT infrastructure environments (systems, storage, and networking), including high availability and disaster recovery solutions and high-performance computing deployments. He has developed materials for and taught over 50 workshops for technical audiences around the world. He has written several IBM publications.

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

Susan Farrell
Les Geer
Dave Surman
Christine Smith
Barbara Weiler

Thanks to Tom Ambrosio and William Lamastro of the Poughkeepsie Competitive Center (PCC), and Robert Haimowitz from IBM Redbooks for their support and contributions to this publication.

Now you can become a published author, too!

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

Find out more about the residency program, browse the residency index, and apply online at: ibm.com/redbooks/residencies.html
Comments welcome

Your comments are important to us!

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

- Use the online Contact us review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:
  http://www.facebook.com/IBMRedbooks
- Follow us on Twitter:
  http://twitter.com/ibmredbooks
- Look for us on LinkedIn:
  http://www.linkedin.com/groups?home=&gid=2130806
- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
- Stay current on recent Redbooks publications with RSS Feeds:
  http://www.redbooks.ibm.com/rss.html
Introduction

This chapter gives a brief overview of the input/output (I/O) channel architecture and introduces the connectivity options available on IBM Z platforms. It includes the following sections:

- I/O channel overview
- FICON Express
- zHyperLink Express
- Open Systems Adapter-Express
- HiperSockets
- Parallel Sysplex and coupling links
- Shared Memory Communications
- I/O feature support
- Special-purpose feature support

**Note:** The link data rates discussed throughout this book do not represent the actual performance of the links. The actual performance depends on many factors including latency through the adapters and switches, cable lengths, and the type of workload using the connection.
1.1 I/O channel overview

I/O channels are components of the IBM mainframe system architecture (IBM z/Architecture®). They provide a pipeline through which data is exchanged between systems or between a system and external devices (in storage or on the networking). z/Architecture channel connections, referred to as channel paths, have been a standard attribute of all IBM Z platforms dating back to the IBM S/360. Over the years, numerous extensions have been made to the z/Architecture to improve I/O throughput, reliability, availability, and scalability.

One of the many key strengths of the Z platform is the ability to deal with large volumes of simultaneous I/O operations. The channel subsystem (CSS) provides the function for Z platforms to communicate with external I/O and network devices and manage the flow of data between those external devices and system memory. This goal is achieved by using a system assist processor (SAP) that connects the CSS to the external devices.

The SAP uses the I/O configuration definitions loaded in the hardware system area (HSA) of the system to identify the external devices and the protocol they support. The SAP also monitors the queue of I/O operations passed to the CSS by the operating system.

Using an SAP, the processing units (PUs) are relieved of the task of communicating directly with the devices, so data processing can proceed concurrently with I/O processing.

Increased system performance demands higher I/O and network bandwidth, speed, and flexibility, so the CSS evolved along with the advances in scalability of the Z platforms. The z/Architecture provides functions for scalability in the form of multiple CSSes that can be configured within the same IBM Z platform, for example:

- IBM z14 and z13 support up to six CSSes
- IBM z14 ZR1 and IBM z13s support up to three CSSes
- IBM zEnterprise EC12 (zEC12) supports up to four CSSes
- IBM zEnterprise BC12 (zBC12) supports up to two CSSes

All of these Z platforms deliver a significant increase in I/O throughput. For more information, see Chapter 2, “Channel subsystem overview” on page 15.

1.1.1 I/O hardware infrastructure

I/O hardware connectivity is implemented through several I/O features, some of which support the Peripheral Component Interconnect Express (PCIe) Gen2 and Gen 3\(^1\) standards. They are housed in PCIe I/O drawers or PCIe+ I/O Drawers, whereas others are housed in either I/O drawers or I/O cages.

In an I/O cage, features are installed in I/O slots, with four I/O slots forming an I/O domain. Each I/O domain uses an InfiniBand Multiplexer (IFB-MP) in the I/O cage and a copper cable that is connected to a Host Channel Adapter (HCA-C) fanout on the front of the book. The interconnection speed is 6 GBps. Each I/O cage supports up to seven I/O domains, for a total of 28 I/O slots. An eighth IFB-MP is installed to provide an alternative path to I/O features in slots 29, 30, 31, and 32 in case of a failure in domain 7.

\(^1\) PCIe Gen3 is not available for the Z family earlier than IBM z13 and z13s. Previously, IBM zEnterprise Systems used PCIe Gen2.
I/O drawers provide more I/O granularity and capacity flexibility than with the I/O cage. The I/O drawer supports two I/O domains, for a total of eight I/O slots, with four I/O features in each I/O domain. Each I/O domain uses an IFB-MP in the I/O drawer and a copper cable to connect to an HCA-C fanout in the CPC drawer. The interconnection speed is 6 Gbps. I/O drawers simplify planning because they can be added concurrently and removed in the field, which is an advantage over I/O cages. I/O drawers were first offered with the IBM z10™ BC.

The PCIe Gen 3 feature can connect to the following drawers:

- PCIe I/O Drawer (FC #4013 and FC #4032)
  PCIe I/O drawers allow a higher number of features (four times more than the I/O drawer and a 14% increase over the I/O cage) and increased port granularity. Each PCIe I/O drawer can accommodate up to 32 features in any combination. They are organized in four hardware domains per drawer, with eight features per domain. The PCIe I/O drawer is attached to a PCIe fanout in the CPC drawer, with an interconnection speed of 8 Gbps with PCIe Gen2 and 16 Gbps with PCIe Gen3. PCIe I/O drawers can be installed and repaired concurrently in the field.

- PCIe+ I/O Drawer (FC #4001)
  With IBM z14 Model ZR1 a new PCIe+ I/O Drawer (FC #4001) has been introduced. The PCIe+ I/O Drawer is fitted in a 19" Rack and allows the installation of 16 PCIe features (features that can be installed in the PCIe I/O Drawers as well). The PCIe+ I/O drawer is attached to a PCIe Gen3 fanout in the CPC drawer, with an interconnection speed of 16 Gbps. PCIe+ I/O drawers can be installed and repaired concurrently in the field.

### 1.1.2 I/O connectivity features

The most common attachment to a Z server I/O channel is a storage control unit (CU), which can be accessed through a Fibre Connection (IBM FICON) channel, for example. The CU controls I/O devices such as disk and tape drives.

System-to-system communications are typically implemented by using the IBM Integrated Coupling Adapter (ICA SR), Coupling Express Long Reach (CE LR), InfiniBand (IFB) coupling links, Shared Memory Communications, and FICON channel-to-channel (FCTC) connections.

The Internal Coupling (IC) channel, IBM HiperSockets™, and Shared Memory Communications can be used for communications between logical partitions within the Z platform.

The Open Systems Adapter (OSA) features provide direct, industry-standard Ethernet connectivity and communications in a networking infrastructure.

The 25GbE RoCE Express2, 10GbE RoCE Express2 and 10GbE RoCE Express features provide high-speed, low-latency networking fabric for IBM z/OS-to-z/OS shared memory communications.

As part of system planning activities, decisions are made about where to locate the equipment (for distance reasons), how it will be operated and managed, and the business continuity requirements for disaster recovery, tape vaulting, and so on. The types of software (operating systems and applications) that will be used must support the features and devices on the Z platform.

---

2 The PCIe+ I/O drawer is available starting with the IBM z14 ZR1. Concurrent add of a PCIe+ I/O drawer depends on the z14 ZR1 CPC drawer feature (configuration).
From a hardware point of view, all of the features in the PCIe I/O drawers, I/O drawers, and I/O cages are supported by the Z Support Elements. This functionality applies to installing and updating Licensed Internal Code (LIC) to features and other operational tasks.

Many features have an integrated processor that handles the adaptation layer functions required to present the necessary features to the rest of the system in a uniform manner. Therefore, all the operating systems have the same interface with the I/O subsystem.

The Z platform supports industry-standard PCIe adapters called native PCIe adapters. For native PCIe adapter features, there is no adaptation layer, but the device driver is present in the operating system. The adapter management functions (such as diagnostics and firmware updates) are provided by Resource Groups.

There are four Resource Groups on z14, z14 ZR1, z13\(^3\), and z13s and there are two Resource Groups zEC12, and zBC12. The Resource Groups are managed by an integrated firmware processor that is part of the system’s base configuration.

The following sections briefly describe connectivity options for the I/O features available on the Z platforms.

### 1.2 FICON Express

The Fibre Channel connection (FICON) Express features were originally designed to provide access to extended count key data (IBM ECKD\(^\text{tm}\)) devices, and FICON channel-to-channel (CTC) connectivity. Then came support for access to Small Computer System Interface (SCSI) devices. This innovation was followed by support for High-Performance FICON for z Systems (zHPF) for OLTP I/O workloads that transfer small blocks of fixed-size data. These OLTP I/O workloads include IBM Db2\(^\circledR\) database, Virtual Storage Access Method (VSAM), partitioned data set extended (PDSE), and IBM z/OS file system (zFS).

IBM Z platforms build on this I/O architecture by offering high-speed FICON connectivity, as shown in Figure 1-1 on page 5.

---

\(^3\) IBM z13 must be at driver level 27 to support four Resource Groups (RGs). While z13 at Driver Level 22 supports only two RGs. The introduction of four RGs with z13/z13s Driver Level 27 has relaxed card availability requirements during an RG firmware update.
Chapter 1. Introduction

The FICON implementation enables full-duplex data transfer. In addition, multiple concurrent I/O operations can occur on a single FICON channel. FICON link distances can be extended by using various solutions. For more information, see 10.2.2, “FICON repeated distance solutions” on page 173.

The FICON features on Z also support full fabric connectivity for the attachment of SCSI devices by using the Fibre Channel Protocol (FCP). Software support is provided by IBM z/VM®, IBM z/VSE® (SCSI disk devices), Linux on Z, and the KVM hypervisor.

1.3 zHyperLink Express

IBM zHyperLink is a technology that provides a low-latency point-to-point connection from a z14 to an IBM DS8880. The transport protocol is defined for reading and writing ECKD data records. It provides a 5-fold reduction in I/O services time for read requests.

The zHyperLink Express feature works as native PCIe adapter and can be shared by multiple LPARs.

On the z14, the zHyperLink Express feature is installed in the PCIe I/O drawer. On the IBM DS8880 side, the fiber optic cable connects to a zHyperLink PCIe interface in I/O bay.

The zHyperLink Express feature has the same qualities of service as do all Z I/O channel features.

Figure 1-2 depicts a point-to-point zHyperLink connection with a z14.
1.4 Open Systems Adapter-Express

The Open Systems Adapter-Express (OSA-Express) features are the pipeline through which data is exchanged between the Z platforms and devices in the network. OSA-Express3, OSA-Express4S, OSA-Express5S, and OSA-Express6S features provide direct, industry-standard LAN connectivity in a networking infrastructure, as shown in Figure 1-3.

The OSA-Express features bring the strengths of the Z family, such as security, availability, and enterprise-wide access to data to the LAN environment. OSA-Express provides connectivity for the following LAN types:

- 1000BASE-T Ethernet (10/100/1000 Mbps)
- 1 Gbps Ethernet
- 10 Gbps Ethernet
- 25 Gbps Ethernet
1.5 HiperSockets

IBM HiperSockets technology provides seamless network connectivity to consolidate servers in an advanced infrastructure intraserver network. HiperSockets creates multiple independent, integrated, virtual LANs within an IBM Z platform.

This technology provides high-speed connectivity between combinations of logical partitions or virtual servers. It eliminates the need for any physical cabling or external networking connection between these virtual servers. This network within the box concept minimizes network latency and maximizes bandwidth capabilities between z/VM, Linux on IBM Z and the KVM hypervisor, IBM z/VSE, and IBM z/OS images, or combinations of these. HiperSockets use is also possible under the IBM z/VM operating system, which enables establishing internal networks between guest operating systems, such as multiple Linux servers.

The z/VM virtual switch can transparently bridge a guest virtual machine network connection on a HiperSockets LAN segment. This bridge allows a single HiperSockets guest virtual machine network connection to directly communicate with other guest virtual machines on the virtual switch and with external network hosts through the virtual switch OSA UPLINK port.

Figure 1-4 shows an example of HiperSockets connectivity with multiple logical partitions (LPARs) and virtual servers.

Figure 1-4   HiperSockets connectivity with multiple LPARs and virtual servers

HiperSockets technology is implemented by Z LIC, with the communication path in system memory and the transfer information between the virtual servers at memory speed.
1.6 Parallel Sysplex and coupling links

IBM Parallel Sysplex is a clustering technology that represents a synergy between hardware and software. It consists of the following components:

- Parallel Sysplex-capable servers
- A coupling facility
- Coupling links (CS5, CL5, IFB, IC)
- Server Time Protocol (STP)
- A shared direct access storage device (DASD)
- Software, both system and subsystem

These components are all designed for parallel processing, as shown in Figure 1-5.

Parallel Sysplex cluster technology is a highly advanced, clustered, commercial processing system. It supports high-performance, multisystem, read/write data sharing, which enables the aggregate capacity of multiple z/OS systems to be applied against common workloads.

The systems in a Parallel Sysplex configuration are linked and can fully share devices and run the same applications. This feature enables you to harness the power of multiple IBM Z platforms as though they are a single logical computing system.

The architecture is centered around the implementation of a coupling facility (CF) that runs the coupling facility control code (CFCC) and high-speed coupling connections for intersystem and intrasystem communications. The CF provides high-speed data sharing with data integrity across multiple Z platforms.
Chapter 1. Introduction

Parallel Sysplex technology provides high availability for business-critical applications. The design is further enhanced with the introduction of System-Managed Coupling Facility Structure Duplexing, which provides the following additional benefits:

- **Availability**: Structures do not need to be rebuilt if a coupling facility fails.
- **Manageability and usability**: A consistent procedure is established to manage structure recovery across users.
- **Configuration benefits**: A sysplex can be configured with internal CFs only.

**Attention**: Parallel Sysplex technology supports connectivity between systems that differ by up to two generations (n-2). For example, an IBM z14 can participate in an IBM Parallel Sysplex cluster with z13, z13s, zEC12, and zBC12 platforms.

**However**, the IBM z14 ZR1 does not support InfiniBand connectivity, as such, an IBM z14 ZR1 does support connectivity to only one generation back (n-1): z14 ZR1 can connect to z14, z13, and z13s, using Integrated Coupling Adapter Short Reach (ICA SR) and Coupling Express Long Reach (CE LR) features.

### 1.7 Shared Memory Communications

Shared Memory Communications (SMC) on Z platforms is a technology that can improve throughput by accessing data faster with less latency. SMC reduces CPU resource consumption compared to traditional TCP/IP communications. Furthermore, applications do not need to be modified to gain the performance benefits of SMC.

SMC allows two peers to send and receive data by using system memory buffers that each peer allocates for its partner’s use. Two types of SMC protocols are available on the Z platform:

- **SMC-Remote Direct Memory Access (SMC-R)**

  SMC-R is a protocol for Remote Direct Memory Access (RDMA) communication between TCP socket endpoints in LPARs in different systems. SMC-R runs over networks that support RDMA over Converged Ethernet (RoCE). It allows existing TCP applications to benefit from RDMA without requiring modifications. SMC-R provides dynamic discovery of the RDMA capabilities of TCP peers and automatic setup of RDMA connections that those peers can use.

  The 25GbE RoCE Express2, 10GbE RoCE Express2, and 10GbE RoCE Express features provide the RoCE support needed for LPAR-to-LPAR communication across Z platforms.

- **SMC-Direct Memory Access (SMC-D)**

  SMC-D implements the same SMC protocol that is used with SMC-R to provide highly optimized intra-system communications. Where SMC-R uses RoCE for communicating between TCP socket endpoints in separate systems, SMC-D uses Internal Shared Memory (ISM) technology for communicating between TCP socket endpoints in the same CPC.

  ISM provides adapter virtualization (virtual functions (VFs)) to facilitate the intra-system communications. Hence, SMC-D does not require any additional physical hardware (no adapters, switches, fabric management, or PCIe infrastructure). Therefore, significant cost savings can be achieved when using the ISM for LPAR-to-LPAR communication within the same Z platform.
Both SMC protocols use shared memory architectural concepts, eliminating TCP/IP processing in the data path, yet preserving TCP/IP quality of service (QoS) for connection management purposes.

Figure 1-6 shows the connectivity for SMC-D and SMC-R configurations.

![Figure 1-6 Connectivity for SMC-D and SMC-R configurations](image)

### 1.8 I/O feature support

Table 1-1 lists the I/O features that are available on Z platforms. Not all I/O features can be ordered on all systems, and certain features are available only for a system upgrade. Depending on the type and version of Z, there might be further restrictions (for example, on the maximum number of supported ports).

**Table 1-1 IBM Z I/O features**

<table>
<thead>
<tr>
<th>I/O feature</th>
<th>Feature code</th>
<th>Maximum number of ports</th>
<th>Ports per feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>zHyperLink Express</td>
<td>0431</td>
<td>16 16 N/A N/A N/A N/A 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express16S+ LX</td>
<td>0427</td>
<td>320 128 320 128 N/A N/A 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S+ SX</td>
<td>0428</td>
<td>320 128 320 128 N/A N/A 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S LX</td>
<td>0418</td>
<td>320 128 320 128 N/A N/A 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S SX</td>
<td>0419</td>
<td>320 128 320 128 N/A N/A 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express8S 10 KM LX</td>
<td>0409</td>
<td>320 128 320 128 128 128 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express8S SX</td>
<td>0410</td>
<td>320 128 320 128 128 128 2</td>
<td></td>
</tr>
<tr>
<td>FICON Express8 10 KM LX</td>
<td>3325</td>
<td>N/A N/A 64 64 176 32/64 a 4</td>
<td></td>
</tr>
<tr>
<td>FICON Express8 SX</td>
<td>3326</td>
<td>N/A N/A 64 64 176 32/64 a 4</td>
<td></td>
</tr>
<tr>
<td>FICON Express4 10 KM LX</td>
<td>3321</td>
<td>N/A N/A N/A N/A N/A 176 32/64 a 4</td>
<td></td>
</tr>
<tr>
<td>FICON Express4 SX</td>
<td>3322</td>
<td>N/A N/A N/A N/A 176 32/64 a 4</td>
<td></td>
</tr>
<tr>
<td>I/O feature</td>
<td>Feature code</td>
<td>Maximum number of ports</td>
<td>Ports per feature</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>z14</td>
<td>z14 ZR1</td>
</tr>
<tr>
<td>FICON Express4-2C SX</td>
<td>3318</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S 25GbE SR</td>
<td>0429</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE LR</td>
<td>0424</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE SR</td>
<td>0425</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE SR</td>
<td>0415</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE SR</td>
<td>0416</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE LR</td>
<td>0406</td>
<td>N/A</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE SR</td>
<td>0407</td>
<td>N/A</td>
<td>48</td>
</tr>
<tr>
<td>OSA-Express3 10 GbE LR</td>
<td>3370</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express3 10 GbE SR</td>
<td>3371</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express6S GbE LX</td>
<td>0422</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express6S GbE SX</td>
<td>0423</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express5S GbE LX</td>
<td>0413</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express5S GbE SX</td>
<td>0414</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express4S GbE LX</td>
<td>0404</td>
<td>N/A</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express4S GbE SX</td>
<td>0405</td>
<td>N/A</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express3 GbE LX</td>
<td>3362</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express3 GbE SX</td>
<td>3363</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express3-2P GbE SX</td>
<td>3373</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express5S 1000BASE-T</td>
<td>0417</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express6S 1000BASE-T</td>
<td>0426</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>OSA-Express4S 1000BASE-T Ethernet</td>
<td>0408</td>
<td>96</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express3 1000BASE-T Ethernet</td>
<td>3367</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OSA-Express3-2P 1000BASE-T Ethernet</td>
<td>3369</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>RoCE Express</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25GbE RoCE Express2</td>
<td>0430</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>10GbE RoCE Express2</td>
<td>0412</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>10GbE RoCE Express</td>
<td>0411</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>HiperSockets</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter 1. Introduction  

Chapter 5, “Open Systems Adapter-Express” on page 71  
Chapter 6, “OSA-Express Console support” on page 119  
Chapter 7, “Shared Memory Communications” on page 125  
Chapter 8, “HiperSockets technology” on page 135
In addition to the I/O connectivity features, several special purpose features are available that can be installed in the PCIe I/O drawers, such as:

- Crypto Express
- Flash Express
- zEDC Express

1.9.1 Crypto Express features

Integrated cryptographic features provide leading cryptographic performance and functions. Reliability, availability, and serviceability support is unmatched in the industry, and the cryptographic solution received the highest standardized security certification (FIPS 140-2 Level 4).

Crypto Express6S, Crypto Express5S, and Crypto Express4S are tamper-sensing and tamper-responding programmable cryptographic features that provide a secure cryptographic environment. Each adapter contains a tamper-resistant hardware security module (HSM).
The HSM can be configured as a secure IBM Common Cryptographic Architecture (CCA) coprocessor, as a secure IBM Enterprise PKCS #11 (EP11) coprocessor, or as an accelerator.

### 1.9.2 Flash Express feature

Flash Express is an innovative optional feature that was introduced with the zEC12 and also available on zBC12, z13, and z13s platforms. It improves performance and availability for critical business workloads that cannot afford any reduction in service levels. Flash Express is easy to configure, requires no special skills, and produces rapid time to value.

Flash Express implements storage-class memory (SCM) through an internal NAND Flash solid-state drive (SSD) in a PCIe adapter form factor. Each Flash Express feature is installed exclusively in the PCIe I/O drawer and occupies one I/O slot.

For availability reasons, the Flash Express feature must be ordered in pairs. A feature pair provides 1.4 TB of usable storage. A maximum of four pairs (4 x 1.4 = 5.6 TB) is supported in a system. One PCIe I/O drawer supports up to two Flash Express pairs.

Flash Express storage is allocated to each partition in a manner similar to main memory allocation. The allocation is specified at the Hardware Management Console (HMC). z/OS can use the Flash Express feature as SCM for paging store to minimize the supervisor call (SVC) memory dump duration.

The z/OS paging subsystem supports a mix of Flash Express and external disk storage. Flash Express can also be used by Coupling Facility images to provide extra capacity for particular structures (for example, IBM MQ shared queues application structures).

**Replacement of Flash Express**

- **Note:** Starting with the IBM z14, Flash Express (FC #0402 and FC 0403) was replaced by Virtual Flash memory (VFM). VFM implements Extended Asynchronous Data Mover (EADM) architecture by using HSA-like memory instead of Flash card pairs.

- The VFM features come in two sizes:
  - For z14, one VFM feature is 1.5 TB (FC #0604). Up to four features per system can be ordered.
  - For z14 ZR1, one VFM feature is 512 GB (FC #0614). Up to four features per system can be ordered.

### 1.9.3 zEDC Express feature

zEDC Express is an optional feature that is supported on the Z platform. It provides hardware-based acceleration of data compression and decompression. That capability improves cross-platform data exchange, reduces processor use, and saves disk space.

A minimum of one feature can be ordered, and a maximum of 16\(^4\) can be installed on the system on the PCIe I/O drawer. Up to two zEDC Express features per domain can be installed. There is one PCIe adapter/compression coprocessor per feature, which implements compression as defined by RFC1951 (DEFLATE). For more information about the DEFLATE Compress Data Format Specification, see the [RFC for DEFLATE Compressed Data Format Specification Version 1.3](https://tools.ietf.org/html/rfc1951).

\(^4\) See Table 1-2 on page 14 for maximum features supported per system.
A zEDC Express feature can be shared by up to 15 LPARs. z/OS version 2.1 and later supports the zEDC Express feature. z/VM 6.3 with PTFs and later also provides guest exploitation.

Table 1-2 lists the special-purpose features that are available on Z platforms. Not all special-purpose features can be ordered on all systems, and certain features are available only with a system upgrade. Except for the Crypto Express3 features, all special-purpose features are installed in the PCIe I/O drawer or the PCIe+ I/O drawer.

Table 1-2   IBM Z special-purpose features

<table>
<thead>
<tr>
<th>Special-purpose feature</th>
<th>Feature codes</th>
<th>Maximum number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>z14</td>
</tr>
<tr>
<td><strong>Crypto Express</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crypto Express6S</td>
<td>0893</td>
<td>16</td>
</tr>
<tr>
<td>Crypto Express5S</td>
<td>0890</td>
<td>16(^a)</td>
</tr>
<tr>
<td>Crypto Express4S</td>
<td>0865</td>
<td>N/A</td>
</tr>
<tr>
<td>Crypto Express3</td>
<td>0864</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Flash Express</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flash Express(^b)</td>
<td>0402</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>0403(^c)</td>
<td></td>
</tr>
<tr>
<td><strong>zEDC Express</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zEDC Express</td>
<td>0420</td>
<td>16</td>
</tr>
</tbody>
</table>

\(^a\) Carry forward during an upgrade only.

\(^b\) Features are ordered in pairs.

\(^c\) Available on the z13 and z13s (new built).

\(^d\) Starting with z14, the Flash Express functionality has been replaced by Virtual Flash Memory (zVFM) feature (real memory backed) - see “Replacement of Flash Express” on page 13.

---

5 PCIe+ I/O drawer (FC #4001) has been introduced with the IBM z14 Model ZR1 server for PCIe I/O features. It is manufactured to fit in the 19” rack, occupying 8U. It cannot be ordered for prior systems. Also, the PCIe I/O drawer cannot be carried forward during an upgrade or MES.
Channel subsystem overview

This chapter describes the channel subsystem (CSS), which handles all the system input/output (I/O) operations for IBM Z platforms. The role of the CSS is to control communication between internal or external channels and control units and devices.

This chapter includes the following sections:

- CSS description
- I/O configuration management
- I/O configuration planning
- References
2.1 CSS description

CSS enables communication from system memory to peripherals by using channel connections. The channels in the CSS allow transfer of data between memory and I/O devices or other servers under the control of a channel program. The CSS allows channel I/O operations to continue independently of other operations in the system, which allows other functions to resume after an I/O operation is initiated. The CSS also provides internal channels for communication between logical partitions in a physical system.

2.1.1 CSS elements

CSS includes the elements that are described in this subsection.

**Channel path**
A channel path is a single interface between a system and one or more control units. Commands and data are sent across a channel path to process I/O requests. A CSS can have up to 256 channel paths.

**Subchannels**
A subchannel provides the logical representation of a device to the program and contains the information that is required to sustain a single I/O operation. One subchannel is assigned per each device that is defined to the logical partition. Subchannel set availability per platform is shown in the following list:
- Four subchannel sets are available on IBM z14 and IBM z13 systems.
- Three subchannel sets are available on IBM z14 ZR1, IBM z13s and IBM zEnterprise EC12 (zEC12).
- Two subchannel sets are available on IBM zEnterprise BC12 (zBC12).

**Channel path identifier**
The channel subsystem communicates with I/O devices through channel paths between the channel subsystem and control units. Each system channel path is assigned a channel path identifier (CHPID) value that uniquely identifies that path. CSS supports a total of 256 CHPIDs.

On IBM Z, a CHPID number is assigned to a physical location (slot or port) by the user through either the HCD or the IOCP.

**Control units**
A control unit provides the logical capabilities that are necessary to operate and control an I/O device. It adapts the characteristics of each device so that it can respond to the standard form of control that is provided by the CSS. A control unit can be housed separately, or it can be physically and logically integrated with the I/O device, the channel subsystem, or in the system itself.

**I/O devices**
An I/O device provides external storage, or a means of communication between data processing systems, or a means of communication between a system and its environment. In the simplest case, an I/O device is attached to one control unit and is accessible through one channel path.
2.1.2 Multiple CSS concept

The IBM Z design offers considerable processing power, memory size, and I/O connectivity. The CSS concept has been scaled up to support the larger I/O capability. IBM Z implements the multiple CSS concept, which provides relief for the number of supported logical partitions, channels, and devices that are available to the system.

Each CSS can have up to 256 channels and can be configured to a maximum of 15 logical partitions\(^1\). Table 2-1 lists the maximum number of CSSes and logical partitions that are supported by Z platforms. CSSes are numbered 0 - 5, with the numbers referred to as **CSS image IDs**.

<table>
<thead>
<tr>
<th>Systems</th>
<th>Maximum number of CSSes</th>
<th>Maximum number of logical partitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>z14</td>
<td>6</td>
<td>85</td>
</tr>
<tr>
<td>z14 ZR1</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>z13</td>
<td>6</td>
<td>85</td>
</tr>
<tr>
<td>z13s</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>zEC12</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>zBC12</td>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>

2.1.3 Multiple CSS structure

The structure of multiple CSSes provides channel connectivity to the defined logical partitions in a manner that is apparent to subsystems and application programs. Z platforms enable you to define more than 256 CHPIDs in the system through the multiple CSSes. As previously noted, the CSS defines CHPIDs, control units, subchannels, and so on. This feature enables you to define a balanced configuration for the processor and I/O capabilities.

For easier management, consider using the Hardware Configuration Definitions (HCD) tool to build and control the IBM Z input/output configuration definitions. HCD support for multiple channel subsystems is available with z/VM and z/OS. HCD provides the capability to make dynamic I/O configuration changes.

Logical partitions cannot be added until at least one CSS is defined. Logical partitions are defined for a CSS, not for a system. A logical partition is associated with only one CSS. CHPID numbers are unique within a CSS, but the same CHPID number can be reused within all CSSes.

All channel subsystems are defined within a single I/O configuration data set (IOCDS). The IOCDS is loaded into the hardware system area (HSA) and initialized during power-on reset.

On the Z platform, the HSA has a fixed size, which is not included in the purchased memory. Table 2-2 lists the HSA sizes.

<table>
<thead>
<tr>
<th>Systems</th>
<th>z14</th>
<th>z14 ZR1</th>
<th>z13</th>
<th>z13s</th>
<th>zEC12</th>
<th>zBC12</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA size</td>
<td>192 GB</td>
<td>64 GB</td>
<td>96 GB</td>
<td>40 GB</td>
<td>32 GB</td>
<td>16 GB</td>
</tr>
</tbody>
</table>

\(^1\) The z14 and z13 support 10 logical partitions in the sixth CSS, while z14 ZR1 and z13s support 10 logical partitions in the third CSS.
Figure 2-1 shows a logical view of these relationships. Each CSS supports up to 15 logical partitions.

<table>
<thead>
<tr>
<th>CSS 0</th>
<th>CSS 1</th>
<th>CSS 2</th>
<th>CSS 3</th>
<th>CSS 4</th>
<th>CSS 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>up to 256 CHPIDs</td>
<td>up to 256 CHPIDs</td>
<td>up to 256 CHPIDs</td>
<td>up to 256 CHPIDs</td>
<td>up to 256 CHPIDs</td>
<td>up to 256 CHPIDs</td>
</tr>
</tbody>
</table>

Logical Partitions

The multiple image facility (MIF) system function enables resource sharing across logical partitions in a single CSS or across multiple CSSes. Sharing a channel across logical partitions in different CSSes is called spanning. See 2.1.8, “Channel spanning” on page 24 for more information.

With multiple CSSes, the IOCDS logical partition MIF image ID is not unique. Therefore, the logical partition identifier value provides a unique value for each logical partition in the same system.

**Logical partition identifier**
The logical partition identifier is a number in the following ranges:

- 00 to 55 (hex) for the z14 and the z13
- 00 to 3F (hex) for the zEC12
- 00 to 28 (hex) for z14 ZR1 and z13s
- 00 to 1F (hex) for the zBC12.

It is assigned in the image profile through the Support Element (SE) or the Hardware Management Console (HMC), and it is unique across the system. It is also referred to as the *user logical partition ID* (or *user LPAR ID*).

**Logical partition name**
The logical partition name is defined through the HCD or IOCP and is the name in the RESOURCE statement in IOCP. The logical partition names must be unique across all CSSes.

**MIF ID**
The MIF ID identifier is defined through HCD or IOCP and is the number that is defined in the RESOURCE statement in IOCP as follows:

- The *MIF ID* is in the range of 1 to F for the zEC12 and zBC12 (all CSSes).
- For z14 and z13, the MIF ID is in the range of 1 to F for CSSes 1 - 5, and from 1 to A for the sixth CSS.

---

2 The sixth CSS (CSS5) on z14 and z13 supports 10 LPARs. Same rule applies for third CSS (CSS2) on z14 ZR1 and z13s.
For z14 ZR1 and z13s, the MIF ID ranges from 1 to F for CSSes 1 and 2, and from 1 to A for the third CSS.

The MIF is unique within a CSS. It is not unique across multiple CSSes because multiple CSSes can specify the same MIF ID, also known as the *image ID*.

You must establish a numbering convention for the logical partition (LPAR) identifiers. As shown in Figure 2-2, use the CSS number concatenated to the MIF image ID, which means that logical partition ID 3A is in CSS 3 with MIF ID A. This number fits within the allowed range of logical partition IDs and conveys information that is useful to system administrators and operators.

![Figure 2-2 Logical partition and MIF definitions](image)

**Dynamic addition or deletion of a logical partition name**

Because the amount of memory reserved for the HSA is a predefined fixed value, all possible CSSes and LPARs are reserved automatically. They can be added or deleted dynamically by using HCD activations to first add the partition and then, through a subsequent HCD activation, to define and add the connectivity resources to that partition.

### 2.1.5 Physical channel ID

A physical channel ID (PCHID) reflects the physical location of a channel-type interface. A PCHID number is based on the PCIe I/O drawer, I/O drawer, or I/O cage location, the channel feature slot number, and the port number of the channel feature. A CHPID does not directly correspond to a hardware channel port, but is assigned to a PCHID in HCD or IOCP.

CHPIDs are not preassigned on Z platforms. You must assign the CHPID numbers by using the CHPID Mapping Tool or directly by using HCD/IOCP. Assigning CHPIDs means that the CHPID number is associated with a physical channel port location (PCHID) or the adapter ID (AID) and a CSS. The CHPID number range is from 00 to FF and must be unique in a CSS. Any CHPID that is not connected to a PCHID fails validation when an attempt is made to build a production IODF or an IOCDS.
Figure 2-3 shows the front view of the PCIe I/O drawer 4 (bottom of the frame), including I/O features in slots 01 - 04, and the PCHID numbers.

Example 2-1 shows a portion of a sample PCHID REPORT for an IBM z13 system.

Example 2-1   PCHID REPORT for a z13

<table>
<thead>
<tr>
<th>Source</th>
<th>Cage</th>
<th>Slot</th>
<th>F/C</th>
<th>PCHID/Ports or AID</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A27/LG07</td>
<td>A27A</td>
<td>LG07</td>
<td>0170</td>
<td>AID=00</td>
<td></td>
</tr>
<tr>
<td>A15/LG04</td>
<td>A15A</td>
<td>LG04</td>
<td>0172</td>
<td>AID=30</td>
<td></td>
</tr>
<tr>
<td>A19/LG15/J01</td>
<td>Z22B</td>
<td>03</td>
<td>0414</td>
<td>108/D1D2</td>
<td></td>
</tr>
<tr>
<td>A19/LG15/J01</td>
<td>Z22B</td>
<td>04</td>
<td>0418</td>
<td>10C/D1 10D/D2</td>
<td>RG2</td>
</tr>
<tr>
<td>A19/LG03/J01</td>
<td>Z08B</td>
<td>20</td>
<td>0411</td>
<td>240/D1D2</td>
<td>RG1</td>
</tr>
<tr>
<td>A27/LG14/J01</td>
<td>Z01B</td>
<td>03</td>
<td>0420</td>
<td>288</td>
<td>RG2</td>
</tr>
<tr>
<td>A23/LG14/J01</td>
<td>Z01B</td>
<td>19</td>
<td>0411</td>
<td>2BC/D1D2</td>
<td>RG1</td>
</tr>
<tr>
<td>A15/LG05/J01</td>
<td>A32B</td>
<td>38</td>
<td>0420</td>
<td>37C</td>
<td></td>
</tr>
</tbody>
</table>

Legend:
Source  Book Slot/Fanout Slot/Jack
A15A    Processor Drawer 1 in A frame
A19A    Processor Drawer 2 in A frame
A23A    Processor Drawer 3 in A frame
A27A    Processor Drawer 4 in A frame
Z22B    PCIe Drawer 1 in Z frame
Z15B    PCIe Drawer 2 in Z frame
Z08B    PCIe Drawer 3 in Z frame
Z01B    PCIe Drawer 4 in Z frame
A32B    PCIe Drawer 5 in A frame
RG1     Resource Group One
RG2     Resource Group Two
0170    HCA3 0 LR PSIFB 1x 4 Links
The following list explains the content of the sample PCHID REPORT:

- Feature code 0170 (HCA3-O LR [1x IFB]) is installed in CPC drawer 4 (location A27A, slot LG07) and has AID 00 assigned.
- Feature code 0172 (Integrated Coupling Adapter [ICA SR]) is installed in CPC drawer 1 (location A15A, slot LG04) and has AID 30 assigned.
- Feature code 0414 (OSA-Express5S GbE short wavelength [SX]) is installed in PCIe I/O drawer 2 (location Z22B, slot 03) and has PCHID 108 assigned. PCHID 108 is shared by ports D0 and D1.
- Feature code 0418 (FICON Express16S long wavelength [LX] 10 km [6.2 miles]) is installed in PCIe I/O drawer 2 (location Z22B, slot 04) and has PCHIDs 10C and 10D assigned.
- Feature code 0411 (10GbE RoCE Express) is installed in PCIe I/O drawer 3 (location Z08B, slot 20) and has PCHID 240 assigned. PCHID 240 is shared by ports D0 and D1.

A resource group (RG) parameter is shown in the PCHID REPORT for native PCIe features. There is a balanced plugging of native PCIe features between two resource groups (RG1 and RG2).

You can get a PCHID report from the IBM account team when ordering.

### 2.1.6 Adapter ID

The adapter ID (AID) number assigns a CHPID to a port by using HCD/IOCP for IBM Parallel Sysplex cluster technology.

On the Z platform, the AID is bound to the serial number of the fanout. If the fanout is moved, the AID moves with it. No IOCDS update is required if adapters are moved to a new physical location.

Table 2-3 lists the assigned AID numbers for a zEC12 new build.

<table>
<thead>
<tr>
<th>Fanout location</th>
<th>Fourth</th>
<th>First</th>
<th>Third</th>
<th>Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>00</td>
<td>08</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>D2</td>
<td>01</td>
<td>09</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>D3</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>D4</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>D5</td>
<td>02</td>
<td>0A</td>
<td>12</td>
<td>1A</td>
</tr>
<tr>
<td>D6</td>
<td>03</td>
<td>0B</td>
<td>13</td>
<td>1B</td>
</tr>
<tr>
<td>D7</td>
<td>04</td>
<td>0C</td>
<td>14</td>
<td>1C</td>
</tr>
</tbody>
</table>
Table 2-4 lists the assigned AID numbers for a zBC12 new build.

**Table 2-4  AID number assignment for zBC12**

<table>
<thead>
<tr>
<th>Fanout location</th>
<th>Processor drawer</th>
<th>Processor drawer 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>08</td>
<td>00</td>
</tr>
<tr>
<td>D2</td>
<td>09</td>
<td>01</td>
</tr>
<tr>
<td>D7</td>
<td>0A</td>
<td>02</td>
</tr>
<tr>
<td>D8</td>
<td>0B</td>
<td>03</td>
</tr>
</tbody>
</table>

Table 2-5 lists the assigned AID numbers for a z13 new build.

**Table 2-5  Fanout AID numbers for z13**

<table>
<thead>
<tr>
<th>Drawer (DX)b</th>
<th>Location</th>
<th>Fanout slot</th>
<th>AIDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (D3) a</td>
<td>A15A</td>
<td>LG02-LG06 (PCIe)</td>
<td>2E-32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG07-LG10 (IFB)</td>
<td>0C-0F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG11-LG15 (PCIe)</td>
<td>33-37</td>
</tr>
<tr>
<td>Second (D2) b</td>
<td>A19A</td>
<td>LG02-LG06 (PCIe)</td>
<td>24-28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG07-LG10 (IFB)</td>
<td>08-0B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG11-LG15 (PCIe)</td>
<td>29-2D</td>
</tr>
<tr>
<td>Third (D1)</td>
<td>A23A</td>
<td>LG02-LG06 (PCIe)</td>
<td>1A-1E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG07-LG10 (IFB)</td>
<td>04-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG11-LG15 (PCIe)</td>
<td>1F-23</td>
</tr>
<tr>
<td>Fourth (D0)</td>
<td>A27A</td>
<td>LG02-LG06 (PCIe)</td>
<td>10-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG07-LG10 (IFB)</td>
<td>00-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG11-LG15 (PCIe)</td>
<td>15-19</td>
</tr>
</tbody>
</table>

a. Indicates the z13 physical CPC drawer installation order.
b. The designation between the parenthesis indicates the logical CPC drawer number.

AIDs are included in the PCHID report that IBM provides for new build systems and upgrades. Example 2-2 shows an AID in a PCHID report.

**Example 2-2  AID assignment in z13 PCHID report**

CHPIDSTART
12345678       PCHID REPORT     Oct 31,2014
Machine: 2964-NE1 SNXXXXXXXX
Chapter 2. Channel subsystem overview

<table>
<thead>
<tr>
<th>Source</th>
<th>Cage</th>
<th>Slot</th>
<th>F/C</th>
<th>PCHID/Ports or AID</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A15/LG02</td>
<td>A15A</td>
<td>LG02</td>
<td>0172</td>
<td>AID=2E</td>
<td></td>
</tr>
<tr>
<td>A19/LG11</td>
<td>A19A</td>
<td>LG11</td>
<td>0172</td>
<td>AID=29</td>
<td></td>
</tr>
</tbody>
</table>

For more information, see 9.2.6, “InfiniBand coupling links” on page 161.

Table 2-6 lists the AID assignment for each fanout slot relative to the drawer location on a z13s new build system.

Table 2-6  AID number assignment for IBM z13s

<table>
<thead>
<tr>
<th>Drawer</th>
<th>Location</th>
<th>Fanout slot</th>
<th>AIDs</th>
</tr>
</thead>
</table>
| First    | A21A     | LG03-LG06 (PCIe)
          |           | LG07-LG10 (IFB)
          |           | LG11-LG14 (PCIe) | 1B-1E  |
|          |          |               | 04-07 |
|          |          |               | 1F-22 |
| Second   | A25A     | LG03-LG06 (PCIe) | 11-14 |
|          |          | LG07-LG10 (IFB) | 00-03 |
|          |          | LG11-LG14 (PCIe) | 15-18 |

a. For model N10, only the first drawer’s LG09-LG10 supported for IFB, and LG11-LG14 supported for PCIe

Table 2-7 lists the assigned AID numbers for a z14 new build system.

Table 2-7  Fanout AID numbers for z14

<table>
<thead>
<tr>
<th>Drawer (DX)b</th>
<th>Location</th>
<th>Fanout slot</th>
<th>AIDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (D3)</td>
<td>A15A</td>
<td>LG03-LG12 (PCIe)</td>
<td>2E-37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG13-LG16 (IFB)</td>
<td>0C-0F</td>
</tr>
<tr>
<td>Second (D2)</td>
<td>A19A</td>
<td>LG03-LG12 (PCIe)</td>
<td>24-2D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG13-LG16 (IFB)</td>
<td>08-0B</td>
</tr>
<tr>
<td>Third (D1)</td>
<td>A23A</td>
<td>LG03-LG12 (PCIe)</td>
<td>1A-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG13-LG16 (IFB)</td>
<td>04-07</td>
</tr>
<tr>
<td>Fourth (D0)</td>
<td>A27A</td>
<td>LG03-LG12 (PCIe)</td>
<td>10-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG13-LG16 (IFB)</td>
<td>00-03</td>
</tr>
</tbody>
</table>

a. Indicates the z14 physical CPC drawer installation order.
b. The designation between the parenthesis indicates the logical CPC drawer number.

Table 2-8 lists the assigned AID numbers for a z14 ZR1 new build system.

Table 2-8  AID number assignment for z14 ZR1

<table>
<thead>
<tr>
<th>Fanout location</th>
<th>CPC Drawer Location</th>
<th>AIDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG01 - LG04</td>
<td>A09B</td>
<td>10 - 13</td>
</tr>
<tr>
<td>LG07 - LG10</td>
<td></td>
<td>14 - 17</td>
</tr>
</tbody>
</table>
2.1.7 Multiple CSS construct example

Each CSS has three logical partitions with their associated MIF image identifiers. In each CSS, the CHPIDs are shared across all logical partitions. The CHPIDs are assigned to their designated PCHIDs. Figure 2-4 shows two CSSes defined as CSS0 and CSS1.

The CHPIDs are mapped to their designated PCHIDs by using the Channel Mapping Tool or manually by using HCD or IOCP.

The output of the Channel Mapping Tool is used as input to HCD or the IOCP to establish the CHPID-to-PCHID assignments. See 2.2, “I/O configuration management” on page 28 for more information about the Channel Mapping Tool.

2.1.8 Channel spanning

Channel spanning extends the MIF concept of sharing channels across logical partitions to sharing channels across channel subsystems and logical partitions.

*Spanning* is the ability for the channel to be configured to multiple channel subsystems. When so defined, the channels can be transparently shared by any or all of the configured logical partitions, regardless of the channel subsystem to which the logical partition is configured.

A channel is considered a spanned channel if the same CHPID number in different CSSes is assigned to the same PCHID in the IOCP or is defined as *spanned* in HCD. The same applies to internal channels such as IBM HiperSockets technology, but there is no PCHID association. Internal channels are defined with the same CHPID number in multiple CSSes.
CHPIDs that span CSSes reduce the total number of channels that are available on IBM Z. The total is reduced because no CSS can have more than 256 CHPIDs. For a zBC12 that supports two CSSes, a total of 512 CHPIDs are supported. For a zEC12 with four CSSes, a total of 1024 CHPIDs are supported. For a z14 and z13 with six CSSes, a total of 1536 CHPIDs are supported. For a z14 ZR1 and z13s with three CSSes, a total of 768 CHPIDs are supported.

If all CHPIDs are spanned across the two, four, or six CSSes depending on the server, only 256 channels can be supported. Table 2-9 shows which channel types can be defined as shared or spanned channels.

<table>
<thead>
<tr>
<th>Channel type</th>
<th>CHPID definition</th>
<th>Shared channels</th>
<th>Spanned channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICON Express4</td>
<td>External FC, FCP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FICON Express8</td>
<td>External FC, FCP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FICON Express8S</td>
<td>External FC, FCP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FICON Express16S</td>
<td>External FC, FCP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FICON Express16S+</td>
<td>External FC, FCP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>zHyperLink Express</td>
<td>External N/A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OSA-Express3a</td>
<td>External OSD, OSE, OSC, OSN, OSX, OSM</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OSA-Express4Sa</td>
<td>External OSD, OSE, OSC, OSN, OSX, OSM</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OSA-Express5Sa</td>
<td>External OSD, OSE, OSC, OSN, OSX, OSM</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OSA-Express6Sa</td>
<td>External OSD, OSE, OSC, OSN, OSX, OSM</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>OSA-Express7Sb</td>
<td>External OSD, OSX</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ICA SR</td>
<td>External CS5</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CE LR</td>
<td>External CL5</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IFB</td>
<td>External CiB</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ISC-3</td>
<td>External CFP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>IC</td>
<td>Internal ICP</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HiperSocketsc</td>
<td>Internal IQD</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

a. Not every type of CHPID is supported by the different OSA-Express features. For more information, see Chapter 5, “Open Systems Adapter-Express” on page 71.

b. OSA-Express7S 25GbE SR

c. The CHPID statement of HiperSocket devices requires the keyword VCHID. VCHID specifies the virtual channel identification number associated with the channel path. The valid range is 7C0 - 7FF. VCHID is not valid on IBM Z products before z13.

---

3 The CHPID statement of HiperSocket devices requires the keyword VCHID. VCHID specifies the virtual channel identification number that is associated with the channel path. The valid range is 7C0 - 7FF. VCHID is not valid on IBM Z products before z13.
In Figure 2-5, CHPID 04 is spanned across CSS 0 and CSS 1. Because it is not an external channel link, no PCHID is assigned. CHPID 06 is a spanned external channel and has PCHID 120 assigned.

2.1.9 Multiple subchannel sets

Do not confuse the multiple subchannel sets (MSS) functions with multiple channel subsystems. In most cases, a subchannel represents an addressable device. For example, a disk control unit with 30 drives uses 30 subchannels. An addressable device is associated with a device number.

Subchannel numbers, including their implied path information to devices, are limited to four hexadecimal digits by hardware and software architectures. Four hexadecimal digits provide 64 thousand addresses, which are known as a set. IBM reserved 256 subchannels, leaving 63.75 thousand subchannels for general use with IBM Z platforms. Parallel access volumes (PAVs) make this limitation of subchannels a challenge for larger installations. A single-disk drive (with PAVs) often uses at least four subchannels.

It was difficult to remove this constraint because the use of four hexadecimal digits for subchannels and device numbers that correspond to subchannels is specified in several places. Simply expanding the field would break too many programs.

The solution allows sets of subchannels (addresses) with a current implementation of four sets with z14 and z13, and three sets with z14 ZR1, z13s, and zEC12. Each set provides 64 thousand addresses minus one. Subchannel set 0, the first set, reserves 256 subchannels for use by IBM. Subchannel set 1 provides a full range of 64 thousand minus one subchannel on the Z platform. Subchannel set 2 provides another full range of 64 thousand, minus one subchannels on z14, z14 ZR1, z13, z13s, and zEC12 systems only. Subchannel set 3, the fourth set, provides another full range of 64 thousand, minus one subchannel on z14 and z13.

---

Four subchannel sets are mostly used in conjunction with PAV. They represent base addresses and three alias addresses.
The first subchannel set (SS0) allows definitions of any type of device that is supported, for example, bases, aliases, secondaries, and devices, other than disks that do not implement the concept of associated aliases or secondaries.

The second, third, and fourth subchannel sets (SS1, SS2, and SS3) are designated for use for disk alias devices, both primary and secondary devices, and for IBM Metro Mirror secondary devices only.

There is no required correspondence between addresses in the three sets. For example, it is possible to have device number 8000 in subchannel set 0 and device number 8000 in subchannel sets 1 or 2, and they can refer to different devices. 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 that are used in the four subchannel sets. Each channel subsystem can have multiple subchannel sets, as shown in Figure 2-6.

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 sets 1, 2, or 3.

**IPL from an alternative subchannel set**

z14 and z13 support IPL from subchannel sets 0, 1, 2, and 3. z14 ZR1, z13s and zEC12 support IPL from subchannel sets 0, 1, and 2, and zBC12 supports IPL from subchannel sets 0 and 1. Devices that are used early during IPL processing can now be accessed by using subchannel set 1, subchannel set 2, or subchannel set 3.

This feature allows the users of Metro Mirror (Peer-to-Peer Remote Copy, or PPRC) secondary devices. These devices are defined by using the same device number and a new device type in an alternative subchannel set to be used for IPL, IODF, and stand-alone memory dump volumes when needed.

IPL from an alternative subchannel set is supported on the Z platform. It is supported by z/OS V2.2, V2.1, and V1.13 and V1.12 with program temporary fixes (PTFs). It applies to the IBM Fibre Channel connection (FICON) and High Performance FICON for IBM Z® (zHPF) protocols.

---

5 The number of subchannels sets in this figure applies to zEC12.
6 z/OS V1.12 is not supported on z14.
Table 2-10 lists the maximum number of CSS elements that are supported per Z platform.

<table>
<thead>
<tr>
<th>CSSes</th>
<th>zBC12</th>
<th>zEC12</th>
<th>z13s</th>
<th>z13</th>
<th>z14 ZR1</th>
<th>z14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partitions</td>
<td>2 per system</td>
<td>4 per system</td>
<td>3 per system</td>
<td>6 per system</td>
<td>3 per system</td>
<td>6 per system</td>
</tr>
<tr>
<td>CSSes</td>
<td>15 per CSS, up to 30 per system</td>
<td>15 per CSS, up to 60 per system</td>
<td>15 for the first two CSSes and 10 for the third, up to 40 per system</td>
<td>15 for the first 5 CSSes and 10 for the sixth, up to 85 per system</td>
<td>15 for the first 5 CSSes and 10 for the sixth, up to 85 per system</td>
<td></td>
</tr>
<tr>
<td>CHPIDs</td>
<td>256 per CSS, up to 512 per system</td>
<td>256 per CSS, up to 1024 per system</td>
<td>256 per CSS, up to 768 per system</td>
<td>256 per CSS, up to 1536 per system</td>
<td>256 per CSS, up to 1536 per system</td>
<td></td>
</tr>
<tr>
<td>Devices</td>
<td>63,750 on SS0; 64,000 - 1 on SS1</td>
<td>63,750 on SS0; 64,000 - 1 on SS1 and SS2</td>
<td>63,750 on SS0; 64,000 - 1 on SS1 and SS2</td>
<td>63,750 on SS0; 64,000 - 1 on SS1, SS2, and SS3</td>
<td>63,750 on SS0; 64,000 - 1 on SS1, SS2, and SS3</td>
<td></td>
</tr>
</tbody>
</table>

2.2 I/O configuration management

CSS controls communication between a configured channel, the control unit, and the device. The IOCDS defines the channels, control units, and devices to the designated logical partitions within the system. This communication is defined by using the IOCP.

The IOCP statements typically are built by using the HCD. An interactive dialog is used to generate the input/output definition file (IODF), start the IOCP program, and then build the production IOCDS. The IOCDS is loaded into the HSA and initialized during power-on reset. In prior Z servers, the HSA storage was allocated based on the size of the IOCDS, partitions, channels, control units, and devices. Additional storage was reserved for dynamic I/O reconfiguration, if enabled.

The HSA has the following fixed sizes:

- On z14, the fixed size is 192 GB.
- On z13, the fixed size is 96 GB.
- On z14 ZR1, the fixed size is 64 GB.
- On z13s, the fixed size is 40 GB.
- On zEC12, the fixed size is 32 GB.
- On zBC12, the fixed size is 16 GB.

For more information, see 2.1.3, “Multiple CSS structure” on page 17.

With Z platforms, CHPIDs are mapped to PCHIDs or AIDs by using the configuration build process through HCD or IOCP (see 2.1.6, “Adapter ID” on page 21). There is no PCHID association for internal channels (that is, IC links and IBM HiperSockets).

The sections that follow describe tools that are used to maintain and optimize the I/O configuration on Z platforms.
2.2.1 Hardware configuration definition

HCD supplies an interactive dialog to generate the IODF and then the IOCDS. Consider using the HCD to generate the IOCDS, as opposed to writing IOCP statements. The validation checking that HCD does as data is entered helps eliminate errors before the I/O configuration is implemented.

2.2.2 CHPID Mapping Tool

The CHPID Mapping Tool helps with IBM Z requirements. It provides a mechanism to map CHPIDS to PCHIDS and identify the best availability options for installed features and defined configuration.

Consider using the mapping tool for all new builds of IBM Z family or when upgrading from one system to another. You can also use it as part of standard hardware changes (for example, MES) for an existing Z mainframe.

The mapping tool takes input from two sources:
- The Configuration Report file (CFreport) produced by the IBM order tool and provided by the IBM account team, or produced by IBM manufacturing and obtained from IBM Resource Link®
- An IOCP statement file

The mapping tool produces the following outputs:
- Tailored reports
  - Save all reports for reference. Supply the port report that is sorted by CHPID number and location to the IBM hardware service representative for IBM Z installations.
  - An IOCP input file with PCHIDs mapped to CHPIDs
    - This IOCP input file can then be migrated back into HCD and used to build a production IODF.

The mapping tool does not automatically map CS5 or CIB CHPIDs to AIDs and ports. This process must be done manually, either in HCD, IOCP, or the mapping tool. The mapping tool provides availability intersects for completely defined CIB CHPIDs. For more information on the CHPID Mapping Tool go to:


2.3 I/O configuration planning

I/O configuration planning for IBM Z requires the availability of physical resources, and must comply with the specific rules of the logical definitions. The following physical resources are the minimum that is required for connectivity:
- Platform frame (or rack)
- PCIe+ I/O drawer, PCIe I/O drawer, I/O drawer, or I/O cage, in a frame (or rack)
- I/O slot in a PCIe+ I/O drawer, PCIe I/O drawer, I/O drawer, or I/O cage
- Channel feature in a slot of a PCIe+ I/O drawer, PCIe I/O drawer, I/O drawer, or I/O cage
- Port on a channel feature
For a system configuration, the Z configurator build process includes all physical resources that are required for a particular I/O configuration, based on the supplied channel type and quantity requirements.

The definition phase starts after the physical resources are ordered. The channels must be defined according to the architecture’s rules, the system’s implementation, and the order.

There are also other definition implications that are imposed by the system implementation, such as the maximum number of subchannels in the HSA:

- 3832.5 thousand subchannels total (63.75 thousand per partition x 30 partitions + 64 thousand per partition x 30 partitions) for the zBC12
- 11,505 thousand subchannels total (63.75 thousand per partition x 60 partitions + 2 x 64 thousand per partition x 60 partitions) for the zEC12
- 21,738.75 thousand subchannels total (63.75 thousand per partition x 85 partitions + 3 x 64 thousand per partition x 85 partitions) for the z14 and the z13
- 7670 thousand subchannels total (63.75 thousand per partition x 40 partitions + 2 x 64 thousand per partition x 40 partitions) for the z14 ZR1 and z13s

The operational characteristics of a particular channel type, along with the addressing capabilities, can affect configuration complexity, topology, infrastructure, and performance.

### 2.3.1 I/O configuration rules

The following sections briefly describe the Z configuration rules, which are identified through the architecture and the specific system that are implemented and enforced by using the HCD and IOCP.

All control units and I/O devices that attach to the system must be defined to a CSS. Specify the following items when defining the I/O configuration for the system through HCD/IOCP:

- Logical partitions (logical partition name, CSS ID, and MIF ID, where appropriate)
- Channel paths on the system, their assignment to CSSes, and logical partitions
- FICON Directors (where appropriate)
- Control units that are attached to the channel paths
- I/O devices that are assigned to the control units

**Cryptographic features:** The cryptographic features on IBM Z do not require a CHPID definition and are configured by using the HMC or SE.
Certain definition characteristics that must be considered for the I/O configuration are found in the architecture (z/Architecture). Other definition characteristics are specified only for the system. Table 2-11 lists general IOCP rules for IBM Z.

### Table 2-11 IBM Z general IOCP rules

<table>
<thead>
<tr>
<th>Constant machine attributes</th>
<th>z14, z14 ZR1, z13, z13s, zEC12, and zBC12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum configurable physical control units (PCUs)</td>
<td></td>
</tr>
<tr>
<td>PCUs per OSD</td>
<td>16</td>
</tr>
<tr>
<td>PCUs per OSE, OSC, OSN⁹</td>
<td>1</td>
</tr>
<tr>
<td>PCUs per OSM, OSX</td>
<td>16</td>
</tr>
<tr>
<td>PCUs per CFP, ICP</td>
<td>1</td>
</tr>
<tr>
<td>PCUs or link addresses per FC</td>
<td>256</td>
</tr>
<tr>
<td>PCUs per FCP</td>
<td>1</td>
</tr>
<tr>
<td>CUs per IQD</td>
<td>64</td>
</tr>
<tr>
<td>Maximum configurable devices</td>
<td></td>
</tr>
<tr>
<td>Per CFP, ICP, CIB (12x IFB, or 1x IFB), CS5, CL5</td>
<td>8</td>
</tr>
<tr>
<td>CIB (1x IFB)</td>
<td>32</td>
</tr>
<tr>
<td>Per CNC</td>
<td>1024</td>
</tr>
<tr>
<td>Per CTC</td>
<td>512</td>
</tr>
<tr>
<td>Per OSC⁰</td>
<td>253</td>
</tr>
<tr>
<td>Per OSD</td>
<td>1920</td>
</tr>
<tr>
<td>Per OSE</td>
<td>254</td>
</tr>
<tr>
<td>Per OSM, OSX</td>
<td>1920</td>
</tr>
<tr>
<td>Per OSN⁹</td>
<td>480</td>
</tr>
<tr>
<td>Per FCP</td>
<td>480</td>
</tr>
<tr>
<td>Per FC</td>
<td>32,000³¹</td>
</tr>
<tr>
<td>For all IQD channel paths</td>
<td>12,000</td>
</tr>
</tbody>
</table>

a. OSN CHPID type NOT supported on z14 or z14 ZR1.

b. A limit of 120 clear sessions and 48 encrypted sessions can be defined at the HMC/SE.

c. zEC12 and zBC12 support 24,000 FC channels.

See *IBM Z Input/Output Configuration Program User’s Guide for ICP IOCP, SB10-7172*, for details about CHPID types and channel configuration rules.

### 2.4 References

The following publications include information that is related to the topics covered in this chapter:

- z/OS Hardware Configuration Definition User’s Guide, SC34-2669
- z/OS Hardware Configuration Definition Planning, GA32-0907
- Hardware Configuration Definition Planning, GC28-1750
- Input/Output Configuration Program User's Guide for ICP IOCP, SB10-7172
- IBM zEnterprise EC12 Technical Guide, SG24-8049
- IBM zEnterprise BC12 Technical Guide, SG24-8138
- IBM zEnterprise System Technical Introduction, SG24-8050
- IBM z13 and IBM z13s Technical Introduction, SG24-8250
- IBM z13 Technical Guide, SG24-8251
- IBM z13s Technical Guide, SG24-8294
- IBM z14 Technical Introduction, SG24-8450
- IBM z14 Technical Guide, SG24-8451
- IBM z14 ZR1 Technical Introduction, SG24-8550
- IBM z14 ZR1 Technical Guide, SG24-8651
Chapter 3. zHyperLink Express

IBM zHyperLink is a technology that can provide up to a 5x reduction in I/O latency through an enhanced Synchronous I/O model. This goal is achieved by using a direct connection between an IBM Z platform and IBM Storage.

This chapter describes the zHyperLink connectivity option, which offered with the z14 and z14 ZR1, and the DS888x.

This chapter includes the following topics:

- Description
- zHyperLink elements
- Connectivity
- References
3.1 Description

The zHyperLink technology was created to provide fast access to data through a low latency, short distance, direct connection between the IBM Z platform and IBM Storage. This short-distance, direct connection is intended to speed up Db2 for z/OS blocking read requests.

Working with the FICON SAN Infrastructure, zHyperLink can improve application response time by cutting I/O-sensitive workload response time by up to 50% without requiring application changes.

The zHyperLink Express feature in the z14 and z14 ZR1 allows you to make synchronous I/O requests for data that is in the cache of the IBM DS888x storage. This process is done by directly connecting a zHyperLink Express port in the z14 or z14 ZR1 to an I/O Bay zHyperLink port of the DS888x.

**Note:** zHyperLink connections work in synergy with FICON channels, they do not replace them (a FICON channel is required to “drive” the zHyperLink). Only z/OS and Extended Count Key Data (ECKD) are supported, and the z/OS image must run in an LPAR, not as a guest under z/VM.

3.2 zHyperLink elements

Synchronous I/O is an I/O model type that is used as part of the zHyperLink technology. It allows the operating system to read data records synchronously, thus avoiding the scheduling and interrupt overhead associated with asynchronous operations.

A new Synchronous I/O command was defined to z/OS to synchronously read one or more data records. In addition, an option is provided to allow z/OS to initiate a Synchronous I/O command and return control to perform other processing requests.

When a traditional I/O operation is requested by a task to start the Input/Output Supervisor (IOS), the I/O operation is not handled by the central processor (CP) assigned to the z/OS LPAR. There are specialized components to do the job: SAPs and channel programs.

An SAP is a special processing unit (PU) that helps set up the I/O operation. It finds an available channel path to the device and ensures that the I/O operation starts. However, an SAP is not responsible for the data movement between the storage control unit and z/OS LPAR memory. It is the channel program that communicates with control unit to manage the data movement. After the I/O operation completes, an I/O interrupt notifies the CP so that IOS can be run again.

For Synchronous I/O, the CP directly issues the I/O request to the storage control unit through a zHyperLink connection with a new z/OS I/O command and new hardware capabilities (zHyperLink Express feature, Z firmware, a DS888x I/O system board). The SAP and channel subsystem are bypassed with the Synchronous I/O model. I/O interrupts and I/O path-lengths are minimized, resulting in improved performance.
3.3 Connectivity

The IBM zHyperLink Express takes up one slot in the PCIe I/O drawer or the PCIe+ I/O drawer, and has two ports. Both ports are on a single PCHID. The zHyperLink Express uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes. It is designed to support distances up to 150 meters at a link data rate of 8 GBps.

Figure 3-1 shows a zHyperLink Express feature in the z14 connecting to a zHyperLink feature in the I/O Bay of a DS888x.

The zHyperLink Express (FC 0431) is available on the z14 and z14 ZR1. It requires 24x MTP-MTP cables to connect to a DS888x and z/OS V2.1 or later.

The zHyperLink Express feature is an IBM-designed PCIe adapter. It is managed by using native z/PCI commands, with enhancements for Synchronous I/O.

Up to 16 zHyperLink Express adapters can be installed in a z14 and z14 ZR1, thus allowing up to 32 links.

The zHyperLink Express feature works as native PCIe adapter that can be shared by multiple LPARs. Each port can support up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This configuration supports a maximum of 254 VFs per adapter.

For the DS888x with firmware R8.3 or above, the I/O Bay system board is updated to support the zHyperLink interface.

**zHyperLink support:** Supported features and functions for IBM zHyperLink Express (FC 0431) require compatible levels of DS8880/F firmware and Db2 12 with PTFs.

**zHyperlink cabling requirements**

A 24x MTP-MTP cable is required for each port of the zHyperLink Express feature. It is a single 24-fiber cable with Multi-fiber Termination Push-on (MTP) connectors. Internally, the single cable houses 12 fibers for transmit and 12 fibers for receive.
Two fiber type options are available with specifications that support different distances for the zHyperLink Express:

- Up to 150 m: OM4 50/125 micrometer multimode fiber optic cable with a fiber bandwidth wavelength of 4.7 GHz-km @ 850 nm.
- Up to 100 m: OM3 50/125 micrometer multimode fiber optic cable with a fiber bandwidth wavelength of 2.0 GHz-km @ 850 nm.

3.4 References

The following publications contain information that is related to the topics covered in this chapter:

- *IBM DS8880 and IBM z Systems® Synergy*, REDP-5196
- *Enhancing Value to Existing and Future Workloads with IBM z13*, REDP-5135
- “Making DB2 Zip with IBM’s Hyperwrite” in *IBM Systems Magazine*. 

---

36  IBM Z Connectivity Handbook
Chapter 4. Fibre Connection Express

This chapter describes the Fibre Connection (FICON) Express features and the protocols that they support on IBM Z platforms.

This chapter includes the following sections:

- FICON Express description
- FICON elements
- Connectivity
- References
4.1 FICON Express description

FICON provides the non-drop distance 9 - 100 km. FICON supports the link data rate of 1, 2, 4, 8, or 16 Gbps depending on the FICON features:

- FICON Express16S+ and FICON Express16S features automatically negotiate to 4, 8, or 16 Gbps.
- FICON Express8S and FICON Express8 features automatically negotiate to 2, 4, or 8 Gbps.
- FICON Express4 features auto-negotiate to 1, 2, or 4 Gbps.

The FICON implementation enables full-duplex data transfer, so data travels in both directions simultaneously. FICON also enables multiple concurrent I/O operations.

Terminology: Throughout this chapter, FICON refers to the FICON Express4, FICON Express8, FICON Express8S, FICON Express16S, and FICON Express16S+ features, except when the function being described is applicable to a specific feature.

The FICON channel matches data storage and access requirements with the latest technology in servers, control units, and storage devices. FICON channels allow faster and more efficient data transfer while allowing you to use their currently installed single mode and multimode fiber optic cabling plant.

FICON uses long wavelength (LX) and short wavelength (SX) transceivers with single mode and multi-mode fiber optic media for data transmission.

4.1.1 FICON modes and topologies

IBM Z supports the FICON channel to operate in one of two modes:

- FICON native mode (CHPID type FC)
- Fibre Channel Protocol (CHPID type FCP)

Note: The z14 and z14 ZR1 supports the FICON Express8S only when carried forward on an upgrade.

The IBM z13 and IBM z13s are the last systems to support FICON Express8 features. The z13 and z13s support the FICON Express8 features only when carried forward on an upgrade. The features cannot be ordered for new builds of z14, z14 ZR1, z13, and z13s platforms. Review the use of your installed FICON Express8 channels and, where possible, upgrade to FICON Express16S+ or FICON Express 16S features.

The IBM zEnterprise EC12 (zEC12) and IBM zEnterprise BC12 (zBC12) are the last platforms to support FICON Express4 features. zEC12 and zBC12 support the FICON Express4 feature only when carried forward on an upgrade. The feature cannot be ordered for new builds of z13, z13s, zEC12, or zBC12 systems. Review the use of your installed FICON Express4 channels and, where possible, upgrade to FICON Express16S or FICON Express8S features.
FICON native mode (FC)
As shown in Figure 4-1, a FICON channel in FICON native mode (CHPID type FC) can access FICON native interface control units by using the following topologies:

- Point-to-point (direct connection)
- Switched point-to-point (through a Fibre Channel switch)
- Cascaded FICON Directors (through *two* Fibre Channel switches)

![Figure 4-1  Native FICON supported topologies](image)

A FICON native channel also supports channel-to-channel communications. The FICON channel at each end of the FICON CTC connection, which supports the FCTC control units, can also communicate with other FICON native control units, such as disk storage devices and tape. At least one end of the FICON CTC connection must be an IBM Z installation.

Fibre Channel Protocol (FCP) mode
A FICON channel in Fibre Channel Protocol mode (CHPID type FCP) can access FCP devices in either of the following ways:

- A FICON channel in FCP mode through a single Fibre Channel switch or multiple switches to a Small Computer System Interface (SCSI) device
- A FICON channel in FCP mode through a single Fibre Channel switch or multiple switches to a Fibre Channel-to-SCSI bridge
The FICON features support Fibre Channel and SCSI devices in IBM z/VM, IBM z/VSE, and Linux on IBM Z and the KVM hypervisor, as shown in Figure 4-2.

On IBM Z, point-to-point connections can be used to access data that is stored on devices without the use of a Fibre Channel switch. In addition, an operating system or other stand-alone program can be IPLed through a point-to-point connection by using the SCSI IPL feature. N_Port ID Virtualization is not supported by FCP point-to-point. For more information, see “Worldwide port name tool” on page 43.

The FCP support allows z/VM, Linux on IBM Z and the KVM hypervisor, and z/VSE operating systems to access industry-standard SCSI devices. For disk applications, these FCP storage devices use fixed block (512-byte) sectors rather than the extended count key data (IBM ECKD) format.

4.1.2 FCP Channel

The FC-FCP standard was developed by the International Committee of Information Technology Standards (INCITS) and published as an ANSI standard. The IBM Z FCP I/O architecture conforms to the FC standards specified by the INCITS. For more information about the FC standards, see the INCITS Technical Committee T11 website and their page for SCSI Storage Interfaces (this is the committee within INCITS that is responsible for the Fibre Channel Interface).
FICON channels in FCP mode provide full fabric attachment of SCSI devices to the operating system images by using the Fibre Channel Protocol and provide point-to-point attachment of SCSI devices. This technique allows z/VM, Linux on IBM Z and the KVM hypervisor, and z/VSE to access industry-standard SCSI storage controllers and devices.

The FCP channel full fabric support means that multiple numbers of directors or switches can be placed between the Z platform and the SCSI device. This technique allows many hops through a storage area network (SAN) and provides improved use of intersite connected resources and infrastructure. This expanded ability to attach storage devices provides more choices for storage solutions and the ability to use existing storage devices. This configuration can facilitate the consolidation of UNIX server farms onto Z platforms, protecting investments in SCSI-based storage.

For a list of switches, storage controllers, and devices that are verified to work in a Fibre Channel network that is attached to FCP channel, and for software requirements to support FCP and SCSI controllers or devices, see the I/O Connectivity website.

FICON channels in FCP mode are based on the Fibre Channel standards that are defined by INCITS and published as ANSI standards. FCP is an upper-layer Fibre Channel mapping of SCSI on a common stack of Fibre Channel physical and logical communication layers.

SCSI is supported by a wide range of controllers and devices, complementing the classical storage attachment capability through FICON channels. FCP is the base for industry-standard Fibre Channel networks or SANs.

Fibre Channel networks consist of servers, storage controllers, and devices as end nodes, which are interconnected by Fibre Channel switches, directors, and hubs. Switches and directors are used to build Fibre Channel networks or fabrics. Fibre Channel Arbitrated Loops (FC-ALs) can be constructed by using Fibre Channel hubs. In addition, different types of bridges and routers can be used to connect devices with different interfaces, such as parallel SCSI. All of these interconnections can be combined in the same network.

SCSI is implemented by many vendors in many different types of storage controllers and devices. These controllers and devices are widely accepted in the marketplace and have proven to be able to meet the reliability, availability, and serviceability (RAS) requirements of many environments.

FICON channels in FCP mode use the queued direct input/output (QDIO) architecture for communication with the operating system. The QDIO architecture for FCP channels derives from the QDIO architecture, which was defined initially for the OSA-Express features and for HiperSockets communications.

FCP channels do not use control devices. Instead, data devices that represent QDIO queue pairs are defined, consisting of a request queue and a response queue. Each queue pair represents a communication path between an operating system and the FCP channel. It allows an operating system to send FCP requests to the FCP channel through the request queue. The FCP channel uses the response queue to pass completion indications and unsolicited status indications to the operating system. Figure 4-3 on page 42 illustrates this process.

HCD/IOCP is used to define the FCP channel type and QDIO data devices. However, there is no definition requirement for the Fibre Channel storage controllers and devices, or for the Fibre Channel interconnect units, such as switches, directors, and bridges. The FCP industry standard architecture requires that the Fibre Channel devices (end nodes) in a Fibre Channel network are addressed by using worldwide names (WWNs), Fibre Channel Identifiers (IDs), and logical unit numbers (LUNs).
These addresses are configured on an operating system level and passed to the FCP channel together with the corresponding Fibre Channel I/O or service request through a logical QDIO device (queue).

Figure 4-3 summarizes the necessary FCP I/O definitions and compares them to FICON I/O definitions.

**Channel and device sharing**

An FCP channel can be shared among multiple Linux operating systems, each running in a logical partition or as a guest operating system under z/VM. To access the FCP channel, each operating system needs its own QDIO queue pair, which is defined as a data device on an FCP channel in the HCD/IOCP.

Each FCP channel can support up to 480 QDIO queue pairs with the Z platform. This support allows each FCP channel to be shared among 480 operating system instances (with a maximum of 252 guests per logical partition).

Host operating systems that share access to an FCP channel can establish up to 2048 concurrent connections to up to 512 different remote Fibre Channel ports that are associated with Fibre Channel controllers. The total number of concurrent connections to end devices, identified by LUNs, must not exceed 4096.

Although multiple operating systems can concurrently access the same remote Fibre Channel port through a single FCP channel, Fibre Channel devices, identified by their LUNs, can be reused only serially. For two or more unique operating system instances to share concurrent access to a single Fibre Channel or SCSI device (LUN), each of these operating systems must access this device through a different FCP channel.
If two or more unique operating system instances attempt to share concurrent access to a single Fibre Channel or SCSI device (LUN) over the same FCP channel, a LUN sharing conflict occurs and errors result. A way to alleviate this sharing conflict on Z platforms is to use N_Port ID Virtualization.

Worldwide port name tool
The worldwide port name tool assigns WWPNs to each FCP channel or port by using the same WWPN assignment algorithms that a system uses when you are assigning WWPNs for channels that use NPIV. Therefore, the SAN can be set up in advance, allowing operations to proceed much faster after the system is installed.

The WWPN tool can calculate and show WWPNs for both virtual and physical ports ahead of system installation. This feature means that the SAN configuration can be retained rather than altered by assigning a WWPN to physical FCP ports when a FICON feature is replaced.

The WWPN tool takes an adhesive file that contains the FCP-specific I/O device definitions and creates the WWPN assignments that are required to set up the SAN. A binary configuration file that can be imported later by the system is also created. The CSV (.csv) file can either be created manually or exported from the hardware configuration definition/hardware configuration manager (HCD/HCM).

The WWPN tool can be downloaded from the Tools section of IBM Resource Link (requires registration).

FCP SCSI IPL feature enabler
This function allows IPLing an operating system from an FCP channel-attached disk to run either in a logical partition or as a guest operating system under z/VM. In particular, SCSI IPL can directly IPL a Linux operating system that was installed previously on a SCSI disk.

The IPL device is identified by its SAN address, consisting of the WWPN of the disk controller and the LUN of the IPL device. SCSI IPL is supported in the following conditions:

- FCP access control
- Point-to-point connections
- N-Port ID Virtualization (NPIV)

A stand-alone-dump program can also be IPLed from an FCP channel that is attached to a SCSI disk. The stand-alone-dump program can also store the generated dumped data on a SCSI disk. z/VM support of SCSI IPL allows Linux and other guest operating systems that support this feature to be IPLed from an FCP-attached SCSI disk when z/VM is running on an IBM Z platform. Therefore, Linux guests can be started and run completely from an FCP channel-attached disk.

For more information about FCP SCSI IPL, see Linux on zSeries: Fibre Channel Protocol Implementation Guide, SG24-6344.

FCP multipathing concept
Multipath access to disk subsystems is a basic feature with the channel subsystem on Z platforms. FICON connections support multiple hardware paths to any physical disk device. The Z platform handles multipathing invisibly through the operating system. With FICON channels, the operating system is presented a single device for I/O operations, and multipathing happens under channel subsystem control.
Multipathing over FCP is different. With FCP multipathing on Linux on IBM Z, each path to each LUN appears to the operating system as a separate device. For example, if there are four paths to five LUNs, the Linux kernel defines 20 SCSI devices. This configuration means that there must be another layer of code between the Linux file system layer and the SCSI subsystem. This extra layer handles all of the coordination between the raw paths and the higher-level file subsystem. Currently, supported distributions use Multiple Device Administration (mdadm) tools. For more information, see the RAID setup website.

**FCP access control**
The ability to control access to nodes and devices is provided as a function in switches and controllers, and is called *LUN masking* and *zoning*. LUN masking and zoning can be used to prevent systems from accessing storage that they are not permitted to access:

- **LUN masking**
  A LUN represents a portion of a controller, such as a disk device. With the use of LUNs, a controller can be logically divided into independent elements or groups of elements. Access to these LUNs can be restricted to distinctive WWPNs as part of the controller configuration. This method is known as *LUN masking*.

- **Zoning**
  Segmentation of a switched fabric is achieved through *zoning*. It can be used to fence off certain portions of the switched fabric, allowing only the members of a zone to communicate within that zone. All others that attempt to access from outside of that zone are rejected.

**I/O devices**
The IBM Z FCP channel implements the FCP standard as defined by the INCITS Fibre Channel Protocol for SCSI (FC-FCP) and Fibre Channel Protocol for SCSI Second Version (FCP-2), and the relevant protocols for the SCSI-2 and SCSI-3 protocol suites. Theoretically, each device that conforms to these protocols works when attached to an IBM Z FCP channel. However, experience shows that there are small deviations in the implementations of these protocols.

Also, for certain types of FCP and SCSI controllers and devices, specific drivers in the operating system might be required to use all of the capabilities of the controller or device. The drivers might also be required to cope with unique characteristics or deficiencies of the device.

**Note:** Do appropriate conformance and interoperability testing to verify that a particular storage controller or device can be attached to an IBM Z FCP channel in a particular configuration. For example, test that it can be attached through a particular type of Fibre Channel switch, director, or point-to-point connection.

**Hardware assists**
A complementary virtualization technology is available for the Z platform. The technology includes these capabilities:

- QDIO Enhanced Buffer-State Management (QEBSM), with two hardware instructions that are designed to eliminate the overhead of hypervisor interception
- Host Page-Management Assist (HPMA), which is an interface to the z/VM main storage management function that allows the hardware to assign, lock, and unlock page frames without z/VM hypervisor assistance
These hardware assists allow a cooperating guest operating system to start QDIO operations directly to the applicable channel, without interception by z/VM, providing additional performance improvements. This support is integrated in the Z platform. Consult the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

Support of T10-DIF for enhanced reliability
Because high reliability is important for maintaining the availability of business-critical applications, the IBM Z FCP implements support of the ANSI T10 Data Integrity Field (DIF) standard. Data integrity protection fields are generated by the operating system and propagated through the SAN. IBM Z helps to provide added end-to-end data protection between the operating system and the storage device.

An extension to the standard, Data Integrity Extensions (DIX), provides checksum protection from the application layer through the host bus adapter (HBA), where cyclical redundancy check (CRC) protection is implemented.

T10-DIF support by the FICON features, when defined as CHPID type FCP, is available on the Z platform. Use of the T10-DIF standard requires support by the operating system and the storage device.

4.1.3 FCP and FICON mode characteristics

The single largest difference between the FICON channel (FC) and FCP channel mode types is the treatment of data access control and security. FICON channels rely on multiple image facility (MIF) to address concerns about shared channels and devices. MIF provides ultra-high access control and security of data so that one operating system image and its data requests cannot interfere with another operating system’s data requests. With the introduction of IBM Z, MIF continues this ultra-high level of access control and security across channel subsystems (CSSes).

FCP and MIF
Linux guest operating systems under z/VM can have access to an FCP channel defined to the z/VM operating system. Using MIF, an FCP channel can also be shared between Linux Logical Partitions and z/VM Logical Partitions with Linux guests.

The FCP industry-standard architecture does not use the data access control and security functions of MIF. As a result, FCP has the following limitations:

- Channel sharing

  When NPIV is not implemented, and if multiple Linux images share an FCP channel and all of the Linux images have connectivity to all of the devices that are connected to the FCP fabric, all of the Linux images use the same worldwide port name (WWPN). They use this name to enter the fabric and are indistinguishable from each other within the fabric. Therefore, the use of zoning in switches and LUN masking in controllers is not effective in creating appropriate access controls among the Linux images.

  By using NPIV, each operating system that shares an FCP channel is assigned a unique WWPN. The WWPN can be used for device-level access control in storage controllers (LUN masking) and in switch-level access control (zoning).
Device sharing

Without using NPIV, an FCP channel prevents logical units from being opened by more than one Linux image at a time. Access is on a first-come, first-served basis. This system prevents problems with concurrent access from Linux images that are share a FCP channel and, therefore, the same WWPN. This serialization means that one Linux image can block other Linux images from accessing the data on one or more logical units unless the sharing images (z/VM guests) are not in contention.

FICON versus FCP

FICON and FCP have other significant differences. Certain differences are fundamental to the IBM Z family, whereas others are fundamental to the two channel architectures. Still others depend on the operating system and the device being attached. Without taking the operating system and the storage device into consideration, I/O connectivity through IBM Z FCP and FICON channels has the following differences:

- Direct connection

With all FICON features on the Z platform, storage controllers can be directly connected to the channel by using point-to-point attachment when in FCP mode. There is no need for a director or switch between the FCP channel and storage controllers.

**Note:** Although N_Port ID Virtualization is supported in switched topology, it is not supported in a point-to-point topology.

- Switch topology

FCP channels support full fabric connectivity, meaning that several directors or switches can be used between a Z platform and the device. With the FICON cascaded director support, the FICON storage network topology is limited to a two-director, single-hop configuration.

- Enterprise fabric

The use of cascaded FICON Directors ensures the implementation of a high-integrity fabric. For FCP, a high-integrity fabric solution is not mandatory, although it must be considered. For example, if an FCP inter-switch link (ISL) must be moved, data potentially might be sent to the wrong path without notification. This type of error does not happen on an enterprise fabric with FICON.

- Transmission data checking

When a transmission is sent through an FCP channel, because of its full fabric capability, FCP checks data for each leg of that transmission. FICON also checks intermediate data.

- Serviceability

  - Licensed Internal Code (LIC) updates and the Z platform itself allow FCP concurrent patches. FICON channels, when configured as CHPID type FCP, support concurrent patches, allowing the application of a LIC without requiring a configuration of off/on. This is an exclusive FCP availability feature that is available with all FICON features.

  - The FICON features have Small Form-factor Pluggable (SFP) optics to permit each channel to be individually serviced during a fiber optic module failure. The traffic on the other channels on the same feature can continue to flow if a channel requires servicing.
Problem determination

- Request Node Identification (RNID)

RNID assists with the isolation of FICON cabling detected errors. Resolving fiber optic cabling problems can be a challenge in a fiber optic environment with extended distances. To facilitate resolution, the operating system can request the RNID data for each device or control unit that is attached to native FICON channels. You can then display the RNID data by using an operator command. RNID is available to the Z platform, and is supported by all FICON features (CHPID type FC) and by z/OS.

- Link incident reporting

Link incident reporting is integral to the FICON architecture. When a problem on a link occurs, this mechanism identifies the two connected nodes between which the problem occurred, leading to faster problem determination and service. For FCP, link incident reporting is not a requirement for the architecture, although it might be offered as an optional switch function. Therefore, important problem determination information might not be available if a problem occurs on an FCP link.

IBM Z allows z/OS to register for all FICON link incident records. This feature improves your ability to capture data for link incident analysis across multiple systems. This function is supported by z/OS version 1.8 or later.

- Simplified problem determination

To more quickly detect fiber optic cabling problems in a storage area network, all FICON channel error information is forwarded to the HMC. This function facilitates detection and reporting trends and thresholds for the channels with aggregate views, including data from multiple systems.

Problem determination can be simplified by using the HMC to pinpoint fiber optic cabling issues in your SAN fabric without involving IBM service personnel.

All FICON channel error information is forwarded to the HMC. In the HMC, it is analyzed to detect and report the trends and thresholds for all FICON channels on the Z platform. The Fibre Channel Analyzer task on the HMC can be used to display analyzed information about errors on FICON channels (CHPID type FC) of attached Support Elements. Data includes information about the PCHID, CHPID, channel type, source link address, and destination link address where the error occurred. This report shows an aggregate view of the data and can span multiple systems.

Starting with z13, similar FICON problem determination tools have been implemented for FCP channels. These channel problem determination tools for FCP channels include analyze functions such as analyze channel information, subchannel data, device status, serial link status, and link error statistic block. In addition to the analyze functions, fabric status login and SAN explorer functions are also available. These FCP problem determination tools are accessed from the HMC in the same way as for the FICON channels.

- FICON purge path extended

The purge path extended function enhances FICON problem determination. FICON purge path error recovery function is extended so that it transfers error-related data and statistics between the channel and entry switch and the control unit and its entry switch to the host operating system. FICON purge path extended use requires a switch or device that supports this function. The purge path extended function for FCP channels is available on z14 and z13.
FICON error recovery

Z platforms, z/OS, and I/O recovery processing are designed to allow the system to detect switch/director fabric problems that might cause FICON links to fail and recover multiple times in a short time.

This feature allows the system to detect these conditions and keep an affected path offline until an operator action is taken. This process is expected to limit the performance impacts of switch/director fabric problems. The FICON error recovery function is available in z/OS.

Forward Error Correction

Forward Error Correction (FEC) is a technique that is used for controlling errors in data transmission over unreliable or noisy communication channels. By adding redundancy and error-correcting code (ECC) to the transmitted information, the receiver detects and corrects a limited number of errors in the information instead of requesting a retransmission. This process improves the reliability and bandwidth utilization of a connection by saving retransmissions due to bit errors. This advantage is true especially for connections across long distances, such as an ISL in an IBM Geographically Dispersed Parallel Sysplex™ (IBM GDPS®) MetroMirror environment.

FICON Express16S+ and FICON Express16S support FEC coding on top of their 64 b/66 b data encoding for 16 Gbps connections. Their FEC design can correct up to 11 bit errors per 2112 bits transmitted. Thus, when connected to devices that support FEC at 16 Gbps connections, the FEC design allows FICON Express16S+ or FICON Express16S channels to operate at higher speeds over longer distances and with reduced power and higher throughput. At the same time, the FEC design maintains the same reliability and robustness that FICON channels are known for.

With the IBM DS8870 or later, the z14, z14 ZR1, z13, and z13s can extend the use of FEC to the fabric N_Ports for a completed end-to-end coverage of 16 Gbps FC links. For more information, see the IBM DS8880 and IBM Z Synergy, REDP-5186.

FICON dynamic routing

With the IBM z14, z14 ZR1, z13, and z13s, FICON channels are no longer restricted to the use of static SAN routing policies for ISLs for cascaded FICON directors. IBM Z now supports dynamic routing in the SAN with the FICON Dynamic Routing (FIDR) feature. It supports the dynamic routing policies provided by the FICON director manufacturers, such as Brocade’s Exchange Based Routing 7 (EBR 7) and Cisco’s Open Exchange ID Routing (OxID).

With FIDR, z14, z14 ZR1, z13, and z13s has advantages for performance and management in configurations with ISL and cascaded FICON directors:

- Support sharing of ISLs between FICON and FCP (PPRC or distributed)
- Better balanced I/O traffic between all available ISLs
- Improved utilization of the FICON director and ISL
- Easier management with a predictable and repeatable I/O performance

FICON dynamic routing can be enabled by defining dynamic routing capable switches and control units in HCD. Also, z/OS has implemented a health check function for FICON dynamic routing.

FICON performance

See the latest FICON and FCP performance information at the IBM I/O Connectivity web page.
4.2 FICON elements

FICON enables multiple concurrent I/O operations to occur simultaneously to multiple control units. FICON channels also permit intermixing of large and small I/O operations on the same link. The data center I/O configuration now has increased flexibility for connectivity because of the increased I/O rate, increased bandwidth, and multiplexing of mixed workloads.

4.2.1 FICON channel

FICON channel architecture is compatible with the following protocols:

- Fibre Channel Physical and Signaling standard (FC-FS)
- Fibre Channel Switch Fabric and Switch Control Requirements (FC-SW)
- Fibre Channel Single-Byte-3 (FC-SB-3) and Fibre Channel Single-Byte-4 (FC-SB-4) standards

Cabling specifications are defined by the Fibre Channel - Physical Interface - 4 (FC-PI-4) standard and used by IBM Z FICON features. Table 4-1 identifies cabling types and link data rates that are supported in the FICON environment, including their allowable maximum distances and link loss budget. The link loss budget is derived from the channel insertion loss budget that is defined by the FC-PI-4 standard (Revision 8.00).

Table 4-1  Fiber optic cabling for FICON: Maximum distances and link loss budget

<table>
<thead>
<tr>
<th>FC-PI-4 Fiber core (light source)</th>
<th>Cable Type Category</th>
<th>2 Gbps Distance in meters</th>
<th>Link-loss budget in dB</th>
<th>4 Gbps Distance in meters</th>
<th>Link-loss budget in dB</th>
<th>8 Gbps Distance in meters</th>
<th>Link-loss budget in dB</th>
<th>16 Gbps Distance in meters</th>
<th>Link-loss budget in dB</th>
<th>10 Gbps ISL a Distance in meters</th>
<th>Link-loss budget in dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 µm SM (10 km LX)</td>
<td>OS1/OS2</td>
<td>10000</td>
<td>7.8</td>
<td>10000</td>
<td>7.8</td>
<td>10000</td>
<td>6.4</td>
<td>10000</td>
<td>6.4</td>
<td>10000</td>
<td>6.4</td>
</tr>
<tr>
<td>9 µm SM (4 km LX)</td>
<td>OS1/OS2</td>
<td>4000</td>
<td>4.8</td>
<td>4000</td>
<td>4.8</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>50 µm MM 4700 MHz·km (SX laser)</td>
<td>OM4</td>
<td>500</td>
<td>3.31</td>
<td>400</td>
<td>2.95</td>
<td>190</td>
<td>2.19</td>
<td>125</td>
<td>1.95</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>50 µm MM 2000 MHz·km (SX laser)</td>
<td>OM3</td>
<td>500</td>
<td>3.31</td>
<td>380</td>
<td>2.88</td>
<td>150</td>
<td>2.04</td>
<td>100</td>
<td>1.86</td>
<td>300</td>
<td>2.6</td>
</tr>
<tr>
<td>50 µm MM 500 MHz·km (SX laser)</td>
<td>OM2</td>
<td>300</td>
<td>2.62</td>
<td>150</td>
<td>2.06</td>
<td>50</td>
<td>1.68</td>
<td>35</td>
<td>1.63</td>
<td>82</td>
<td>2.3</td>
</tr>
<tr>
<td>62.5 µm MM 200 MHz·km (SX laser)</td>
<td>OM1</td>
<td>150</td>
<td>2.1</td>
<td>70</td>
<td>1.78</td>
<td>21</td>
<td>1.58</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

a. ISL between two FICON Directors.

Note: IBM does not support a mix of 50 µm and 62.5 µm fiber optic cabling in the same physical link.
When an application performs an I/O operation to a device represented by a unit control block (UCB), it initiates an I/O request by using macros or Supervisor Call to the input/output supervisor (IOS). The application or access method also provides the channel program (channel command words (CCWs)) and additional parameters in the operation request block (ORB). This request is queued on the UCB. The IOS services the request from the UCB on a priority basis.

The IOS then issues a start subchannel (SSCH) instruction, with the subsystem identifier (SSID) that represents the device and the ORB as operands. The CSS is signaled to perform the operation. This flow is shown in Figure 4-4.

The most appropriate FICON channel is selected by the CSS. The FICON channel fetches channel programs (CCWs) prepared by the application, fetches data from memory (write) or stores data into memory (read), and presents the status of the operation to the application (I/O interrupt).

The z/Architecture channel commands, data, and status are packaged by the FICON channel into FC-SB-3 or FC-SB-4 (FC-4 layer) Information Units (IUs). IUs from several different I/O operations to the same or different control units and devices are multiplexed or demultiplexed by the FC-2 layer (framing). These FC-2 frames, with encapsulated FC-SB-3 or FC-SB-4 IUs, are encoded or decoded by the FC-1 layer (encode or decode) and sent to or received from the FC-0 fiber optic medium (optics).
On a FICON channel, CCWs are transferred to the control unit without waiting for the first command response from the control unit or for a CE/DE after each CCW execution. The device presents a logical end to the control unit after each CCW execution. If the last CCW of the CCW chain has been run by the device, the control unit presents CE/DE to the channel. Figure 4-5 shows a CCW operation on a FICON channel that uses CCW chaining.

![Figure 4-5  CCW chaining](image)

### 4.2.2 High Performance FICON for IBM Z

High Performance FICON for IBM Z (zHPF) is an enhancement of the FICON channel architecture that is compatible with these protocols:

- Fibre Channel Physical and Signaling standard (FC-FS)
- Fibre Channel Switch Fabric and Switch Control Requirements (FC-SW)
- Fibre Channel Single-Byte-4 (FC-SB-4) standards

Using zHPF with the FICON channel, the z/OS operating system, and the control unit reduces the FICON channel overhead. This goal is achieved by protocol optimization and reducing the number of IUs processed, resulting in more efficient use of the fiber link.

The FICON Express16S+, FICON Express16S, FICON Express8S, FICON Express8, and FICON Express4 features support both the existing FICON architecture and the zHPF architecture. From the z/OS point of view, the existing FICON architecture is called command mode, and the zHPF architecture is called transport mode. A parameter in the ORB is used to determine whether the FICON channel is running in command or transport mode.

The mode that is used for an I/O operation depends on the control unit that is supporting zHPF and the settings in the z/OS operating system. An IEIOSxxx parameter and SETIOS commands in z/OS can enable or disable zHPF. Support is also added for the D IOS, ZHPF system command to indicate whether zHPF is enabled, disabled, or not supported on the system.

During link initialization, both the channel and the control unit indicate whether they support zHPF. The Process Login (PRLI) support indicator is presented in response to the RNID Extended Link Services (ELS). If PRLI is supported, the channel sends a PRLI ELS. The PRLI response then indicates that zHPF is supported by the CU.
Similar to the existing FICON channel architecture described previously, the application or access method provides the channel program (CCWs) and parameters in the ORB. Bit 13 in word 1 of the ORB specifies how to handle the channel program in either command mode or transport mode.

The way that zHPF transport mode manages CCW operation is significantly different from the CCW operation for the existing FICON architecture command mode, as shown in Figure 4-6. In command mode, each single CCW is sent to the control unit for execution. In transport mode, all CCWs are sent over the link in one single frame to the control unit. Certain complex CCW chains are not supported by zHPF. Figure 4-6 shows an example of the optimization by a zHPF transport mode read operation.

The channel sends all the required CCWs and read operations of 4 KB of data in one single frame to the control unit. The control unit transfers the requested data over the link to the channel, followed by a CE/DE if the operation was successful. Less overhead is generated compared with the existing FICON architecture.
Figure 4-7 shows the same reduction of frames and open exchanges for a zHPF transport mode write operation.

![Diagram of FICON Channel and Control Unit with CCWs and Data Transfers]

**Figure 4-7  High-performance FICON write operation**

The channel sends all the required CCWs and write operations of 4 KB of data in one frame to the control unit. The control unit responds with XFER when it is ready to receive the data. The channel then sends the 16 KB of data to the control unit. If the control unit successfully receives the data and finishes the write operation, the CE/DE status is sent by the control unit to indicate the completion of the write operation.

zHPF supports multitrack operations. It allows the channel to operate at rates that fully use the bandwidth of a FICON Express channel. The zHPF fully supports multiple tracks of data that can be transferred in a single operation.

For more information about the FICON channel protocol and zHPF, see *FICON Planning and Implementation Guide*, SG24-6497.

### 4.2.3 Platform and name server registration in FICON channel

All FICON features on the Z platform support platform and name server registration to the fabric. That support exists only if the FICON feature is defined as channel-path identifier (CHPID) type FC.

Information about the channels that are connected to a fabric, if registered, allow other nodes or SAN managers to query the name server to determine what is connected to the fabric. The following attributes are registered for the Z platform:

- **Platform information:**
  - Worldwide node name (WWNN). This is the node name of the platform and is the same for all channels that belong to the platform.
  - Platform type.
  - Host computer type.
  - Platform name. The platform name includes vendor ID, product ID, and vendor-specific data from the node descriptor.

- **Channel information.**
- Worldwide port name (WWPN).
- Port type (N_Port_ID).
- FC-4 types supported.
- Classes of service that are supported by the channel.

The platform and name server registration service are defined in the Fibre Channel - Generic Services 4 (FC-GS-4) standard.

### 4.2.4 Open exchanges

An open exchange, part of FICON and FC terminology, represents an I/O operation in progress over the channel. Many I/O operations can be in progress over FICON channels at any one time. For example, a disk I/O operation might temporarily disconnect from the channel when performing a seek operation or while waiting for a disk rotation. During this disconnect time, other I/O operations can be managed as follows:

- **Command mode open exchanges**
  
  In command mode, the number of open exchanges is limited by the FICON Express feature. FICON Express16S+, FICON Express16S, FICON Express8S, FICON Express8, and FICON Express4 allow up to 64 open exchanges. One open exchange (an exchange pair, actually) in command mode is the same as one I/O operation in progress.

- **Transport mode open exchanges**
  
  In transport mode, one exchange is sent from the channel to the control unit. Then, the same exchange ID is sent back from the control unit to the channel to complete the I/O operation. The maximum number of simultaneous exchanges that the channel can have open with the CU is 750 exchanges. The CU sets the maximum number of exchanges in the status area of the transport mode response IU. The default number is 64, which can be increased or decreased.

In addition, FICON channels can multiplex data transfer for several devices at the same time. This feature also allows workloads with low to moderate control unit cache hit ratios to achieve higher levels of activity rates per channel.

If the open exchange limit is reached, additional I/O operations are refused by the channel, which can result in queues and retries by the operating system.

### Extended distances

Degradation of performance at extended distances can be avoided by implementing an enhancement to the industry standard FICON architecture (FC-SB-3). This enhancement is a protocol for persistent IU pacing. Control units that use the architecture can increase the pace count, meaning the number of information units that are allowed to be underway between the channel and the control unit. Extended distance FICON channels remember the last pacing information and use this information for subsequent operations. This feature avoids performance degradation at the start of a new operation.

Information unit pacing helps to optimize the link use and simplifies the requirements for channel extension equipment because more commands can be in flight. Extended distance FICON is apparent to the operating systems and is applicable to all FICON features defined with CHPID type FC.
Modified Indirect Data Address Word
On IBM Z, the Modified Indirect Data Address Word (MIDAW) provides alternatives to using CCW data chaining in channel programs. The MIDAW facility has been added to z/Architecture and can coexist with the current CCW IDAW facility.

MIDAW decreases channel, fabric, and control unit overhead by reducing the number of CCWs and frames that are processed and allows scattering of data in memory for noncontiguous real pages. Although the CCW IDAW function requires all but the first and last IDAW in a list to deal with complete 2 thousand or 4 thousand units of data, the MIDAW facility allows page boundary crossing on either 2 thousand or 4 thousand boundaries. This feature allows access to data buffers anywhere in a 64-bit buffer space.

Figure 4-8 shows an example of MIDAW use.

The use of MIDAWs is indicated by the MIDAW bit (flag bit 7) in the CCW. The last MIDAW in the list has a last flag set, indicated by L in Figure 4-8. MIDAW provides significant performance benefits, especially when processing extended format data sets with FICON channels.

FICON link
The transmission medium for the FICON interface is a fiber optic cable. Physically, it is a pair of optical fibers that provide two dedicated, unidirectional, serial-bit transmission lines. Information in a single optical fiber flows, bit by bit, in one direction. At any link interface, one optical fiber is used to receive data, and the other is used to transmit data. Full-duplex capabilities are used for data transfer. The Fibre Channel Standard (FCS) protocol specifies that for normal I/O operations frames flow serially in both directions, allowing several concurrent read and write I/O operations on the same link.
These are the link data rates:

- 1, 2, or 4 Gbps for FICON Express4 channels
- 2, 4, or 8 Gbps for FICON Express8 and FICON Express8S channels
- 4, 8, or 16 Gbps for FICON Express16S and FICON Express16S+ channels

Whether these link data rates can be achieved depends on many factors, such as transfer sizes and access methods used. The link speed capability is automatically negotiated between the system and FICON Director, and the director and CUs, and is apparent to the operating system and the application.

In general, the FICON channels and the FICON Director or CU communicate and agree upon either a 1, 2, 4, 8, or 16 Gbps (that is, 100 MBps, 200 MBps, 400 MBps, 800 MBps, or 1600 MBps) link speed. This speed determination is based on the supported speeds in the system feature, FICON Director, and CU.

**Note:** The link speed is the theoretical maximum unidirectional bandwidth capability of the fiber optic link. The actual link data rate, whether it is measured in I/O operations per second or MBps, depends on the type of workload, fiber infrastructure, and storage devices in place.

FICON LX features use LX transceivers and 9 µm single mode fiber optic media (cables and trunks) for data transmission. FICON SX features use SX transceivers and 50 or 62.5 µm multimode fiber optic media (cables and trunks) for data transmission. A transceiver is a transmitter and receiver. The transmitter converts electrical signals to optical signals to be sent over the fiber optic media. The receiver converts optical signals to electrical signals to be sent through the system, director, or CU.

**FICON to ESCON conversion**

For the requirement of connecting to ESCON devices, see Appendix D, “Channel conversion solutions”.

**FICON and Fibre Channel switch**

The Fibre Channel switch fabric (FC-SW) supports packet-switching. It allows up to 64 simultaneous concurrent I/O operations (read and write) from multiple FICON-capable systems to multiple FICON control units.
Figure 4-9 shows a conceptual view of frame processing in a switched point-to-point configuration for multi-system and multi-control unit environments.

**Figure 4-9  FICON switch function**

**FICON support of cascaded Fibre Channel Directors**

FICON native channels on IBM Z platforms support cascaded FICON Directors. This support is for a two-director configuration only. With cascading, a FICON native channel, or a FICON native channel with channel-to-channel (FCTC) function, you can connect a system to a device or other system through two native connected Fibre Channel Directors. This cascaded director support is for all native FICON channels that are implemented on IBM Z platforms.

FICON support of cascaded Fibre Channel Directors, sometimes referred to as *cascaded switching* or *two-switch cascaded fabric*, is for single-vendor fabrics only.
Cascaded support is important for disaster recovery and business continuity solutions, as shown in Figure 4-10. It can provide high availability connectivity and the potential for fiber infrastructure cost savings for extended storage networks. FICON two-director cascaded technology allows for shared links, so it improves use of connected site resources and infrastructure. Solutions such as IBM GDPS can benefit from the reduced intersite configuration complexity that is provided by FICON support of cascaded directors.

![Figure 4-10 Two-site FICON connectivity with cascaded directors](image)

Generally, organizations that have data centers that are separated between two sites can reduce the number of cross-site connections by using cascaded directors. Specific cost savings vary depending on the infrastructure, workloads, and size of data transfers. Further savings can be realized by reducing the number of channels and switch ports. Another important value of FICON support of cascaded directors is its ability to provide high-integrity data paths. The high-integrity function is an integral component of the FICON architecture when you are configuring FICON channel paths through a cascaded fabric.

To support the introduction of FICON cascaded switching, IBM worked with Fibre Channel Director vendors. IBM and the vendors worked to ensure that robustness in the channel-to-control unit path is maintained to the same high standard of error detection, recovery, and data integrity as the initial implementation of FICON.

End-to-end data integrity is maintained through the cascaded director fabric. Data integrity helps ensure that any changes to the data streams are always detected and that the data frames (data streams) are delivered to the correct endpoint. The endpoint is a FICON channel port or a FICON control unit port. For FICON channels, CRCs and longitudinal redundancy checks (LRCs) are bit patterns added to the data streams to allow for detection of any bit changes in the data stream. With FICON support of cascaded switching, integrity features are introduced within the FICON channel and the FICON cascaded switch fabric. This feature helps to ensure the detection and reporting of any incorrect cabling actions that occur within the fabric during operation that might cause a frame to be delivered to the wrong endpoint.
A FICON channel, when configured to operate with a cascaded switch fabric, requires that the switch fabric supports high integrity. During initialization, the FICON channel queries the switch fabric to determine that it supports high integrity. If it does, the channel completes the initialization process, allowing the channel to operate with the fabric.

After a FICON switched fabric is customized to support FICON cascaded directors and the required WWNN and domain IDs are added in the fabric membership list, the director checks that its ISLs are attached to the correct director before they are operational. If an accidental cable swap occurs, the director starts logical path testing, reporting, isolation, and recovery. The high integrity fabric feature for cascaded FICON Directors protects against miscabling and misdirecting of data streams, as shown in Figure 4-11. The checking process follows this sequence:

1. Channel initialization completes.
2. Later, miscabling occurs (for example, cables are swapped at a patch panel).
3. The director port enters invalid attachment state and notifies state change back to the Z platform.
4. The Z platform starts the channel logical path testing, reporting, isolation, and error recovery.
5. Any I/O requests to the invalid route are discarded until the error is corrected.
6. Data is protected. Channel initialization completes.

High-integrity fabric architecture support includes these capabilities:

- Fabric binding support

  This is the ability of the fabric to prevent a switch from being added to the fabric that is not configured to support the high integrity fabric. For example, all switches in the required fabric must be defined to the fabric by using a fabric membership list.
Insistent domain IDs support

This support does not allow a switch address to be automatically changed when a duplicate switch address is added to the enterprise fabric. It requires operator action to change a switch address.

FICON Dynamic Routing

FICON Dynamic Routing is a feature on the z14, z14 ZR1, z13, and z13s that enables exploitation of SAN dynamic routing policies in the fabric to lower cost and improve performance for supporting I/O devices.

A static SAN routing policy typically assigns the ISL routes according to the incoming port and its destination domain (port-based routing), or the source and destination ports pairing (device-based routing).

Port-based routing (PBR) assigns the ISL routes statically based on first come, first served when a port starts a fabric login (FLOGI) to a destination domain. The ISL is round robin selected for assignment. Thus, I/O flow from the same incoming port to the same destination domain is always assigned the same ISL route, regardless of the destination port of each I/O. This process can result in some ISLs being overloaded at the same time that others are under utilized. The ISL routing table is changed every time that a Z server undergoes a power-on-reset (POR). Because of this POR, the ISL assignment is unpredictable.

Device-base routing (DBR) assigns the ISL routes statically based on a hash of the source and destination port. That I/O flow from the same incoming port to the same destination is assigned with the same ISL route. Compared to PBR, DBR can better spread load across ISLs for I/O flow from the same incoming port to different destination ports within the same destination domain.

When using a static SAN routing policy, the FICON director has limited capability to assign ISL routes based on workload. There are also chances of ISL overloaded or under-utilization.

With dynamic routing, ISL routes are dynamically changed based on the Fibre Channel exchange ID, which is unique for each I/O operation. The ISL is assigned at I/O request time so that different I/Os from the same incoming port to the same destination port are assigned different ISLs. See Figure 4-12 for an illustration of FICON dynamic routing.

![Figure 4-12  FICON dynamic routing](image-url)
Dynamic routing (Brocade EBR or Cisco OxID) dynamically changes the routing between the channel and control unit based on the Fibre Channel Exchange ID. Each I/O operation has a unique exchange ID. The following are some of the benefits of FIDR:

- Reduces cost by allowing sharing of ISLs between FICON and FCP (PPRC or distributed)
- Provides better balanced I/O traffic between all available ISLs
- Improves utilization of switch and ISL hardware
- Provides easier management
- Allows for easier capacity planning for ISL bandwidth requirements
- Provides predictable, repeatable I/O performance

**FICON control units and devices**

The control unit receives commands from the channel, receives data from and transmits data to the channel, and controls execution of commands and data transfer at associated devices. The control unit attaches to one or multiple links through port interfaces.

FICON does not allow multiple physical control units, or control unit link interfaces, to be on the same link. A control unit can contain multiple images with dedicated or shared physical control unit facilities. The FICON I/O architecture provides addressing for these multiple CU images.

**Channel device-addressing support**

The FC device addressing for the zEC12 and the zBC12 increases to 24,000 devices\(^2\) and 32,000 devices\(^3\) for the IBM z14, z14 ZR1, z13, and z13s.

**Multiple image facility**

MIF enables sharing FICON channels among logical partitions that are running on Z platforms. For more information, see Chapter 2, “Channel subsystem overview” on page 15.

### 4.2.5 Spanned channels

*Spanning* is the ability to configure channels to multiple channel subsystems. When so defined, the channels can be transparently shared by any or all of the configured logical partitions, regardless of the channel subsystem to which the logical partition is configured.

FICON channels can be spanned across multiple CSSes in the Z platform. See 2.1.2, “Multiple CSS concept” on page 17, for more details about MIF and spanned channels.

### 4.2.6 Control unit port

The CUP function allows z/OS to manage a FICON Director with the greater level of control and security. Host communication includes control functions like blocking or unblocking ports, monitoring, and error-reporting functions.

IBM Tivoli® System Automation for z/OS (SA for z/OS) includes support for FICON channels and FICON Directors. You can find additional information, updates, extensions, tools, and demonstrations at the IBM DevOps web page.

---

\(^2\) Applies to the FICON Express8S, FICON Express8, and FICON Express4 features supported by z/OS, z/VM, and Linux on IBM Z and the KVM hypervisor.

\(^3\) Applies to the FICON Express16S+, FICON Express16S and FICON Express8S features supported by z/OS, z/VM, and Linux on IBM Z and the KVM hypervisor.
Before using SA for z/OS in your FICON environment, check the latest maintenance recommendations in the appropriate z/OS subset of the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

### 4.2.7 z/OS discovery and auto-configuration

IBM z/OS Discovery and Automatic Configuration (zDAC) is a function that is supported on the Z platform. It automatically performs several I/O configuration definition tasks for new and changed disk and tape controllers that are connected to a FICON Director. It helps simplify I/O configuration of IBM Z CPCs running z/OS and helps reduce complexity and setup time.

The zDAC function is integrated into the existing HCD tool. A policy can be defined in the HCD according to the availability and bandwidth requirements, including parallel access volume (PAV) definitions, control unit numbers, and device number ranges. The zDAC proposed configurations are created as work I/O definition files (IODF) that can be converted to production IODF and activated.

zDAC provides real-time discovery for the FICON fabric, subsystem, and I/O device resource changes from z/OS. By exploring the discovered control units for defined LCUs and devices, zDAC compares the discovered controller information with the current system configuration to determine delta changes to the configuration for a proposed configuration.

All new added or changed logical control units and devices are added to the proposed configuration with proposed control unit and device numbers and channel paths based on the defined policy. zDAC uses a channel path chosen algorithm to minimize single point of failure. zDAC applies to all FICON features supported on the Z platform when configured as CHPID type FC.

### 4.3 Connectivity

The connectivity options for the FICON I/O interface environment are discussed in this section. Table 1-1 on page 10 lists the maximum number of FICON channels that are supported, based on each system.

Table 4-2 lists the available FICON features and their respective specifications. All FICON features use LC duplex connectors. For long wavelength FICON features that can use a data rate of 1 Gbps, mode-conditioning patch cables (MCP), either 50 or 62.5 MM, can be used. The maximum distance for this connection is reduced to 550 m at a link data rate of 1 Gbps. Details for each feature follow the table.

<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distancea</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICON Express4 10KM LX</td>
<td>3321</td>
<td>1, 2, or 4 Gbps</td>
<td>SM 9 µm</td>
<td>10 km/20 km</td>
<td>zEC12, zBC12b</td>
</tr>
<tr>
<td>FICON Express4 4KM LX</td>
<td>3324</td>
<td>4 Gbps</td>
<td>SM 9 µm</td>
<td>4 km</td>
<td>zEC12, zBC12b</td>
</tr>
</tbody>
</table>

---

a. Distances are reduced to 550 m at 1 Gbps.
<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distance</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICON Express4 SX</td>
<td>3322</td>
<td>4 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>70 m (200) 150 m (500) 380 m (2000)</td>
<td>zEC12, zBC12b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>300 m (200) 500 m (500) 860 m (2000)</td>
<td></td>
</tr>
<tr>
<td>FICON Express4-2C SX</td>
<td>3318</td>
<td>4 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>70 m (200) 150 m (500) 380 m (2000)</td>
<td>zBC12b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>300 m (200) 500 m (500) 860 m (2000)</td>
<td></td>
</tr>
<tr>
<td>FICON Express8 10KM LX</td>
<td>3325</td>
<td>2, 4, or 8 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z13, z13s, zEC12, zBC12</td>
</tr>
<tr>
<td>FICON Express8 SX</td>
<td>3326</td>
<td>8 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>21 m (200) 50 m (500) 150 m (2000)</td>
<td>zEC12, zBC12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>70 m (200) 150 m (500) 380 m (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
</tr>
<tr>
<td>FICON Express8S 10KM LX</td>
<td>0409</td>
<td>2, 4, or 8 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14, z14 ZR1b, z13, z13s, zEC12, zBC12</td>
</tr>
<tr>
<td>FICON Express8S SX</td>
<td>0410</td>
<td>8 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>21 m (200) 50 m (500) 150 m (2000)</td>
<td>zEC12, zBC12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>70 m (200) 150 m (500) 380 m (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm, MM 50 µm</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S 10KM LX</td>
<td>0418</td>
<td>4, 8, or 16 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14, z14 ZR1b, z13, z13s</td>
</tr>
</tbody>
</table>
The ports on a single FICON Express16S, FICON Express8S, FICON Express8, and FICON Express4 feature can be configured individually and can be defined in different channel modes (FC and FCP).

FICON Express16S+: With FICON Express16S+ both ports must be defined as the SAME channel type (either FC or FCP). Mixing channel types on a FICON Express16S+ is not supported.

Note: FICON Express4 features were the last FICON features to support 1 Gbps link data rates.

For more information about extended distances for FICON, see Chapter 10, “Extended distance solutions” on page 169.
4.3.1 FICON Express16S+

The FICON Express16S+ features have two independent ports. Each feature occupies a single I/O slot, using one CHPID per channel. Each channel supports 4 Gbps, 8 Gbps, and 16 Gbps link data rates with auto-negotiation.

These features are supported on IBM z14. FICON Express16S+ has increased performance compared to FICON Express 16S.

Each FICON Express16S+ feature has two ports. Each port has a PCHID and can be defined as FC or FCP type. However, both ports must be same CHPID type. FICON Express16S+ CHPIDs can be defined as a spanned channel and can be shared among logical partitions within and across CSSes.

All FICON Express16S+ features reside in the PCIe I/O drawer or PCIe+ I/O drawer and use Small Form-factor Pluggable (SFP) optics to permit each channel to be individually serviced during a fiber optic module failure. Traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

The FICON Express16S+ features are ordered in two channel increments and are designed to be added concurrently. This concurrent update capability allows you to continue to run workloads through other channels when the FICON Express16S+ features are being added.

The FICON Express16S+ features are designed for connectivity to systems, switches, directors, disks, tapes, and printers, and can be defined in two ways:

- **CHPID type FC**
  - Native FICON, zHPF, and FICON channel-to-channel (CTC) traffic
  - Supported in the IBM z/OS, IBM z/VM hypervisor, IBM z/VSE V6.2 (earlier z/VSE versions - no zHPF support), IBM z/Transaction Processing Facility (z/TPF), and Linux on IBM Z and the KVM hypervisor environments

- **CHPID type FCP**
  - Fibre Channel Protocol traffic for communication with SCSI devices
  - Supported in IBM z/VM, z/VSE, and Linux on IBM Z and the KVM hypervisor environments

**KVM hypervisor notes:** KVM for IBM z Systems Version 1.1.2 (with a fixpack) was the last release delivered by IBM. It has been withdrawn from marketing and is no longer in service. With Announcement Letter 917-04, dated March 7, 2017, IBM announced that it is working with Linux distributions to support the KVM component on IBM Z. Refer to your distribution’s documentation regarding KVM support for the Z platform.

4.3.2 FICON Express16S

The FICON Express16S features have two independent ports. Each feature occupies a single I/O slot, using one CHPID per channel. Each channel supports 4 Gbps, 8 Gbps, and 16 Gbps link data rates with auto-negotiation.

---

4 PCIe+ I/O drawer (FC #4001) is available on IBM z14 Model ZR1 and replaces PCIe I/O drawer. All PCIe features that can be carried forward during an upgrade to z14 ZR1 can be installed in the PCIe+ I/O drawer. PCIe I/O drawer (FC #4032) CANNOT be carried forward during an upgrade to z14 ZR1.
These features are supported on IBM z14 and z14 ZR1 (carry forward only); and z13 and z13s platforms, and are designed to deliver increased performance in comparison to FICON Express8S features. The FICON Express16S features include half the number of ports per feature in comparison with the FICON Express8 features. This design facilitates purchasing the correct number of ports to help satisfy your application requirements and to better optimize for redundancy.

All FICON Express16S features reside in the PCIe I/O drawer or PCIe+ I/O drawer and use SFP optics to permit each channel to be individually serviced during a fiber optic module failure. Traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

The FICON Express16S features are ordered in two channel increments and are designed to be added concurrently. This concurrent update capability allows you to continue to run workloads through other channels when the FICON Express16S features are being added.

FICON Express16S CHPIDs can be defined as a spanned channel and can be shared among logical partitions within and across CSSes.

The FICON Express16S features are designed for connectivity to systems, switches, directors, disks, tapes, and printers and can be defined in two ways:

- **CHPID type FC**
  - Native FICON, zHPF, and FICON CTC traffic
  - Supported in the IBM z/OS, IBM z/VM hypervisor, IBM z/VSE V6.2 (earlier z/VSE versions - no zHPF support), IBM z/TPF, and Linux on IBM Z and the KVM hypervisor environments

- **CHPID type FCP**
  - Fibre Channel Protocol traffic for communication with SCSI devices
  - Supported in IBM z/VM, z/VSE, and Linux on IBM Z and the KVM hypervisor environments

### 4.3.3 FICON Express8S

The FICON Express8S features have two independent ports. Each feature occupies a single I/O slot, using one CHPID per channel. Each channel supports 2 Gbps, 4 Gbps, and 8 Gbps link data rates with auto-negotiation.

These features are supported on the Z platform (carry forward only on z14 and z14 ZR1), and are designed to deliver increased performance in comparison to FICON Express8 features. The FICON Express8S features include half the number of ports per feature in comparison with the FICON Express8 features. This design facilitates purchasing the correct number of ports to help satisfy your application requirements and to better optimize for redundancy.

All FICON Express8S features reside in the PCIe I/O drawer or PCIe+ I/O drawer and use SFP optics to permit each channel to be individually serviced during a fiber optic module failure. Traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

The FICON Express8S features are ordered in two channel increments and are designed to be added concurrently. This concurrent update capability allows you to continue to run workloads through other channels when the FICON Express8S features are being added.
FICON Express8S CHPIDs can be defined as a spanned channel and can be shared among logical partitions within and across CSSes.

The FICON Express8S features are designed for connectivity to systems, switches, directors, disks, tapes, and printers, and can be defined in two ways:

- **CHPID type FC**
  - Native FICON, zHPF, and FICON CTC traffic
  - Supported in the IBM z/OS, IBM z/VM hypervisor, IBM z/VSE V6.2 (earlier z/VSE versions - no zHPF support), IBM z/TPF, and Linux on IBM Z and the KVM hypervisor environments

- **CHPID type FCP**
  - Fibre Channel Protocol traffic for communication with SCSI devices
  - Supported in z/VM, z/VSE, and Linux on IBM Z and the KVM hypervisor environments

### 4.3.4 FICON Express8

The FICON Express8 features have four independent ports. Each feature occupies a single I/O slot, using one CHPID per channel. Each channel supports 2 Gbps, 4 Gbps, and 8 Gbps link data rates with auto-negotiation.

All FICON Express8 features use SFP optics so that each channel can be serviced individually during a fiber optic module failure. Traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

The FICON Express8 features are ordered in four channel increments and are designed to be added concurrently. This concurrent update capability enables you to continue to run workloads through other channels when the FICON Express8 features are being added.

FICON Express8 CHPIDs can be defined as a spanned channel and can be shared among logical partitions within and across CSSes.

The FICON Express8 features are designed for connectivity to systems, switches, directors, disks, tapes, and printers, and can be defined in these two ways:

- **CHPID type FC**
  - Native FICON, zHPF, and FICON CTC traffic
  - Supported in the z/OS, z/VM, z/VSE V6.2 (earlier z/VSE versions - no zHPF support), z/TPF, and Linux on IBM Z and the KVM hypervisor environments

- **CHPID type FCP**
  - Fibre Channel Protocol traffic for communication with SCSI devices
  - Supported in z/VM, z/VSE, and IBM Linux on IBM Z and the KVM hypervisor environments

**Note:** IBM z13 and z13s are the last IBM Z platforms to support FICON Express8. These features are supported as carry-forward only with an I/O drawer, for a maximum of eight features (32 ports).
### 4.3.5 FICON Express4

The FICON Express4 features have four independent ports (or two for the two-port features). Each feature occupies a single I/O slot, using one CHPID per channel. Each channel supports 1 Gbps, 2 Gbps, and 4 Gbps link data rates with auto negotiation.

All FICON Express4 features use SFP optics so that each channel can be serviced individually during a fiber optic module failure. Traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

The FICON Express4 features are ordered in four-channel (or two-channel) increments and are designed to be added concurrently. This concurrent update capability enables you to continue to run workloads through other channels when the FICON Express4 features are being added.

FICON Express4 CHPIDs can be defined as a spanned channel and can be shared among logical partitions in and across CSSes.

The FICON Express4 features are designed for connectivity to systems, switches, directors, disks, tapes, and printers, and can be defined in these ways:

- **CHPID type FC**
  - Native FICON, zHPF, and FICON CTC traffic
  - Supported in the z/OS, z/VM, z/VSE V6.2 (earlier z/VSE versions - no zHPF support), z/TPF, and Linux on IBM Z

- **CHPID type FCP**
  - Fibre Channel Protocol traffic for communication with SCSI devices
  - Supported in z/VM, z/VSE, and Linux on IBM Z environments

**Note:** zEC12 and zBC12 are the last IBM Z platforms to support FICON Express4. These features were withdrawn from the market and are supported only as carry-forward features.

### 4.3.6 Qualified FICON and FCP products

See IBM Resource Link for information about IBM Z qualified FICON and FCP products, and products that support intermixing of FICON and FCP within the same physical FC switch or FICON Director (requires registration).

On the left, select **Library** and locate the listing for “Hardware products for servers” in the middle of the web page. Then, select **Switches and directors that are qualified for IBM Z® FICON and FCP channels**.

### 4.3.7 Software support

For more information about the operating systems that are supported on IBM Z platforms see this website.

**Note:** Certain functions might require specific levels of an operating system, program temporary fixes (PTFs), or both. That information is provided when necessary within this chapter.
Consult the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

4.3.8 Resource Measurement Facility

IBM Resource Measurement Facility™ (IBM RMF™) reporting is available for all FICON features. This application enables you to capture performance data through the following reports:

- Channel path activity report (of primary interest)
- Device activity report
- FICON Director activity report
- I/O queuing activity report

With these reports, you can analyze possible bandwidth bottlenecks to determine the cause. For further performance information, see the IBM Z I/O connectivity web page.

4.4 References

The following publications contain information that is related to the topics covered in this chapter:

- *FICON I/O Interface Physical Layer*, SA24-7172
- *IBM Cabling System Optical Fiber Planning and Installation*, GA27-3943
- *FICON Planning and Implementation Guide*, SG24-6497

For Fibre Channel Standard publications, see the Technical Committee T11 website.

For more information about the SCSI Storage Interface standards, see the Technical Committee T10 website.
Open Systems Adapter-Express

This chapter describes the IBM Open Systems Adapter-Express features. These features provide connectivity to other servers and clients in 1000BASE-T Ethernet (10, 100, and 1000 Mbps), Gigabit Ethernet (GbE), 10-Gigabit Ethernet, and 25-Gigabit Ethernet environments.

Terminology: If not stated otherwise, the term OSA applies to all OSA-Express features throughout this book.

The following OSA features are described in this chapter:

- OSA-Express7S
- OSA-Express6S
- OSA-Express5S
- OSA-Express4S
- OSA-Express3

Unless otherwise stated, they are referred to as “OSA-Express”.

The following topics are covered in this chapter:

- Functional description
- OSA capabilities
- Connectivity

Statement of Direction: This chapter makes references to IBM Statements of Direction (SoDs). All statements regarding IBM plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these SoDs is at the relying party’s sole risk and will not create liability or obligation for IBM.
5.1 Functional description

The seven generations of the OSA-Express features (OSA-Express, OSA-Express2, OSA-Express3, OSA-Express4S, OSA-Express5S, OSA-Express6S, and OSA-Express7S) integrate several hardware features and support many networking transport protocols. Every new OSA-Express feature provides enhancements in the areas of function, connectivity, bandwidth, data throughput, network availability, reliability, and recovery.

Notes: Consider the following points:

- Throughout this chapter, we provide Operating System (OS) support information for the functions that are described. Not every OS supports all features. The OS support information is provided for IBM z/OS, IBM z/VM, IBM z/VSE, IBM z/TPF, and Linux on IBM Z (supported distributions).
- KVM is included with Linux Distributions. For more information about the latest support, contact Linux Distribution providers. At the time of this writing (October 2018), the KVM hypervisor is supported with the following minimum Linux distributions:
  - SLES 12 SP2 with service.
  - RHEL 7.5 with kernel-alt package (kernel 4.14).
  - Ubuntu 16.04 LTS with service and Ubuntu 18.04 LTS with service.

5.1.1 Operating modes

The integration of a channel path with network ports makes the OSA a unique channel or CHPID type, which is recognized by the hardware I/O configuration on a port-by-port basis as one of the following types:

- Queued Direct Input/Output (OSD)
- Non-Queued Direct Input/Output (OSE)
- OSA Integrated Console Controller (OSC)
- Open Systems Adapter for Network Control Program (OSN)
- OSA-Express for zEnterprise BladeCenter Extension (OSX)
- OSA-Express for Management of an ensemble (OSM)
- OSA-Express for Dynamic Partition Manager (OSM)

Not all features support all CHPID types. Table 5-1 provides an overview of the type of traffic that is supported. It also indicates whether the Open Systems Adapter Support Facility (OSA/SF) is required to configure the OSA-Express6S, OSA-Express5S, or OSA-Express4S ports that are based on the supported modes of operation (CHPID types).

OSA and Dynamic Partition Manager

OSM channel type can support Dynamic Partition Manager and Ensemble Management. Both functions can be configured and ordered, but they cannot be enabled at the same time on the system. Dynamic Partition Manager requires two OSA-Express 1000BASE-T Ethernet features for primary and backup connectivity.

The OSA-Express5S 1000BASE-T Ethernet or OSA-Express6S 1000BASE-T Ethernet features cannot be shared and must be dedicated for usage only by Dynamic Partition Manager (DPM) or only by Ensemble Membership (depending on which function is enabled on the system). A CPC in DPM mode cannot be part of an Ensemble.
Table 5-1  Supported CHPID types

<table>
<thead>
<tr>
<th>CHPID type</th>
<th>Feature</th>
<th>SNA/APPN/HPR traffic</th>
<th>IP traffic</th>
<th>TN3270E traffic</th>
<th>OSA/SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSD</td>
<td>OSA-Express7S 25GbE</td>
<td></td>
<td></td>
<td></td>
<td>Optional</td>
</tr>
<tr>
<td></td>
<td>OSA-Express6S 10GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 10GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 10GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 10GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express6S GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express6S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>OSE</td>
<td>OSA-Express6S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Required</td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 1000BASE-T</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>OSC^c</td>
<td>OSA-Express6S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>OSM</td>
<td>OSA-Express6S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 1000BASE-T</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>OSX</td>
<td>OSA-Express7S 25GbE</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>OSA-Express6S 10 GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express5S 10 GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express4S 10 GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OSA-Express3 10 GbE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. SNA over IP with the use of Enterprise Extender or TN3270. See 5.2.17, “Enterprise Extender” on page 94 and 5.2.18, “TN3270E Server” on page 94.
b. Layer 2 support allows for non-IP protocols, such as SNA. See 5.2.13, “Layer 2 support” on page 88.
c. OSA-ICC (OSC Channel) now supports Secure Sockets Layer for z14, z14 ZR1, z13, and z13s. See the IBM Z HMC driver level for feature support.

Open Systems Adapter Support Facility
OSA/SF is a host-based tool that is used to customize and manage all OSA features:

- OSA/SF is not required for the OSA feature that is configured for the QDIO mode or the default IP Passthru non-QDIO mode. However, it can be used for problem determination purposes.
- OSA/SF is not required for OSA port CHPID type OSN, although information about channel use can be displayed through OSA/SF for OSN CHPIDs.
- OSA/SF is a required facility when the OSA feature is being configured for shared non-QDIO mode and where SNA definitions are involved.
With the Z platform and OSA/SF, the HMC is enhanced to provide configuration, validation, activation, and display support for the OSA-Express6S, OSA-Express5S, and OSA-Express4S features:

- OSA/SF on the HMC is required for OSA-Express6S and OSA-Express5S features.
- Either OSA/SF on the HMC or the OSA/SF in the operating system component can be used for the OSA-Express4S feature.

One OSA/SF application can communicate with all OSA features in an IBM Z platform. OSA/SF communicates with an OSA feature through a device (type OSD) defined by using HCD/IOCP. For more information, see 5.1.5, “OSA/SF support” on page 79.

**QDIO versus non-QDIO**

Figure 5-1 shows the much shorter I/O process when in QDIO mode rather than non-QDIO mode. I/O interrupts and I/O path-lengths are minimized, resulting in improved performance versus non-QDIO mode, reduced system assist processor (SAP) use, improved response time, and reduced system use.

OSA-Express features use direct memory access (DMA) and a data router model to eliminate store and forward delays.

Also, in QDIO mode, all OSA features dynamically receive configuration information from the host. This process reduces configuration and setup time, eliminates duplicate data entry, and reduces the possibility of data entry errors and incompatible definitions.
5.1.2 QDIO mode

QDIO is a highly efficient data transfer mechanism that dramatically reduces system overhead and improves throughput by using system memory queues and a signaling protocol to directly exchange data between the OSA microprocessor and network software. QDIO is the interface between the operating system and the OSA hardware.

The following components make up QDIO:

- Direct memory access (DMA)
- Data router
- Priority queuing
- Dynamic OSA address table building
- LPAR-to-LPAR communication
- Internet Protocol (IP) Assist functions

QDIO supports IP and non-IP traffic with the OSA-Express features. The following features support two transport modes:

- Layer 2 (Link Layer) for IP (IPv4, IPv6) and non-IP (AppleTalk DECnet, IPX, NetBIOS, or SNA) traffic
- Layer 3 (Network Layer) for IP traffic only

A more detailed description of the Layer 2 support is provided in 5.2.13, “Layer 2 support” on page 88.

Direct memory access

OSA and the operating system share a common storage area for memory-to-memory communication, reducing system overhead, and improving performance. Data can move directly from the OSA microprocessor to system memory and vice versa by using a store-and-forward technique in DMA. There are no read or write channel programs for data exchange. For write processing, no I/O interrupts must be handled. For read processing, the number of I/O interrupts is minimized.

Data router

With OSA-Express4S, OSA-Express5S, OSA-Express6S, and OSA-Express7S, what was previously done in Licensed Internal Code (LIC) is now performed on hardware. Additional logic was added in the IBM ASIC processor of the OSA feature to handle packet construction, inspection, and routing. This feature allows packets to flow between host memory and the LAN at line speed without firmware intervention. With the data router, the store and forward technique in DMA is no longer used, which enables a direct host memory-to-LAN flow and avoids a hop. It reduces latency and increases throughput for standard frames (1492 bytes) and jumbo frames (8992 bytes).

Priority queuing

Priority queuing is supported by the QDIO architecture and introduced with the Service Policy Server (for z/OS environments only). It sorts outgoing IP message traffic according to the service policy that is set up for the specific priority that is assigned in the IP header.

This capability is an alternative to the best-effort priority assigned to all traffic in most IP networks. Priority queuing allows the definition of four different priority levels for IP network traffic through the OSA features defined for QDIO. For example, the highest priority can be granted to interactive communications, and the lowest can be granted to batch traffic, with two more categories in between, perhaps based on particular user groups or projects.
QDIO uses four write (outbound) queues and one read (inbound) queue for each IP network stack that is sharing the OSA feature.

OSA signals the z/OS Communications Server when there is work to do. z/OS Communications Server puts outbound packets in one of the four queues, based on priority settings.

At a certain time, z/OS Communications Server signals the OSA feature that there is work to do. The OSA feature searches the four possible outbound queues by priority and sends the packets to the network, giving more priority to queues 1 and 2, and less priority to queues 3 and 4. For example, if there is data on every queue, queue 1 is served first, then portions of queue 2, then fewer portions of queue 3, then even fewer portions of queue 4, and then back to queue 1. This process means that four transactions are running across the four queues. Over time, queue 1 finishes first, queue 2 finishes second, and so on.

**Note:** With OSA-Express7S, OSA-Express6S, OSA-Express5S, OSA-Express4S, and OSA-Express3, priority queuing is enabled by default. This feature reduces the total number of supported IP network stacks and devices. For more information, see “Maximum IP network stacks and subchannels” on page 79.

**Dynamic OSA Address Table update**

With QDIO, the dynamic OSA Address Table (OAT) update process simplifies installation and configuration tasks. The definition of IP addresses is done in one place, the IP network profile, thus removing the requirement to enter the information into the OAT by using the OSA/SF.

The OAT entries are dynamically built when the corresponding IP device in the IP network stack is started.

At device activation, all IP addresses contained in the IP network stack’s IP HOME list are downloaded to the OSA port. Corresponding entries are built in the OAT. Subsequent changes to these IP addresses cause an update of the OAT.

**LPAR-to-LPAR communication**

Access to an OSA port can be shared among the system images that are running in the logical partitions to which the channel path is defined to be shared. Also, access to a port can be shared concurrently among IP network stacks in the same logical partition or in different logical partitions.

When sharing ports, an OSA port operating in QDIO mode can send and receive IP traffic between logical partitions without sending the IP packets to the LAN and then back to the destination logical partition.

For outbound IP packets, the OSA port uses the next-hop IP address within the packet to determine where it is sent. If the next-hop IP address was registered by another IP network stack that is sharing the OSA port, the packet is sent directly to that IP network stack, not onto the LAN. This feature makes the forwarding of IP packets possible within the same host system.

**Internet Protocol Assist functions**

OSA QDIO helps IP processing and offloads the IP network stack functions for the following processes:

- Multicast support (see “IP network multicast and broadcast support” on page 85)
- Broadcast filtering (see “IP network multicast and broadcast support” on page 85)
- Building MAC and LLC headers
- Address Resolution Protocol (ARP) processing (see “Address Resolution Protocol cache management” on page 86)
- Checksum offload (see “Checksum offload support for z/OS and Linux on IBM Z” on page 87)

**QDIO functions**

The QDIO functions that are discussed in this section are supported on the Z platform.

**IP network functions**

The following IP network functions are available:

- Large send for IP network traffic for OSA-Express (for more information, see “Large send for IP network traffic” on page 82)
- 640 IP network stacks (for more information, see “Maximum IP network stacks and subchannels” on page 79)

**Hardware assists**

Complementary virtualization technology is available that includes these capabilities:

- QDIO Enhanced Buffer-State Management (QEBSM). Two hardware instructions help eliminate the overhead of hypervisor interception.
- Host Page-Management Assist (HPMA). An interface to the IBM z/VM operating system main storage management function to allow the hardware to assign, lock, and unlock page frames without z/VM hypervisor assistance.

These hardware assists allow a cooperating guest operating system to start QDIO operations directly to the applicable channel, without interception by z/VM, which helps improve performance. Support is integrated in IBM Z Licensed Internal Code.

**QDIO Diagnostic Synchronization for z/OS**

QDIO Diagnostic Synchronization is exclusive to IBM Z, and the OSA-Express features when configured as CHPID type OSD (QDIO). It provides the system programmer and network administrator with the ability to coordinate and simultaneously capture both operating system (software) and OSA (hardware) traces at the same instance of a system event.

This process allows the host operating system to signal the OSA-Express features to stop traces and capture the current trace records. By using existing tools (traps) and commands, the operator can capture both hardware and software traces at the same time and then correlate the records during post processing.

**OSA-Express Network Traffic Analyzer for z/OS**

The OSA-Express Network Traffic Analyzer is exclusive to IBM Z and the OSA-Express features when configured as CHPID type OSD (QDIO). It allows trace records to be sent to the host operating system to improve the capability to capture data for both the system programmer and the network administrator. This function allows the operating system to control the sniffer trace for the LAN and capture the records in host memory and storage. It uses existing host operating system tools to format, edit, and process the sniffer records.
5.1.3 Non-QDIO mode

Similar to any other channel-attached control unit and device, an OSA port can run channel programs (CCW chains) and present I/O interrupts to the issuing applications. For non-QDIO mode, the OSA ports are defined as channel type OSE. The non-QDIO mode requires the use of the OSA/SF for setup and customization of the OSA features.

The 1000BASE-T features support non-QDIO mode. This mode supports SNA, APPN, HPR, and IP network traffic simultaneously through the OSA port. This section discusses the non-QDIO mode types.

IP network Passthru

In IP network Passthru mode, an OSA feature transfers data between an IP network stack to which it is defined and clients on an Ethernet 10/100/1000 Mbps LAN. The LAN is attached to the port on a 1000BASE-T feature and supports one of the following frame protocols:

- Ethernet II using the DEC Ethernet V 2.0 envelope
- Ethernet 802.3 using the 802.2 envelope with SNAP

For IP network Passthru mode, the default OAT can be used. In that case, no configuration or setup is required.

SNA/APPN/HPR support

In this mode, an OSA feature acts as an SNA Passthru agent to clients that use the SNA protocol on the LAN that is directly attached to the OSA. If an OSA feature is running in the SNA mode, it is viewed by IBM VTAM® as an external communication adapter (XCA) that can use either switched or non-switched lines of communication.

5.1.4 OSA addressing support

This section describes the maximum number IP addresses, MAC addresses, and subchannels that are supported by the OSA features.

Maximum IP addresses per OAT

The OAT is a component of an OSA feature’s configuration. An OAT entry defines the data path between an OSA feature port and a logical partition (LP) and device unit address. That is, it manages traffic through the OSA CHPID.

OSA-Express features support up to 4096 IP addresses per port.

When the OSA port is defined in QDIO mode, the OAT table entries are built and updated dynamically.

Maximum number of MAC addresses

When configured as an OSD CHPID type, up to 2048 (up to 4096 for the z14) Media Access Control (MAC) or virtual MAC (VMAC) addresses are supported for each port of the OSA feature. Included in the maximum number of MAC addresses is the “burned-in” MAC address of the OSA port.

The MAC or VMAC addresses are added to the Layer 2 table of the OAT when the IP network stacks (in which the addresses are defined) are started.

For more information, see 5.2.13, “Layer 2 support” on page 88 and 5.2.16, “Layer 3 VMAC for z/OS” on page 93.
**Maximum IP network stacks and subchannels**

A subchannel is a logical representation of a device. One subchannel is assigned for each device that is defined to the logical partition. Therefore, if an OSA CHPID is being shared across 15 LPARs and one device is defined, that device uses 15 subchannels.

These are the maximum number of supported IP network stacks and subchannels:

- **OSA port in non-QDIO mode (CHPID type OSE)**
  
  An OSA port in non-QDIO mode can support up to 120 IP network stacks and 240 subchannels for all IBM Z platforms.

- **OSA port in QDIO mode (CHPID type OSD)**
  
  The OSA features support 640 IP network stack connections per dedicated CHPID, or 640 total stacks across multiple logical partitions when defined as a shared or spanned CHPID. The maximum number of subchannels that are allowed is 1920 (1920 subchannels / 3 = 640 stacks).

**Note:** By default, OSA-Express features have multiple priorities for outbound queues enabled (four QDIO priorities). Therefore, the maximum number of supported subchannels is reduced to 480 (1920 subchannels / 4 = 480 subchannels), which reduces the total number of supported IP network stacks to 160 (480 subchannels / 3 = 160 stacks). Priority queues can be disabled through HCD or IOCP. For example, in IOCP use the `CHPARM=02` value to disable priority queuing.

### 5.1.5 OSA/SF support

OSA/SF includes a graphical user interface (GUI) that is based on Java to support the client application. The Java GUI is independent of any operating system or server, and is expected to operate wherever the current Java runtime support is available.

Use of the GUI is optional. A REXX command interface is also included with OSA/SF. OSA/SF is integrated in z/OS, z/VM, and z/VSE, and runs as a host application. For OSA/SF, Java GUI communication is supported through IP networks only. This version of OSA/SF is not being offered as a separately licensed product.

Starting with driver 22 installed on z13 and with next driver up to 32 on z14 server, the HMC is enhanced to use OSA/SF function for the OSA-Express6S, OSA-Express5S, and OSA-Express4S features. Either OSA/SF on the HMC or the OSA/SF in the operating system component can be used for the OSA-Express4S features. For the OSA-Express6S and OSA-Express5S features, OSA/SF on the HMC is required.

OSA/SF is used primarily for the following purposes:

- Manage all OSA ports.
- Configure all OSA non-QDIO ports.
- Configure local MAC addresses.
- Display registered IPv4 addresses (in use and not in use). This display is supported on IBM Z platforms for QDIO ports.
- Display registered IPv4 or IPv6 Virtual MAC and VLAN ID associated with all OSA Ethernet features configured as QDIO Layer 2.
- Provide status information about an OSA port and its shared or exclusive use state.

This support is applicable to all OSA features on IBM Z platforms.
OSA/SF is not always required to customize an OSA feature. However, it can be used to gather operational information, to help on problem determination, to monitor, and to control ports. The OSA/SF Query function provides performance information about the OSA CHPIDs. As shown in Table 5-1 on page 73, OSA/SF is not required to configure the OSA features in any operating modes except OSE.

5.2 OSA capabilities

This section describes the capabilities that use the OSA-Express6S, OSA-Express5S, and OSA-Express4S features.

5.2.1 Virtual IP address

In the IP network environment, virtual IP address (VIPA) frees IP network hosts from dependence on a particular network attachment, allowing the establishment of primary and secondary paths through the network. VIPA is supported by all of the OSA features.

An IP address traditionally ties to a physical link at one end of a connection. If the associated physical link goes down, it is unreachable. The virtual IP address exists only in software and has no association to any physical link. The IP network stack is the destination IP address rather than the network attachment.

VIPA provides for multiple IP addresses to be defined to an IP network stack, allowing fault-tolerant, redundant backup paths to be established. Applications become insensitive to the condition of the network because the VIPAs are always active. This configuration enables users to route around intermediate points of failure in the network.

VIPA takeover and takeback

Because a VIPA is associated with an IP network stack and not a physical network attachment, it can be moved to any IP network stack within its network. If the IP network stack that the VIPA is on fails (because of an outage), the same VIPA can be brought up automatically on another IP network stack (VIPA takeover). This process allows users to reach the backup server and applications. The original session between the user and original server is not disrupted. After the failed IP network stack is restored, the same VIPA can be moved back automatically (VIPA takeback).

5.2.2 Primary/secondary router function

The primary/secondary router function enables an OSA port to forward packets with unknown IP addresses to an IP network stack for routing through another IP network interface, such as IBM HiperSockets or another OSA feature.

For an OSA port to forward IP packets to a particular IP network stack for routing to its destination, the PRIRouter must be defined on the DEVICE statement in the IP network profile.

If the IP network stack that has an OSA port that is defined as PRIRouter becomes unavailable, the following process occurs: A second IP network stack that is defined as the secondary router (SECRouter on the DEVICE statement in the IP network profile) receives the packets for unknown IP addresses.
For enhanced availability, the definition of one primary router and multiple secondary routers for devices on an OSD-type CHPID is supported. However, only one secondary router is supported for devices on an OSE-type CHPID.

**Important:** Sharing a single OSA port (with a single MAC address) can fail in load-balancing solutions. A workaround is to use generic routing encapsulation (GRE) or network address translation (NAT), which can have a negative effect on performance. Layer 3 virtual MAC is a function available on IBM Z platforms with OSA features that allows multiple MAC addresses on a single OSA port. For more information, see 5.2.16, “Layer 3 VMAC for z/OS” on page 93.

### 5.2.3 IPv6 support

Internet Protocol version 6 (IPv6) is supported by the OSA features when configured in QDIO mode. IPv6 is the protocol that is designed by the Internet Engineering Task Force (IETF) to replace Internet Protocol version 4 (IPv4). IPv6 provides improved traffic management in the following areas:

- **128-bit addressing**
  
  This improvement eliminates all practical limitations on global address ability. Private address space, along with the NATs used between a private intranet and the public internet, are no longer needed.

- **Simplified header formats**
  
  This improvement allows for more efficient packet handling and reduced bandwidth cost.

- **Hierarchical addressing and routing**
  
  This feature keeps routing tables small and backbone routing efficient by using address prefixes rather than address classes.

- **Improved support for options**
  
  This improvement changes the way IP header options are encoded, allowing more efficient forwarding and greater flexibility.

- **Address auto-configuration**
  
  This change allows stateless IP address configuration without a configuration server. In addition, IPv6 brings greater authentication and privacy capabilities through the definition of extensions, and integrated quality of service (QoS) through a traffic class byte in the header.
5.2.4 Large send for IP network traffic

Large send (also referred to as TCP segmentation offload) can improve performance by offloading TCP packet processing from the host to the OSA-Express features that are running in QDIO mode. Offload allows the host to send large blocks of data (up to 64 KB) directly to the OSA. The OSA then fragments those large blocks into standard Ethernet frames (1492 bytes) to be sent on the LAN (see Figure 5-2).

![Figure 5-2 Large send versus standard Ethernet and jumbo frame sizes](image)

Large send support reduces host processor use, so it returns processor cycles for application use and increases network efficiencies. For OSA-Express features, large send supports outbound IPv4 traffic only and applies solely to unicasts. Large send support for IPv6 packets applies to the OSA-Express7S, OSA-Express6S, OSA-Express5S, and OSA-Express4S features (CHPID types OSD and OSX) available on the Z platform.

**Note:** Large send for IPv6 packets is not supported for LPAR-to-LPAR packets.

5.2.5 VLAN support

Virtual local area network (VLAN) is supported by the OSA-Express features when configured in QDIO mode. This support is applicable to z/OS, z/VM, and Linux on IBM Z environments.

The IEEE standard 802.1q describes the operation of virtual bridged local area networks. A VLAN is defined to be a subset of the active topology of a local area network. The OSA features provide for the setting of multiple unique VLAN IDs per QDIO data device. They also provide for both tagged and untagged frames to flow from an OSA port. The number of VLANs supported is specific to the operating system.

---

1 For more information about the KVM hypervisor, see: [https://www.ibm.com/support/knowledgecenter/linuxonibm/liaaf/lnz_r_kvm.html](https://www.ibm.com/support/knowledgecenter/linuxonibm/liaaf/lnz_r_kvm.html)
VLANs facilitate easy administration of logical groups of stations, which can communicate as though they were on the same LAN. They also facilitate easier administration of moves, adds, and changes in members of these groups. VLANs are also designed to provide a degree of low-level security by restricting direct contact with a server to only the set of stations that comprise the VLAN.

With IBM Z platforms where multiple IP network stacks exist and potentially share one or more OSA features, VLAN support provides a greater degree of isolation (see Figure 5-3).

**Figure 5-3  VLAN support**

**VLAN support for z/OS**

Full VLAN support is offered for all OSA Ethernet features that are available on IBM Z platforms. The IBM z/OS Communications Server supports virtual local area network Identifications (VLAN IDs). With z/OS version 13.1, support is offered for up to 32 global VLAN IDs per OSA port for IPv4 and IPv6.

**VLAN support for z/VM**

z/VM uses the VLAN technology and conforms to the IEEE 802.1q standard. Support is offered for one global VLAN ID per OSA port for IPv4 and IPv6. Each port can be configured with a different VLAN ID.

**VLAN support for Linux on IBM Z**

VLAN support in a Linux on IBM Z environment is available for the OSA Ethernet features that operate in QDIO mode.

**VLAN support of GARP VLAN Registration Protocol**

Generic Attribute Registration Protocol (GARP) VLAN Registration Protocol (GVRP) is defined in the IEEE 802.1p standard for the control of IEEE 802.1q VLANs. It can be used to simplify networking administration and management of VLANs.
With GVRP support, an OSA-Express port can register or unregister its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change (see Figure 5-4).

Support of GVRP is exclusive to IBM Z. It is applicable to all of the OSA-Express features when in QDIO mode (CHPID type OSD), and is supported by z/OS and z/VM.

5.2.6 SNMP support for z/OS and Linux on IBM Z

SNMP is supported for all of the OSA features when configured in the QDIO mode (CHPID type OSD). The OSA features LIC includes the following support for the OSA SNMP subagent:

- **Get and GetNext requests**
  This support applies to all OSA features supported on IBM Z platforms.

- **dot3StatsTable**
  Ethernet data for dot3StatsTable applies to all of the Ethernet features that are supported on IBM Z platforms. It implements the SNMP EtherLike Management Information Base (MIB) module in RFC 2665, which provides statistics for Ethernet interfaces. These statistics can help in the analysis of network traffic congestion.

- **Performance data**
  This support applies to all of the OSA features supported on IBM Z platforms. The performance data reflects the OSA use.

- **Traps and Set**
  This support applies to all of the OSA features supported on IBM Z.
SNMP support for LAN channel station (LCS) applies to all of the OSA features that are supported on IBM Z, along with IP network applications only. It supports the same SNMP requests and alerts that are offered in QDIO mode (Get, GetNext, Trap, and Set) and is exclusive to the z/OS environment.

**Tip:** You can subscribe to the “OSA-Express Direct SNMP MIB module” document through Resource Link to receive email notifications of document changes.

OSA/SF is not required to manage SNMP data for the OSA features. An SNMP subagent exists on an OSA feature, which is part of a direct path between the z/OS or Linux on IBM Z master agent (IP network stacks) and an OSA-Express MIB.

The OSA features support an SNMP agent by providing data for use by an SNMP management application, such as IBM Tivoli NetView® for z/OS. This data is organized into MIB tables that are defined in the IP network enterprise-specific MIB and standard RFCs. The data is supported by the SNMP IP network subagent (see Figure 5-5).

![Figure 5-5  SNMP support, z/OS example](image)

### 5.2.7 IP network multicast and broadcast support

Multicast and broadcast support are part of the Internet Protocol Assist (IPA) function of the OSA feature.

**Multicast support**

For sending data to multiple recipients, OSA features support IP multicast destinations only in QDIO or IP Passthru mode.

**IP network broadcast support for z/OS, z/VM, and Linux on IBM Z**

Broadcast support is included for all of the OSA features when configured in QDIO mode and supporting the Routing Information Protocol (RIP) version 1. Broadcast is also supported for all of the OSA features when carrying IP network traffic and configured in the non-QDIO mode (LAN Channel Station in LCS mode).
A broadcast simultaneously transmits data to more than one destination. Messages are transmitted to all stations in a network (for example, a warning message from a system operator). The broadcast frames can be propagated through an OSA feature to all IP network applications that require broadcast support, including applications that use RIP V1.

5.2.8 Address Resolution Protocol cache management

The query and purge Address Resolution Protocol (ARP) enhancements are supported for all OSA features when configured in QDIO mode. The OSA feature maintains a cache of recently acquired IP-to-physical address mappings (or bindings). When the binding is not found in the ARP cache, a broadcast (an ARP request: “How can I reach you?”) to find an address mapping is sent to all hosts on the same physical network. Because a cache is maintained, ARP does not have to be used repeatedly, and the OSA feature does not have to keep a permanent record of bindings.

Query ARP table for IPv4 for Linux on IBM Z

The Query ARP table is supported by using Internet Protocol version 4 (IPv4). The IP network stack already has an awareness of Internet Protocol version 6 (IPv6) addresses.

Purge ARP entries in cache for IPv4 for z/OS and Linux on IBM Z

Purging of entries in the ARP cache is supported by using IPv4. The IP network stack already has an awareness of IPv6 addresses.

ARP takeover

ARP takeover provides the capability of switching OSA port operations from one OSA to another OSA running in the same mode in z/OS environments.

When a z/OS IP network is started in QDIO mode, it downloads all of the home IP addresses in the stack and stores them in each OSA feature to which it has a connection. This is a service of the QDIO architecture and occurs automatically only for OSD channels. For OSA ports set up as OSE channels (non-QDIO), multiple IP addresses must be defined in the OAT by using OSA/SF. The OSA then responds to ARP requests for its own IP address and for VIPAs.

If an OSA feature fails and a backup OSA available on the same network or subnetwork, IP network informs the backup OSA which IP addresses (real and VIPA) to take over, and the network connection is maintained. For this technique to work, multiple paths must be defined to the IP network stack. For example, MULTIPATH must be defined to the IPCONFIG statement of the IP network profile in z/OS.

ARP statistics

QDIO includes an IPA function, which gathers ARP data during the mapping of IP addresses to Media Access Control (MAC) addresses. CHPIDs that are defined as OSD maintain ARP cache information in the OSA feature (ARP offload). This data is useful in problem determination for the OSA feature.

5.2.9 IP network availability

There are several ways to ensure network availability if failure occurs at either the logical partition or the network connection level. Port sharing, redundant paths, and the use of primary and secondary ports all provide some measure of recovery. A combination of these can ensure network availability, regardless of the failing component.
5.2.10 Checksum offload support for z/OS and Linux on IBM Z

z/OS and Linux on IBM Z environments provide the capability of calculating and validating the Transmission Control Protocol/User Datagram Protocol (TCP/UDP) and IP header checksums. Checksums are used to verify the contents of files when transmitted over a network, such as these examples:

- OSA validates the TCP, UDP, and IP header checksums for inbound packets.
- OSA calculates the TCP, UDP, and IP header checksums for outbound packets.

Checksum offload is supported by all OSA Ethernet features when operating in QDIO mode.

By offloading checksum processing to the supporting OSA features, host system cycles are reduced, which can result in improved performance for most IPv4 packets.

**Note:** Linux on IBM Z supports only inbound checksum offload (inbound packets).

When checksum is offloaded, the OSA feature calculates the checksum. For OSA-Express3 feature, this function applies only to packets that go onto the LAN or come in from the LAN. When multiple IP stacks share an OSA port and an IP stack sends a packet to a next-hop IP address that is owned by another IP stack that shares the same OSA port, OSA sends the IP packet directly to the other IP stack without placing it on the LAN. Checksum offload does not apply to such IP packets.

The checksum offload is performed for IPv6 packets and IPv4 packets. This process occurs regardless of whether the traffic goes to the LAN, comes in from the LAN, or flows from one logical partition to another logical partition through the OSA feature.

Checksum offload for IPv6 packets is with CHPID type OSD on the Z platform. Checksum offload for LPAR-to-LPAR traffic in the z/OS environment is included in the OSA-Express7S, OSA-Express6S, OSA-Express5S, and OSA-Express4S design for both IPv4 and IPv6 packets.

Checksum offload support is available with z/OS and z/VM. For more information, see 5.3.3, “Software support” on page 116.

5.2.11 Dynamic LAN idle for z/OS

Dynamic LAN idle is exclusive to IBM Z platforms and applies to the OSA-Express features (QDIO mode). It is supported by z/OS.

Dynamic LAN idle reduces latency and improves networking performance by dynamically adjusting the inbound blocking algorithm. When enabled, the z/OS IP network stack adjusts the inbound blocking algorithm to best match the application requirements.

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 z/OS IP network stack dynamically detects the application requirements and makes the necessary adjustments to the blocking algorithm. The monitoring of the application and the blocking algorithm adjustments are made in real time, so they dynamically adjust the application’s LAN performance.
System administrators can authorize the z/OS IP network stack to enable a dynamic setting, which was previously a static setting. The z/OS IP network stack dynamically determines the best setting for the current running application, based on system configuration, system, inbound workload volume, processor use, traffic patterns, and related items.

5.2.12 QDIO optimized latency mode

QDIO optimized latency mode (OLM) can help improve performance for applications that have a critical requirement to minimize response times for inbound and outbound data.

OLM optimizes the interrupt processing according to direction:

- For inbound processing, the IP network stack looks more frequently for available data to process. This ensures that any new data is read from the OSA feature without requiring more program-controlled interrupts (PCIs).
- For outbound processing, the OSA features also look more frequently for available data to process from the IP network stack. Therefore, a Signal Adapter instruction (SIGA) is not required to determine whether more data is available.

OLM is supported by the Communications Server for z/OS with any OSA feature on the Z platform.

5.2.13 Layer 2 support

The OSA Ethernet features on IBM Z platforms can support two transport modes of the OSI model:

- Layer 2 (Link Layer or MAC Layer)
- Layer 3 (Network Layer)

The Layer 2 transport mode allows for communication with IP and non-IP protocols. OSA works with either z/VM IP network or Linux on IBM Z Layer 2 support that is running in a logical partition or as a z/VM guest.
The z/VM virtual switch can also be used to enable Layer 2 function for guest systems (see Figure 5-6).

![Figure 5-6  Layer 2 support for OSA](image)

The virtual switch uses both Layer 2 and Layer 3 support in the z/VM Control Program. For Layer 2 support, the z/VM Control Program owns the connection to the OSA feature and manages the MAC addresses and VLAN connectivity of the attached guest systems. The virtual switch automatically generates the MAC address and assignment to ensure that the z/VM guest systems are unique. MAC addresses can also be locally administered.

The virtual switch uses each guest system's unique MAC address to forward frames. Data is transported and delivered within Ethernet frames. This process provides the ability to transport both IP and non-IP frames (for example, NetBIOS and SNA) through the fabric that the virtual switch supports. Through the address-resolution process, each guest system’s MAC address becomes known to hosts on the physical side of the LAN segment. All inbound or outbound frames that pass through the OSA port have the guest system’s corresponding MAC address as the source or destination address.

The OSA Ethernet features can filter inbound frames by virtual local area network identification (VLAN ID, IEEE 802.1q), the Ethernet destination MAC address, or both. Filtering can reduce the amount of inbound traffic that is being processed by the operating system, which reduces processor use. Filtering by VLAN ID or MAC address can also enable you to isolate portions of your environment that have sensitive data to provide a degree of low-level security.

**Link aggregation for z/VM in Layer 2 mode**

Link aggregation is exclusive to IBM Z and is applicable to the OSA-Express features in Layer 2 mode when configured as CHPID type OSD (QDIO). Link aggregation is supported by z/VM.
z/VM virtual switch-controlled (VSWITCH-controlled) link aggregation (IEEE 802.3ad) allows you to do the following when the port is participating in an aggregated group and is configured in Layer 2 mode:

- Dedicate an OSA-Express port to the z/VM operating system
- Share an OSA-Express7S, OSA-Express6S, OSA-Express5S, or OSA-Express4S port within a z/VM operating system
- Share an OSA-Express7S, OSA-Express6S, OSA-Express5S, or OSA-Express4S port between any z/VM operating system that is running on the CPC

Link aggregation (trunking) allows you to combine multiple physical OSA-Express ports into a single logical link. This configuration allows increased throughput and nondisruptive failover if a port becomes unavailable.

Link aggregation for z/VM in Layer 2 mode functions in these ways:

- Aggregated links are viewed as one logical trunk and contain all of the VLANs that are required by the LAN segment.
- Link aggregation is between a VSWITCH and the physical network switch.
- Load balance communications are across multiple links in a trunk to prevent a single link from being overrun.
- Up to eight OSA-Express ports are supported in one aggregated link.
- OSA ports can be added or removed to provide on-demand bandwidth.
- It operates in full-duplex mode (send and receive) for increased throughput.

**Note:** Target links for aggregation must be of the same type (for example, Gigabit Ethernet to Gigabit Ethernet).

### 5.2.14 QDIO data connection isolation for z/VM

The QDIO data connection isolation function provides a higher level of security when sharing a OSA-Express port and virtual switch (VSWITCH) across multiple z/VM guest systems. The VSWITCH is a virtual network device that provides switching between OSA-Express ports and the connected guest systems.

Two modes allow for connection isolation: **Port isolation** and **Virtual Ethernet Port Aggregator (VEPA)**. Port isolation and VEPA are mutually exclusive.

**Port isolation**

z/VM allows you to disable guest system-to-guest system communication through a virtual switch, preserving each guest system’s ability to communicate with hosts or routers in the external network. When a virtual switch is defined to be isolated, all communication between the guest systems’ ports on that virtual switch (VSWITCH) are disabled.

An isolated virtual switch is unable to communicate directly with other logical partitions that share the OSA-Express port. Communications must be relayed through another network-based device, such as a router or firewall, in the external network.

As shown in Figure 5-7 on page 91, when in isolation mode, data traffic that is destined for a guest system port in the VSWITCH is blocked. However, traffic that is going to an external network device is sent to the OSA-Express port for delivery. The isolation options (ON or OFF) can be set by using the `SET VSWITCH ISOLATION` command.
Virtual Ethernet Port Aggregation

VEPA is part of the IEEE 802.1Qbg standard. It provides the capability of sending all attached guest systems’ data traffic to an external Ethernet switch for further processing. This mode does not allow any direct guest system-to-guest system communications through the VSWITCH. In tandem with the VSWITCH, the OSA-Express feature prevents any data traffic between the VSWITCH and any connected systems that share that OSA-Express port. Hence, the isolation is provided within both the VSWITCH and the OSA-Express feature. However, VEPA mode does allow data traffic from a guest system to be sent to a router or similar network-based devices and come back through the same VSWITCH to another guest system. This is known as hairpinning.

For a VSWITCH to enter VEPA mode, the external Ethernet switch must be in Reflective Relay mode.
As shown in Figure 5-8, when in VEPA mode, data traffic that is destined for a guest system in the VSWITCH is forced to go to an external Ethernet switch through an OSA-Express port for further processing. VEPA mode (ON or OFF) can be set by using the SET VSWITCH VEPA command.

QDIO data connection isolation is supported by all OSA-Express7S, OSA-Express6S features on z14 and, subsequently, all OSA-Express5S features on the z13, all OSA-Express4S features on the z13 and zEnterprise CPCs, and all OSA-Express3 features on the zEnterprise CPCs.

5.2.15 QDIO interface isolation for z/OS

Some environments require strict controls for routing data traffic between systems or nodes. In certain cases, the LPAR-to-LPAR capability of a shared OSA port can prevent such controls from being enforced. With interface isolation, internal routing can be controlled on an LPAR basis. When interface isolation is enabled, the OSA discards any packets that are destined for a z/OS LPAR that is registered in the OAT as isolated.
For example, in Figure 5-9, LPAR 1 has interface isolation enabled. Therefore, data traffic from LPAR 2 that is destined for LPAR 1 is dropped by the OSA.

QDIO interface isolation is supported by the Communications Server for z/OS with OSA features on the Z platform.

### 5.2.16 Layer 3 VMAC for z/OS

To help simplify the infrastructure and provide load balancing when a logical partition is sharing an OSA MAC address with another logical partition, each operating system instance can now have its own unique VMAC address. All IP addresses associated with an IP network stack are accessible by using their own VMAC address rather than sharing the MAC address of an OSA port. This applies to Layer 3 mode and to an OSA port shared among logical partitions.

This support has several benefits:

- Improves IP workload balancing
- Dedicates a Layer 3 VMAC to a single IP network stack
- Removes the dependency on GRE tunnels
- Improves outbound routing
- Simplifies configuration setup
- Allows z/OS to use a standard interface ID for IPv6 addresses
- Allows IBM WebSphere® Application Server content-based routing to support an IPv6 network
- Eliminates the need for PRIROUTER/SECROUTER function in z/OS
OSA Layer 3 VMAC for z/OS is applicable to OSA Ethernet features when configured as CHPID type OSD (QDIO).

**Note:** OSA layer 3 Virtual MAC support was first introduced with z/VM V4.4

### 5.2.17 Enterprise Extender

The Enterprise Extender (EE) function of z/OS Communications Server allows you to run SNA applications and data on IP networks and IP-attached clients. It can be used with any OSA feature that is running IP traffic. EE is a simple set of extensions to the open High-Performance Routing technology that integrates HPR frames into UDP/IP packets, which provide these advantages:

- SNA application connectivity by using an IP backbone support for:
  - SNA-style priority
  - SNA Parallel Sysplex
- Improved throughput and response times
- Compatible support for TCP and UDP traffic on the IP portion of the application traffic path (SNA/HPR and UDP/IP traffic can coexist on an EE connection)

The EE function is an IP network *encapsulation* technology. It carries SNA traffic from an endpoint over an IP network (for example, through the OSA port to the Communications Server) to another endpoint where it is *de-encapsulated* and presented to an SNA application.

EE requires APPN/HPR at the endpoints. To enable EE, you must configure the IP network stack with a virtual IP address and define an XCA major node. The XCA major node is used to define the PORT, GROUP, and LINE statements for the EE connections.

See the *Enterprise Extender Implementation Guide*, SG24-7359, for more information.

### 5.2.18 TN3270E Server

The TN3270E Server is supported by z/OS. It allows desktop users to connect through an IP network directly to SNA applications.

The following support is provided:

- Secure access using SSL and Client Authentication by using IBM Resource Access Control Facility (IBM RACF®)
- Over 64,000 sessions per server when using multiple ports
- Over 2000 transactions/second with subsecond response time
- Reconnect 16,000 sessions in less than a minute by using VIPA takeover support
- IP QoS using Service Policy server
- Host Print support
- Tivoli support provides IP visibility to VTAM
- Manage with your data center resources
- More robust than external TN3270 servers
z/OS Communications Server also supports a secure, RACF-based single sign-on logic called Express Logon Facility (ELF). ELF works with IBM TN3270 clients to securely authenticate the client, acquire a pass token, and then pass it on to the TN3270E server for replacement or submission to the application.

5.2.19 Adapter interruptions for QDIO

Linux on IBM Z and z/VM work together to provide performance improvements by using extensions to the QDIO architecture. Adapter interruptions, which were first added to IBM z/Architecture with HiperSockets, provide an efficient, high-performance technique for I/O interruptions. This technique reduces path lengths and overhead in both the host operating system and the adapter when using type OSD CHPID.

In extending the use of adapter interruptions to OSD (QDIO) channels, the programming overhead to process a traditional I/O interruption is reduced. This technique benefits OSA IP network support in Linux on IBM Z, z/VM, and z/VSE.

Adapter interruptions apply to all of the OSA features on a the Z platform when in QDIO mode (CHPID type OSD).

5.2.20 Inbound workload queuing

Inbound workload queuing (IWQ) helps reduce overhead and latency for inbound z/OS network data traffic and implements an efficient way for initiating parallel processing. This goal is achieved by using an OSA-Express feature in QDIO mode (CHPID types OSD and OSX) with multiple input queues and by processing network data traffic that is based on workload types.
Figure 5-10 shows the IWQ concept of using multiple inbound queues for different types of workloads (T1 through T4) compared to a single inbound queue.

The data from a specific workload type is placed in one of four input queues, and a process is created and scheduled to run on one of multiple processors, independent from the three other queues. This process greatly improves performance because IWQ can use the symmetric multiprocessor (SMP) architecture of the Z platform.

A primary objective of IWQ is to provide improved performance for business-critical interactive workloads by reducing contention that is created by other types of workloads. These types of z/OS workloads are identified and assigned to unique input queues:

- **z/OS Sysplex Distributor traffic**
  
  Network traffic that is associated with a distributed VIPA is assigned to a unique input queue, allowing the Sysplex Distributor traffic to be immediately distributed to the target host.

- **z/OS bulk data traffic**
  
  Network traffic that is dynamically associated with a streaming (bulk data) TCP connection is assigned to a unique input queue. This technique allows most of the data processing to be assigned the appropriate resources and isolated from critical interactive workloads.

- **Enterprise Extender traffic**
  
  Inbound Enterprise Extender traffic is differentiated and separated to a new input queue. This separation and processing provides improved scalability and performance for Enterprise Extender. It is supported on the Z platform with OSA features.
Chapter 5. Open Systems Adapter-Express

5.21 Network management: Query and display OSA configuration

As complex functions have been added to OSA, the job of the system administrator to display, monitor, and verify the specific current OSA configuration unique to each operating system has become more complex. The operating system can directly query and display the current OSA configuration information (similar to OSA/SF). z/OS and z/VM use this OSA capability with an IP network operator command, which allows the operator to monitor and verify the current OSA configuration. This feature improves the overall management, serviceability, and usability of the OSA features.

Display OSAINFO is possible on OSA-Express features (CHPID types OSD, OSM, and OSX).

5.3 Connectivity

The transmission medium and cabling requirements for the OSA ports depend on the OSA feature, OSA port type, and LAN. Acquiring cables and other connectivity items is the IBM client’s responsibility.

Removal of support for IEEE 802.3 Ethernet frame types: The zEC12 and zBC12 were the last IBM Z platforms to support IEEE 802.3 Ethernet frame types on OSA-Express QDIO interfaces in Layer 3 mode.

This statement applies to CHPID types OSD and OSX when they are used in Layer 3 mode. These OSA-Express CHPID types in Layer 3 mode are planned to support Ethernet DIX version 2 (DIX V2) exclusively on future IBM Z platforms.

OSA-Express non-QDIO (CHPID type OSE) supporting SNA, APPN, HPR with Link Station Architecture (LSA), TCP/IP Passthru environments with LAN channel station (LCS), and QDIO CHPID types OSD and OSX running in Layer 2 mode are not affected by this change.

OSA devices

The different types of OSA channels (CHPID types) require the following device types:

- OSA devices for QDIO (OSD) and non-QDIO (OSE) CHPID types
- 3270-X and 3287 devices (consoles) for the OSA-ICC (OSC) CHPID type
- 3745 and OSN device for the OSA for NCP (OSN)² CHPID type
- OSA devices for ensemble network connectivity (OSX) and ensemble management (OSM) CHPID types

OSA/SF requires one device (defined through HCD) to be associated with the OSA CHPID as device type OSAD (UNITADD=FE). OSA/SF uses this device to communicate with the OSA feature.

The OSA-Express Network Traffic Analyzer for z/OS requires one or more data paths devices for the OSAENTA trace interface, depending on the configuration.

**Multiple image facility**

Multiple image facility (MIF) enables OSA ports that are installed on IBM Z platforms to be shared across logical partitions. For more information, see Chapter 2, “Channel subsystem overview” on page 15.

**Spanned channels**

Spanning is the ability to configure channels to multiple channel subsystems. When defined that way, the channels can be transparently shared by any or all of the configured logical partitions, regardless of the channel subsystem to which the logical partition is configured.

OSA ports can be spanned across multiple channel subsystems (CSSes) on IBM Z platforms. See 2.1.2, “Multiple CSS concept” on page 17, for more information about spanned channels.

### 5.3.1 OSA features

Table 5-2 lists the OSA features that are supported on IBM Z platforms along with the maximum number of OSA ports that are supported by each system. For all optical links, the connector type is LC duplex unless otherwise specified. The electrical Ethernet cable for the OSA connectivity is through an RJ 45 jack. Details for each feature that is supported on the Z platform are provided in subsequent sections.

**Table 5-2 IBM Z OSA features**

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Feature code</th>
<th>Cable type</th>
<th>Maximum unrepeated distance</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express7S 25GbE SR</td>
<td>0429</td>
<td>MM 50 µm</td>
<td>70 m (2000) 100 m (4700)</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td>OSA-Express6S GbE LX</td>
<td>0422</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td>OSA-Express6S GbE SX</td>
<td>0423</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200) 220 m (160)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE LR</td>
<td>0424</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE SR</td>
<td>0425</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200) 220 m (160)</td>
<td></td>
</tr>
</tbody>
</table>

2 All IBM products supporting CHPID Type OSN are withdrawn from marketing as of 9 March 2015.
<table>
<thead>
<tr>
<th>Feature name</th>
<th>Feature code</th>
<th>Cable type</th>
<th>Maximum unrepeated distancea</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express6S</td>
<td>0426</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14, z14 ZR1</td>
</tr>
<tr>
<td>1000BASE-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S</td>
<td>0413</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>z14b, z14 ZR1b, z13, z13s, zEC12, zBC12</td>
</tr>
<tr>
<td>GbE LX</td>
<td></td>
<td>MCPc</td>
<td>550 m (500)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S</td>
<td>0414</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>z14b, z14 ZR1b, z13, z13s, zEC12, zBC12b</td>
</tr>
<tr>
<td>GbE SX</td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>220 m (160)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S 10</td>
<td>0415</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14b, z14 ZR1b, z13, z13s, zEC12, zBC12b</td>
</tr>
<tr>
<td>GbE LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S 10</td>
<td>0416</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>z14b, z14 ZR1b, z13, z13s, zEC12, zBC12b</td>
</tr>
<tr>
<td>GbE SR</td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>220 m (160)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S</td>
<td>0417</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14b, z14 ZR1b, z13, z13s, zEC12, zBC12b</td>
</tr>
<tr>
<td>1000BASE-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S</td>
<td>0404</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>z14 ZR1b, z13b, z13sb, zEC12, zBC12</td>
</tr>
<tr>
<td>GbE LX</td>
<td></td>
<td>MCPb</td>
<td>550 m (500)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S</td>
<td>0405</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>z14 ZR1b, z13b, z13sb, zEC12, zBC12</td>
</tr>
<tr>
<td>GbE SX</td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>220 m (160)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S 10</td>
<td>0406</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14 ZR1b, z13b, z13sb, zEC12, zBC12</td>
</tr>
<tr>
<td>GbE LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S 10</td>
<td>0407</td>
<td>MM 50 µm</td>
<td>300 m (2000)</td>
<td>z14 ZR1b, z13b, z13sb, zEC12, zBC12</td>
</tr>
<tr>
<td>GbE SR</td>
<td></td>
<td>MM 62.5 µm</td>
<td>82 m (500)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>33 m (200)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express4s</td>
<td>0408</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14 b, z14 ZR1b, z13b, z13sb, zEC12</td>
</tr>
<tr>
<td>1000BASE-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express3</td>
<td>3362</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>zEC12b, zBC12b</td>
</tr>
<tr>
<td>GbE LX</td>
<td></td>
<td>MCPb</td>
<td>550 m (500)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express3</td>
<td>3363</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>zEC12b, zBC12b</td>
</tr>
<tr>
<td>GbE SX</td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>220 m (166)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express3-2P</td>
<td>3373</td>
<td>MM 50 µm</td>
<td>550 m (500)</td>
<td>zBC12b</td>
</tr>
<tr>
<td>GbE SX</td>
<td></td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>220 m (166)</td>
<td></td>
</tr>
<tr>
<td>OSA-Express3 10</td>
<td>3370</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>zEC12b, zBC12b</td>
</tr>
<tr>
<td>GbE LR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Maximum number of OSA features

- On a z14 or z14 ZR1, the maximum combined number of OSA-Express features is 48 with any mix of OSA-Express7S, OSA-Express6S, OSA-Express5S (carry-forward only), and supported OSA Express4S (carry-forward only) features.
- On a z13 system, the maximum combined number of OSA-Express5S features and OSA-Express4S features (carry-forward only) is 48.
- On a z13s system, the maximum number of OSA-Express5S and OSA-Express4S combined features (carry-forward only) is 48.
- On a zEC12, the maximum OSA-Express3, OSA-Express4S, and OSA-Express 5S features depends on the mix of those features. On a zEC12, the maximum number of OSA-Express3 features is 24 (carry forward only). The maximum number of OSA-Express4S features alone is 48. The maximum number of OSA-Express3, OSA-Express4S, and OSA-Express5S that are mixed is determined based on a maximum of 48 PCHIDs.
- On a zBC12, the maximum number of OSA-Express3 features is 8 (carry forward only) or 16 with RPQ 8P2733 for the second I/O drawer (carry forward only). The maximum OSA-Express4S (carry forward only) and OSA-Express5S features is 48.

Table 5-3 shows the number of OSA I/O features supported on z14 systems.

<table>
<thead>
<tr>
<th>I/O feature: z14 / z14 ZR1</th>
<th>Ports per feature</th>
<th>Ports per CHPID</th>
<th>Max. number of Ports</th>
<th>CHPID definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express7S 25GbE SR</td>
<td>1</td>
<td>1</td>
<td>48</td>
<td>48 OSD, OSX</td>
</tr>
<tr>
<td>OSA-Express6S GbE LX/SX</td>
<td>2</td>
<td>2</td>
<td>96 / 96</td>
<td>48 / 48 OSD</td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE LR/SR</td>
<td>1</td>
<td>1</td>
<td>48 / 48</td>
<td>48 / 48 OSD, OSX</td>
</tr>
<tr>
<td>OSA-Express6S 1000BASE-T</td>
<td>2</td>
<td>2</td>
<td>96 / 96</td>
<td>48 / 48 OSE, OSD, OSC(^b), OSM</td>
</tr>
</tbody>
</table>

\(a\) Minimum fiber bandwidth in MHz*km for multimode fiber optic links is included in parentheses where applicable.

\(b\) Available on carry-forward only.

\(c\) Mode-conditioning patch (MCP) cables enable the 1 Gbps single mode features to connect to multimode fiber.
Table 5-4 shows the number of OSA I/O features supported on z13 systems.

Table 5-4  

<table>
<thead>
<tr>
<th>I/O feature: z14 / z14 ZR1</th>
<th>Ports per feature</th>
<th>Ports per CHPID</th>
<th>Max. number(^a) of Ports</th>
<th>Max. number(^a) of I/O slots</th>
<th>CHPID definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express5S GbE LX/SX</td>
<td>2</td>
<td>2</td>
<td>96 / 96</td>
<td>48 / 48</td>
<td>OSD</td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE LR/SR</td>
<td>1</td>
<td>1</td>
<td>48 / 48</td>
<td>48 / 48</td>
<td>OSD, OSX</td>
</tr>
<tr>
<td>OSA-Express5S 1000BASE-T</td>
<td>2</td>
<td>2</td>
<td>96 / 96</td>
<td>48 / 48</td>
<td>OSE, OSD, OSC(^b), OSM</td>
</tr>
<tr>
<td>OSA-Express4S GbE SX/LX</td>
<td>2</td>
<td>2</td>
<td>NA / 96</td>
<td>NA / 48</td>
<td>OSD</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE SR/LR</td>
<td>1</td>
<td>1</td>
<td>NA / 48</td>
<td>NA / 48</td>
<td>OSD, OSX</td>
</tr>
<tr>
<td>OSA-Express4S 1000BASE-T</td>
<td>2</td>
<td>2</td>
<td>96 / NA</td>
<td>48 / NA</td>
<td>OSE, OSD, OSC(^b), OSM</td>
</tr>
</tbody>
</table>

\(^a\) The maximum number OSA-Express that are mixed is determined based on a maximum of 48 PCHIDs.
\(^b\) OSA-ICC (OSC Channel) now supports Secure Sockets Layer. Up to 48 secure sessions per CHPID are supported (the overall maximum of 120 connections is unchanged).

Table 5-4  

<table>
<thead>
<tr>
<th>I/O feature: z13 and z13s systems supported OSA I/O features</th>
<th>Ports per feature</th>
<th>Ports per CHPID</th>
<th>Maximum number(^a) of Ports</th>
<th>Maximum number(^a) of I/O Slots</th>
<th>CHPID definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express4S GbE LX/SX</td>
<td>2</td>
<td>2</td>
<td>96</td>
<td>48</td>
<td>OSD</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE LR/SR</td>
<td>1</td>
<td>1</td>
<td>48</td>
<td>48</td>
<td>OSD, OSX</td>
</tr>
<tr>
<td>OSA-Express4S 1000BASE-T</td>
<td>2</td>
<td>2</td>
<td>96</td>
<td>48</td>
<td>OSE, OSD, OSC(^b), OSM</td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE LR/SR</td>
<td>1</td>
<td>1</td>
<td>48</td>
<td>48</td>
<td>OSD</td>
</tr>
<tr>
<td>OSA-Express5S GbE LX/SX</td>
<td>2</td>
<td>2</td>
<td>96</td>
<td>48</td>
<td>OSD</td>
</tr>
<tr>
<td>OSA-Express5S 1000BASE-T</td>
<td>2</td>
<td>2</td>
<td>96</td>
<td>48</td>
<td>OSE, OSD, OSC(^b), OSM</td>
</tr>
</tbody>
</table>

\(^a\) The maximum number OSA-Express4S that are mixed is determined based on a maximum of 48 PCHIDs.
\(^b\) OSA-ICC (OSC Channel) now supports Secure Sockets Layer (z13s and z13 GA2). Up to 48 secure sessions per CHPID are supported (the overall maximum of 120 connections is unchanged).
OSA-Express7S 25-Gigabit Ethernet Short Reach (25GbE SR) feature
Feature code 0429 is supported on the z14 M0x (MT 3906) and z14 ZR1 (MT 3907) servers and can be installed only in the PCIe I/O drawer and PCIe+ I/O drawer, respectively. It occupies one slot and has one port with an SFP transceiver and an LC duplex receptacle. This feature connects to a 25 Gbps Ethernet LAN through a 50-micron multi-mode fiber optic cable that is terminated with an LC duplex connector.

The following maximum unrepeated distances are supported:
- 70 meters (229 feet) on a 50-micron multimode (2000 MHz*km) fiber optic cable
- 100 meters (328 feet) on a 50-micron multimode (4700 MHz*km) fiber optic cable

The SFP allows a concurrent repair or replace action.

The OSA-Express7S 25GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express7S 25GbE SR supports 64b/66b encoding.

The one port of the OSA-Express7S 25GbE SR feature has one CHPID assigned, and can be defined as type OSD or OSX. For more information about modes of operation and supported traffic types, see 5.1.1, “Operating modes” on page 72.

The following Ethernet standards are applicable for the 25GBASE-SR (standard transmission scheme) feature:
- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.
- DIX version 2.

OSA-Express6S 10-Gigabit Ethernet Long Reach (10GbE LR) feature
Feature code 0424 is supported exclusively on the z14 server and can be installed only in the PCIe I/O drawer. It occupies one slot and has one port with an SFP transceiver and an LC duplex receptacle. This feature connects to a 10 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable that is terminated with an LC duplex connector and supports an unrepeated maximum distance of 10 km (6.2 miles). The SFP allows a concurrent repair or replace action.

The OSA-Express6S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express6S 10 GbE LR supports 64b/66b encoding, whereas GbE supports 8b/10 encoding, making auto-negotiation to any other speed impossible.

The one port of the OSA-Express6S 10 GbE LR feature has one CHPID assigned, and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-LR (standard transmission scheme) feature:
- IEEE 802.3 (ISO/IEC 802.3)
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing
- DIX version 2
OSA-Express6S 10-Gigabit Ethernet Short Reach (10GbE SR) feature

Feature code 0425 is supported exclusively on the z14 server and can be installed only in the PCIe I/O drawer. It occupies one slot and has one port with a small form-factor pluggable (SFP) transceiver and an LC duplex receptacle. This feature connects to a 10 Gbps Ethernet LAN through a 62.5-micron or 50-micron multimode fiber optic cable that is terminated with an LC duplex connector.

The following are the maximum supported unrepeated distances:
- 33 meters (108 feet) on a 62.5 micron multimode (200 MHz) fiber optic cable
- 82 meters (269 feet) on a 50-micron multimode (500 MHz) fiber optic cable
- 300 meters (984 feet) on a 50-micron multimode (2000 MHz) fiber optic cable

The SFP allows a concurrent repair or replace action.

The OSA-Express6S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express5S 10 GbE SR supports 64b/66b encoding, whereas GbE supports 8b/10 encoding, which makes auto-negotiation to any other speed impossible.

The one port of the OSA-Express6S 10 GbE SR feature has one CHPID assigned, and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-SR (standard transmission scheme) feature:
- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.
- DIX version 2.

OSA-Express6S Gigabit Ethernet Long Wavelength (GbE LX) feature

Feature code 0422 is supported on the z14 server and can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports with an SFP transceiver and an LC duplex receptacle. This feature connects to a 1 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable that is terminated with an LC duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles).

A multimode (62.5- or 50-micron) fiber optic cable can be used with these features. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link (see Table E-1 on page 200). Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet). The SFP allows a concurrent repair or replace action.

The two ports of the OSA-Express6S GbE LX feature share a channel path identifier (CHPID type OSD exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express6S GbE LX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
The following Ethernet standards are applicable for the 1000BASE-LX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express6S Gigabit Ethernet short wavelength (GbE SX) feature**

Feature code 0423 is supported on the z14 and can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports with an SFP transceiver and an LC duplex receptacle. This feature connects to a 1 Gbps Ethernet LAN through a 50-micron or 62.5-micron multimode fiber optic cable. The cable is terminated with an LC duplex connector over an unrepeated distance of 550 meters (for 50 µm fiber) or 220 meters (for 62.5 µm fiber). The SFP allows a concurrent repair or replace action.

The two ports of the OSA-Express6S GbE SX feature share a channel path identifier (CHPID type OSD exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express6S GbE SX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

The following Ethernet standards are applicable for the 1000BASE-SX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express6S 1000BASE-T Ethernet feature**

Feature code 0426 is exclusive to the z14 server and can be installed only in the PCIe I/O drawer. It occupies one slot in the PCIe I/O drawer and has two ports that connect to a 1000 Mbps (1 Gbps) or 100 Mbps Ethernet LAN. Each port has a Small Form-factor Pluggable (SFP) transceiver with an RJ-45 receptacle for cabling to an Ethernet switch. The SFP allows concurrent repair or replacement on each SFP. The RJ-45 receptacle must be attached by using EIA/TIA category 5 or 6 unshielded twisted pair (UTP) cable with a maximum length of 100 meters (328 feet).

The two ports of the OSA-Express6S 1000BASE-T feature have one CHPID assigned, so the two ports share one CHPID number. The use of both ports on a two-port CHPID requires support by the operating system.

The OSA-Express6S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed to default to auto-negotiation, the OSA port and the attached router or switch auto-negotiate the LAN speed between them and connect at the highest common performance speed of interoperation. If the attached Ethernet router or switch does not support automatic negotiation, the OSA port examines the signal that it is receiving and connects at the speed and full-duplex mode of the device at the other end of the cable.

You can choose any one of the following settings for the OSA-Express6S 1000BASE-T Ethernet feature port:

- Auto-negotiate
- 100 Mbps full-duplex
- 1000 Mbps full-duplex
If you are not using auto-negotiate, the OSA port attempts to join the LAN at the specified speed. If this setting does not match the speed and duplex mode of the signal on the cable, the OSA port does not connect.

LAN speed can be set explicitly by using OSA/SF or the OSA Advanced Facilities function of the HMC. The explicit settings override the OSA feature port’s ability to auto-negotiate with its attached Ethernet switch.

Each OSA-Express6S 1000BASE-T CHPID can be defined as CHPID type OSD, OSE, OSC, or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for this feature:
- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-ICC (CHPID type OSC) SSL Support**

TLS/SSL with Certificate Authentication was added to the OSC CHPID to provide a secure and validated method for connecting clients to the IBM Z host. Up to 48 secure sessions per CHPID can be defined (the overall maximum of 120 connections is unchanged).

**OSA-Express5S 10-Gigabit Ethernet Long Reach (GbE LR) feature**

Feature code 0415 is supported on the Z platform (carry forward only on z14 and z14 ZR1), and can be installed only in the PCIe I/O drawer. It occupies one slot and has one port with an SFP transceiver and an LC duplex receptacle. This feature connects to a 10 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable that is terminated with an LC duplex connector and supports an unrepeated maximum distance of 10 km (6.2 miles). The SFP allows a concurrent repair or replace action.

The OSA-Express5S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express5S 10 GbE LR supports 64b/66b encoding, whereas GbE supports 8b/10 encoding, making auto-negotiation to any other speed impossible.

The one port of the OSA-Express5S 10 GbE LR feature has one CHPID assigned and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-LR (standard transmission scheme) feature:
- IEEE 802.3 (ISO/IEC 802.3)
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing
- DIX version 2

---

3 Supported on z14. See the latest HMC driver level for feature support.
**OSA-Express5S 10-Gigabit Ethernet Short Reach (GbE SR) feature**

Feature code 0416 is supported on the Z platform (carry forward only on z14 and z14 ZR1), and can be installed only in the PCIe I/O drawer. It occupies one slot and has one port with an SFP transceiver and an LC duplex receptacle. This feature connects to a 10 Gbps Ethernet LAN through a 62.5-micron or 50-micron multimode fiber optic cable that is terminated with an LC duplex connector.

The following are the maximum supported unrepeated distances:

- 33 meters (108 feet) on a 62.5 micron multimode (200 MHz) fiber optic cable
- 82 meters (269 feet) on a 50-micron multimode (500 MHz) fiber optic cable
- 300 meters (984 feet) on a 50-micron multimode (2000 MHz) fiber optic cable

The SFP allows a concurrent repair or replace action.

The OSA-Express5S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express5S 10 GbE SR supports 64b/66b encoding, whereas GbE supports 8b/10 encoding, which makes auto-negotiation to any other speed impossible.

The one port of the OSA-Express5S 10 GbE SR feature has one CHPID assigned and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-SR (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing
- DIX version 2

**OSA-Express5S Gigabit Ethernet Long Wavelength (GbE LX) feature**

Feature code 0413 is supported on the Z platform (carry forward only on z14 and z14 ZR1), and can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports with an SFP transceiver and an LC duplex receptacle. This feature connects to a 1 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable that is terminated with an LC duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles).

A multimode (62.5- or 50-micron) fiber optic cable can be used with these features. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link (see Table F-1 on page 210). Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet). The SFP allows a concurrent repair or replace action.

The two ports of the OSA-Express5S GbE LX feature share a channel path identifier (CHPID type OSD exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express5S GbE LX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
The following Ethernet standards are applicable for the 1000BASE-LX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express5S Gigabit Ethernet short wavelength (GbE SX) feature**

Feature code 0414 is supported on the Z platform (carry forward only on z14 and z14 ZR1), and can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports with an SFP transceiver and an LC duplex receptacle. This feature connects to a 1 Gbps Ethernet LAN through a 50-micron or 62.5-micron multimode fiber optic cable. The cable is terminated with an LC duplex connector over an unrepeated distance of 550 meters (for 50 µm fiber) or 220 meters (for 62.5 µm fiber). The SFP allows a concurrent repair or replace action.

The two ports of the OSA-Express5S GbE SX feature share a channel path identifier (CHPID type OSD exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express5S GbE SX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

The following Ethernet standards are applicable for the 1000BASE-SX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express5S 1000BASE-T Ethernet feature**

Feature code 0417 is exclusive to the Z platform (carry forward only on z14 and z14 ZR1), and can be installed only in the PCIe I/O drawer. It occupies one slot in the PCIe I/O drawer and has two ports that connect to a 1000 Mbps (1 Gbps) or 100 Mbps Ethernet LAN. Each port has an SFP transceiver with an RJ-45 receptacle for cabling to an Ethernet switch. The SFP allows concurrent repair or replacement on each SFP. The RJ-45 receptacle must be attached by using EIA/TIA category 5 or 6 UTP cable with a maximum length of 100 meters (328 feet).

The two ports of the OSA-Express5S 1000BASE-T feature have one CHPID assigned, so the two ports share one CHPID number. The use of both ports on a two-port CHPID requires support by the operating system.

The OSA-Express5S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed to default to auto-negotiation, the OSA port and the attached router or switch auto-negotiate the LAN speed between them and connect at the highest common performance speed of interoperation. If the attached Ethernet router or switch does not support automatic negotiation, the OSA port examines the signal that it is receiving and connects at the speed and full-duplex mode of the device at the other end of the cable.

You can choose any one of the following settings for the OSA-Express5S 1000BASE-T Ethernet feature port:

- Auto-negotiate
- 100 Mbps full-duplex
- 1000 Mbps full-duplex
If you are not using auto-negotiate, the OSA port attempts to join the LAN at the specified speed. If this does not match the speed and duplex mode of the signal on the cable, the OSA port does not connect.

LAN speed can be set explicitly by using OSA/SF or the OSA Advanced Facilities function of the HMC. The explicit settings override the OSA feature port’s ability to auto-negotiate with its attached Ethernet switch.

Each OSA-Express5S 1000BASE-T CHPID can be defined as CHPID type OSD, OSE, OSC, OSN, or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for this feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-ICC (CHPID type OSC) SSL Support**

TLS/SSL with Certificate Authentication was added\(^4\) to the OSC CHPID to provide a secure and validated method for connecting clients to the IBM Z host. Up to 48 secure sessions per CHPID can be defined (the overall maximum of 120 connections is unchanged).

**OSA-Express4S 10-Gigabit Ethernet Long Reach (10 GbE LR) feature**

Feature code 0406 is supported on the zEnterprise CPCs and on the z14 Model ZR1, z13 and z13s as a carry-forward. It can be installed only in the PCIe I/O drawer. It occupies one slot and has one port that connects to a 10 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable. The cable is terminated with an LC duplex connector and supports an unrepeated maximum distance of 10 km (6.2 miles).

The OSA-Express4S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express4S 10 GbE LR supports 64B/66B encoding, whereas GbE supports 8B/10 encoding, making auto-negotiation to any other speed impossible.

The one port of the OSA-Express4S 10 GbE LR feature has one CHPID assigned and can be defined as type OSD or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-LR (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.
- DIX version 2.

**OSA-Express4S 10-Gigabit Ethernet Short Reach (10 GbE SR) feature**

Feature code 0407 is supported on the zEnterprise CPCs and on the IBM z14 Model ZR1, z13 and z13s as a carry-forward. It can be installed only in the PCIe I/O drawer. It occupies one slot and has one port that connects to a 10 Gbps Ethernet LAN through a 62.5-micron or 50-micron multimode fiber optic cable that is terminated with an LC duplex connector.

The following are the maximum supported unrepeated distances:

\(^4\) Supported on z14, z14 ZR1, z13, and z13s. See the latest HMC driver level for feature support.
The one port of the OSA-Express4S 10 GbE SR feature has one CHPID assigned and can be defined as type OSD or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-SR (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.
- DIX version 2.

OSA-Express4S Gigabit Ethernet Long Wavelength (GbE LX) feature

Feature code 0404 is supported on the zEnterprise CPCs and as a carry-forward on the IBM z14 Model ZR1, z13, and z13s as a carry forward. It can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports that connect to a 1 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable. The cable is terminated with an LC duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles). A multimode (62.5-micron or 50-micron) fiber optic cable can be used with these features. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link (see Table F-1 on page 210). Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet).

The two ports of the OSA-Express4S GbE LX feature share a channel path identifier (CHPID type OSD, exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express4S GbE LX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

The following Ethernet standards are applicable for the 1000BASE-LX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX Version 2

OSA-Express4S 10-Gigabit Ethernet Short Reach (10 GbE SR) feature

Feature code 0407 is supported on the zEnterprise CPCs and on the IBM z14 Model ZR1, z13 and z13s as a carry-forward. It can be installed only in the PCIe I/O drawer. It occupies one slot and has one port that connects to a 10 Gbps Ethernet LAN through a 62.5-micron or 50-micron multimode fiber optic cable that is terminated with an LC duplex connector.
The following are the maximum supported unrepeated distances:

- 33 meters (108 feet) on a 62.5-micron multimode (200 MHz) fiber optic cable
- 82 meters (269 feet) on a 50-micron multimode (500 MHz) fiber optic cable
- 300 meters (984 feet) on a 50-micron multimode (2000 MHz) fiber optic cable

The OSA-Express4S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express4S 10 GbE SR supports 64B/66B encoding, whereas GbE supports 8B/10 encoding, making auto-negotiation to any other speed impossible.

The one port of the OSA-Express4S 10 GbE SR feature has one CHPID assigned and can be defined as type OSD or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-SR (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.
- DIX version 2.

**OSA-Express4S Gigabit Ethernet Long Wavelength (GbE LX) feature**

Feature code 0404 is supported on the zEnterprise CPCs and as a carry-forward on the IBM z14 Model ZR1, z13 and z13s as a carry forward. It can be installed only in the PCIe I/O drawer. It occupies one slot and has two ports that connect to a 1 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable. The cable is terminated with an LC duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles). A multimode (62.5-micron or 50-micron) fiber optic cable can be used with these features. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link (see Table F-1 on page 210). Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet).

The two ports of the OSA-Express4S GbE LX feature share a channel path identifier (CHPID type OSD, exclusively). The use of both ports on a two-port CHPID requires support by the operating system. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The OSA-Express4S GbE LX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

The following Ethernet standards are applicable for the 1000BASE-LX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX Version 2

**OSA-Express4S 1000BASE-T Ethernet feature**

Feature code 0408 was introduced with zEC12 and can be installed only in the PCIe I/O drawer. It is supported on the IBM zBC12, and on IBM z14, z13, and z13s systems as a carry-forward only. It occupies one slot in the PCIe I/O drawer and has two ports that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle must be attached by using EIA/TIA category 5 or 6 UTP cable with a maximum length of 100 meters (328 feet).
The two ports of the OSA-Express4S 1000BASE-T feature have one CHPID assigned, so the two ports share one CHPID number. The use of both ports on a two-port CHPID requires support by the operating system.

The OSA-Express4S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed and duplex mode to default to automatic negotiation, the OSA port and the attached router or switch automatically negotiate the LAN speed and duplex mode settings. They connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

You can choose any one of the following settings for the OSA-Express4S 1000BASE-T Ethernet feature port:

- Auto-negotiate
- 10 Mbps half-duplex
- 10 Mbps full-duplex
- 100 Mbps half-duplex
- 100 Mbps full-duplex
- 1000 Mbps full-duplex

**Important:** The OSA-Express4S 1000BASE-T Ethernet feature is the last copper Ethernet feature to support half-duplex operation and a 10 Mbps link data rate.

If you are not using auto-negotiate, the OSA port attempts to join the LAN at the specified speed and duplex mode. If this does not match the speed and duplex mode of the signal on the cable, the OSA port does not connect.

LAN speed and duplex mode can be set explicitly by using OSA/SF or the OSA Advanced Facilities function of the HMC. The explicit settings override the OSA feature port’s ability to auto-negotiate with its attached Ethernet switch.

Each OSA-Express4S 1000BASE-T CHPID can be defined as CHPID type OSD, OSE, OSC, OSN, or OSM. See 5.1.1, “Operating modes” on page 72, for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for this feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX Version 2

**OSA-ICC (CHPID type OSC) SSL Support**

STLS/SSL with Certificate Authentication was added to the OSC CHPID to provide a secure and validated method for connecting clients to the IBM Z host. Up to 48 secure sessions per CHPID can be defined. The overall maximum of 120 connections is unchanged.

**OSA-Express3 10-Gigabit Ethernet Long Reach (10 GbE LR) feature**

Feature code 3370 is supported on the zEnterprise CPCs. It occupies one slot in the I/O cage or I/O drawer and has two ports that connect to a 10 Gbps Ethernet LAN through a 9-micron single-mode fiber optic cable. The cable is terminated with an LC duplex connector and supports an unrepeated maximum distance of 10 km (6.2 miles).

---

5 Supported on z14, z14 ZR1, z13, and z13s. See the latest HMC driver level for feature support.
The OSA-Express3 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express3 10 GbE LR supports 64B/66B encoding, whereas GbE supports 8B/10 encoding, making auto-negotiation to any other speed impossible.

The two ports of the OSA-Express3 10 GbE LR feature have two CHPIDs assigned and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-LR (standard transmission scheme):
- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.

**OSA-Express3 10-Gigabit Ethernet Short Reach (10 GbE SR) feature**

Feature code 3371 is supported on the zEnterprise CPC. It occupies one slot in the I/O cage or I/O drawer and has two ports that connect to a 10 Gbps Ethernet LAN through a 62.5-micron or 50-micron multimode fiber optic cable that is terminated with an LC duplex connector. The following are the maximum supported unrepeated distances:
- 33 meters (108 feet) on a 62.5-micron multimode (200 MHz) fiber optic cable
- 82 meters (269 feet) on a 50-micron multimode (500 MHz) fiber optic cable
- 300 meters (984 feet) on a 50-micron multimode (2000 MHz) fiber optic cable

The OSA-Express3 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. OSA-Express3 10 GbE LR supports 64B/66B encoding, whereas GbE supports 8B/10 encoding, making auto-negotiation to any other speed impossible.

The two ports of the OSA-Express3 10 GbE SR feature have two CHPIDs assigned and can be defined as type OSD or OSX. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 10GBASE-SR (standard transmission scheme) feature:
- IEEE 802.3 (ISO/IEC 802.3).
- Ethernet V2.0, including jumbo frames (those larger than 1518 bytes). Larger frame sizes increase efficiency for data intensive applications by reducing frame transmission processing.

**OSA-Express3 Gigabit Ethernet Long Wavelength (GbE LX) feature**

Feature code 3362 is supported on the zEnterprise CPCs. It occupies one slot in the I/O cage or I/O drawer and has four ports that connect to a 1 Gbps Ethernet LAN through a 9-micron single mode fiber optic cable. The cable is terminated with an LC duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles). A multimode (62.5-micron or 50-micron) fiber optic cable can be used with this feature. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link (see Table F-1 on page 210). Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet).
The four ports of the OSA-Express3 GbE LX feature have two CHPIDs assigned, so two corresponding ports share one CHPID number. The use of both ports on a two-port CHPID requires support by the operating system.

The OSA-Express3 GbE LX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

Each OSA-Express3 GbE LX port can be defined as CHPID type OSD or OSN. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for the 1000BASE-LX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express3 Gigabit Ethernet short wavelength (GbE SX) feature**

Feature code 3363 is supported on the zEnterprise CPCs. It occupies one slot in the I/O cage or I/O drawer and has four ports that connect to a 1 Gbps Ethernet LAN through a 50-micron or 62.5-micron multimode fiber optic cable. The cable is terminated with an LC duplex connector over an unrepeated distance of 550 meters (for 50 µm fiber) or 220 meters (for 62.5 µm fiber).

The four ports of the OSA-Express3 GbE SX feature have two CHPIDs assigned, so two corresponding ports share one CHPID number. The use of both ports on a two-port CHPID requires support by the operating system.

The OSA-Express3 GbE SX feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

Each OSA-Express3 GbE SX port can be defined as CHPID type OSD or OSN. See 5.1.1, “Operating modes” on page 72 for more about modes of operation and traffic types.

The following Ethernet standards are applicable for the 1000BASE-SX (standard transmission scheme) feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express3 1000BASE-T Ethernet feature**

Feature code 3367 is supported on the zEnterprise CPCs. It occupies one slot in the I/O cage or I/O drawer and has four ports that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle must be attached by using EIA/TIA Category 5 or 6 UTP cable with a maximum length of 100 meters (328 feet).

The four ports of the OSA-Express3 1000BASE-T feature have two CHPIDs assigned, so two corresponding ports share one CHPID number. Use of both ports on a two-port CHPID requires operating system support. When a CHPID is defined as OSM, only port 0 is usable.
The OSA-Express3 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If LAN speed and duplex mode are allowed to default to auto-negotiation, the OSA port and the attached router or switch auto-negotiate the LAN speed and duplex mode settings between them. They connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

You can choose any one of the following settings for the OSA-Express3 1000BASE-T Ethernet feature port:

- Auto-negotiate
- 10 Mbps half-duplex
- 10 Mbps full-duplex
- 100 Mbps half-duplex
- 100 Mbps full-duplex
- 1000 Mbps full-duplex

If auto-negotiate is not being used, the OSA port attempts to join the LAN at the specified speed and duplex mode. If this does not match the speed and duplex mode of the signal on the cable, the OSA port does not connect.

LAN speed and duplex mode can be set explicitly by using OSA/SF or the OSA Advanced Facilities function of the HMC. The explicit settings override the OSA feature port’s ability to auto-negotiate with the attached Ethernet switch.

Each OSA-Express3 1000BASE-T port can be defined as CHPID type OSD, OSE, OSC, OSN, or OSM. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for this feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

**OSA-Express3-2P 1000BASE-T Ethernet feature**

Feature code 3369 is supported on the zBC12 and occupies one slot in the I/O drawer. It has two ports that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle must be attached by using EIA/TIA category 5 or 6 UTP cable with a maximum length of 100 meters (328 feet).

The two ports of the OSA-Express3-2P 1000BASE-T feature have one CHPID assigned, so the two ports share one CHPID number. The use of both ports requires support by the operating system.

The OSA-Express3-2P 1000BASE-T feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed and duplex mode to default to auto-negotiation, the OSA port and the attached router or switch auto-negotiate the LAN speed and duplex mode settings between them. They connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

You can choose any one of the following settings for the OSA-Express3-2P 1000BASE-T feature port:

- Auto-negotiate
If auto-negotiate is not being used, the OSA port attempts to join the LAN at the specified speed and duplex mode. If this does not match the speed and duplex mode of the signal on the cable, the OSA port does not connect.

LAN speed and duplex mode can be set explicitly by using OSA/SF or the OSA Advanced Facilities function of the HMC. The explicit settings override the OSA-Express3-2p 1000BASE-T feature port’s ability to auto-negotiate with its attached Ethernet switch.

Each OSA-Express3-2P 1000BASE-T port can be defined as CHPID type OSD, OSE, OSC, or OSN. See 5.1.1, “Operating modes” on page 72 for details about modes of operation and supported traffic types.

The following Ethernet standards are applicable for this feature:

- IEEE 802.3 (ISO/IEC 802.3)
- DIX version 2

### 5.3.2 OSA function support

Table 5-5 lists the functions that are supported based on the OSA feature.

<table>
<thead>
<tr>
<th>Function</th>
<th>OSA-Express4S and OSA-Express5S</th>
<th>OSA-Express6S</th>
<th>OSA-Express7S 25GbE SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jumbo frame support (8992-byte frame size)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Network Traffic Analyzer for z/OS^a</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>QDIO Diagnostic Synchronization for z/OS^a</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>640 IP network (with priority queues disabled)^a</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Virtual IP address (VIPA)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Primary/secondary router function</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Internet Protocol version 6 (IPv6)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Large send support for IPv4</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Large send support for IPv6</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>VLAN (IEEE 802.1q)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>VLAN support of GVRP (IEEE 802.1 p)^a</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SNMP support for z/OS and Linux on IBM Z^a</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
### 5.3.3 Software support

For more information about the operating systems that are supported on IBM Z platforms, see this website.

Note: Certain functions might require specific levels of an operating system, program temporary fixes (PTFs), or both. That information is provided when necessary within this chapter.
Consult the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

5.3.4 Resource Measurement Facility

Resource Measurement Facility (RMF) reporting (z/OS only) supports the OSA features. It can capture performance data for these features:

- Microprocessor use (per logical partition image, if it applies)
- Physical Peripheral Component Interconnect (PCI) bus use
- Bandwidth per port (both read and write directions) per logical partition image

For example, with this support, possible bandwidth bottlenecks and root causes can be analyzed.

5.4 Summary

OSA features provide direct LAN connectivity as integrated features of the IBM Z platforms. This feature brings the strengths of IBM Z and z/Architecture to the modern network environment.

Table 5-6 summarizes the OSA features for the different modes of operation and maximum addressing ranges that are supported by IBM Z platforms.

<table>
<thead>
<tr>
<th>Addresses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IP addresses per channel path (IPv4, IPv6, VIPA)</td>
<td>4096</td>
</tr>
<tr>
<td>Multicast addresses (IPv4 + IPv6)</td>
<td>16384</td>
</tr>
<tr>
<td>ARP table size</td>
<td>16384</td>
</tr>
<tr>
<td>MAC addresses</td>
<td>2048(^a)</td>
</tr>
</tbody>
</table>

\(^a\): 4096 on the z14.

<table>
<thead>
<tr>
<th>Non-QDIO (OSE)(^b)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subchannels per IP network link</td>
<td>2</td>
</tr>
<tr>
<td>IP network stacks per channel path</td>
<td>120</td>
</tr>
<tr>
<td>SNA PUs per port</td>
<td>4096</td>
</tr>
<tr>
<td>Subchannels per channel path</td>
<td>240</td>
</tr>
<tr>
<td>CUs per channel path</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QDIO (OSD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subchannels per IP network link</td>
<td>3</td>
</tr>
<tr>
<td>IP network stacks per channel path</td>
<td>640(^c)</td>
</tr>
<tr>
<td>Subchannels per channel path</td>
<td>1920(^c)</td>
</tr>
<tr>
<td>CUs per channel path</td>
<td>16</td>
</tr>
</tbody>
</table>

\(^b\): 1000BASE-T feature only.

\(^c\): \text{architecture specific}.
5.5 References

For more information about the OSA features and configuration, see:

- *OSA-Express Implementation Guide*, SG24-5948
- *Communications Server: SNA Network Implementation Guide*, SC31-8563
- *Communications Server: IP Configuration*, SC31-8513
- *IBM Communication Controller Migration Guide* SG24-6298
- *IBM Communication Controller for Linux on System z V1.2.1 Implementation Guide*, SG24-7223
- *Enterprise Extender Implementation Guide*, SG24-7359

Moreover, if multiple priority for queues is enabled, the maximum is reduced to 160 IP network stacks/480 devices.
OSA-Express Console support

This chapter describes the IBM Open Systems Adapter-Express Integrated Console Controller (OSA-ICC) function of the OSA-Express6S, OSA-Express5S, OSA-Express4S, and OSA-Express3 1000BASE-T features.

The OSA-ICC function supports TN3270 console (TN3270E), local non-SNA DFT 3270 emulation, and 328x printer emulation. This emulation support for console session connections is integrated in IBM Z platforms. It can help reduce cost and complexity by eliminating the requirement for external console controllers, such as the IBM 2074 and IBM 3174 controllers.

The following topics are covered in this chapter:
- A description of the OSA-ICC
- Migration considerations
- Minimum software support
6.1 Description

OSA-ICC support is a no-charge function that is included in Licensed Internal Code (LIC) on the Z platform. With z14 and z14 ZR1, ICC is available only by using the OSA-Express6S 1000BASE-T feature. For z13, z13s, zEC12, and zBC12, it is available through the OSA-Express5S, OSA-Express4S, and OSA-Express3 1000BASE-T Ethernet features.

The OSA-ICC supports Ethernet-attached TN3270E consoles and provides a system console function at IPL time and operating systems support for multiple logical partitions. Console support can be used by IBM z/OS, IBM z/VM, IBM z/VSE, and IBM z/Transaction Processing Facility (z/TPF) software.

The OSA-ICC also supports local non-SNA DFT 3270 and 328x printer emulation for IBM Time Sharing Option Extensions (TSO/E), IBM CICS®, IBM Information Management System (IMS™), or any other 3270 application that communicates through VTAM. For z14 with FC 0034, 3215 emulation is provided for z/TPF 1.1. For previous server generations, 3215 emulation for z/TPF 1.1 is provided by way of RPQ 8P2339.

With the 1000BASE-T Ethernet features, the OSA-ICC is configured on a port-by-port basis, by using the channel path identifier (CHPID) type OSC. When the CHPID shares two ports the following server definition rules apply:

- Each physical port is defined to a unique IP network port number.
- Different subnets are defined for each physical port host IP.
- There is a single defined common gateway router with an interface on each IP subnet. Only one of the IP network ports defines the IP default gateway.

Each CHPID can support up to 120 console session connections. These connections can be shared among logical partitions by using the multiple image facility (MIF) and can be spanned across multiple channel subsystems (CSSes).

**OSA-ICC (CHPID type OSC) SSL Support**

TLS/SSL with Certificate Authentication was added\(^1\) to the OSC CHPID to provide a secure and validated method for connecting clients to the IBM Z host. Up to 48 secure sessions per CHPID can be defined (the overall maximum of 120 connections unchanged).

**OSA-ICC enhancements with HMC 2.14.1**

The following OSA-ICC enhancements were introduced with HMC 2.14.1:

- The IPv6 communications protocol is supported by OSA-ICC 3270 so that clients can comply with regulations that require all computer purchases to support IPv6.
- The supported TLS protocol levels for the OSA-ICC 3270 client connection can be now specified. Supported protocol levels are: TLS 1.0, TLS 1.1, and TLS 1.2
- In addition to a single certificate for all OSA-ICC PCHIDs in the system, separate and unique OSA-ICC 3270 certificates are now supported for the benefit of customers who host workloads across multiple business units or data centers where cross-site coordination is required. Customers can avoid interruption of all the TLS connections at the same time when having to renew expired certificates.

\(^1\) Supported on z13 and z13s. See the latest HMC driver level for feature support.
Figure 6-1 shows an example of the OSA-Express Integrated Console Controller in a single system configuration.

Figure 6-1   Example of an OSA-ICC configuration

The following features are included in base support for the OSA-ICC:
- Local and remote session connectivity, depending on the provided security and networking environment
- Local and remote connections for configuration changes by using security features of the IBM Z Hardware Management Console environment
- Coexistence in configurations with IBM 2074 and IBM 3174 console controllers

6.2 Connectivity

IBM Z platforms have base Licensed Internal Code support for the OSA-ICC function. At least one 1000BASE-T Ethernet feature must be installed.

The Hardware Management Console (HMC) or Support Element (SE) is used to create the configuration source file for the OSA-ICC CHPID and for operation and diagnosis.

The OSA-Express6S 1000BASE-T (FC 0417) and OSA-Express5S 1000BASE-T (FC 0426) Ethernet features support the OSC CHPID. FC 0426 is supported only on z14 server. FC 0417 is supported on the Z platform (carry forward only on z14 and z14 ZR1). The feature has two ports with one CHPID associated to the two ports.

The OSA-Express4S 1000BASE-T Ethernet feature (FC 0408) supports the OSC CHPID. FC 0408 is supported on the Z14, z14 ZR1, z13, and z13s as a carry forward feature, and on the zEC12 and zBC12. It has two ports, with one CHPID associated to the two ports.
The OSA-Express3 1000BASE-T Ethernet features (FC 3367 and FC 3369) support the OSC CHPID as carried forward features. FC 3367 has four ports, with one CHPID associated to each of the two corresponding ports and is supported on the zEC12 and zBC12. FC 3369 has two ports, with one CHPID associated to the two ports and supported on the zBC12.

### 6.3 Software support

For more information about the operating systems that are supported on IBM Z platforms, see this website.

**Note:** Certain functions might require specific levels of an operating system, program temporary fixes (PTFs), or both. That information is provided when necessary within this chapter.

Consult the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

### 6.3.1 TN3270E emulation

The TN3270E emulation program that is used with the OSA-ICC must comply with the TN3270E (TN3270 Enhancements) protocol (RFC 2355), which allows for the following functions:

- The ability to request that a connection is associated with a specific 3270 LU or pool name.
- Access to the SNA bind information. Until it receives the bind information, the session runs at the SSCP-LU level.
- Handling of SNA positive and negative responses.
- Universal support of the 3270 ATTN and SYSREQ keys.
- Support for the emulation of the 328x class of printer.

IBM Personal Communications V5.6 and later supports TN3270 Extensions that are used for OSA-ICC support. Personal Communications is a component of the IBM Access Client Package for Multiplatforms program suite. For more information, see this web page.

### 6.4 Summary

These are the key characteristics of the Open Systems Adapter-Express Integrated Console Controller (OSA-ICC) function:

- A no-charge function that is integrated in z14, z14 ZR1, z13, z13s, and zEC12, and zBC12 LIC.
- Support in IBM z13 and zEnterprise CPCs by using the OSA-Express4S and OSA-Express3 1000BASE-T features.\(^2\)
- Each OSC CHPID can support up to 120 console session connections, be shared among logical partitions, and be spanned across multiple CSSes.

---

\(^2\) OSA-Express2 is not available on z13, zEC12, and zBC12. OSA-Express3S is available only on zEC12 and zBC12 as a carry-forward.
Chapter 6. OSA-Express Console support

- Support for TN3270 Enhancements (TN3270E) and local non-SNA DFT 3270 emulation by using Ethernet-attached workstations.
- Console support that can be used by z/OS, z/VM, z/VSE, and z/TPF.
- Local non-SNA DFT 3270 and 328x printer emulation support that can be used by TSO/E, CICS, IMS, or any other 3270 application that communicates through VTAM.
- TLS/SSL with Certificate Authentication was added³ to the OSC CHPID to provide a secure and validated method for connecting clients to the IBM Z host. Up to 48 secure sessions per CHPID can be defined (the overall maximum of 120 connections is unchanged).

The OSA-ICC function provides several potential benefits:

- **Simplicity:**
  - Uses installed 1000BASE-T Ethernet for both console sessions and networking traffic
  - External controllers are no longer needed

- **Scalable capacity:**
  - Facilitates addition of partitions and operations support pools
  - Can be shared by up to 85 logical partitions on z14 and z13
  - Can be shared by up to 40 partitions on z14 ZR1 and z13s
  - Can be shared by up to 60 logical partitions on zEC12
  - Can be shared by up to 30 logical partitions on zBC12

- **Improved availability**
  - Can enable *lights out* operation
  - Hot plug OSA availability characteristics
  - OSA features are a highly integrated component of the IBM Z platform, with the reliability, availability, and serviceability characteristics inherent in IBM Z

- **Low operating cost versus external console controller:**
  - Power, cooling, cabling, and floor space requirements are minimized
  - No rack necessary

### 6.5 References

For more information about the OSA-ICC function, see the following publications:


³ Supported on z13 and z13s. See the latest HMC driver level for feature support.
Shared Memory Communications

Shared Memory Communications (SMC) on IBM Z platforms is a technology that can improve throughput by accessing data faster with less latency, while reducing CPU resource consumption compared to traditional TCP/IP communications. Furthermore, applications do not need to be modified to gain the performance benefits of SMC.

This chapter includes the following topics:

- SMC overview
- SMC over Remote Direct Memory Access
- SMC over Direct Memory Access
- Software support
- Reference material
7.1 SMC overview

SMC is a technology that allows two peers to send and receive data by using system memory buffers that each peer allocates for its partner’s use. Two types of SMC protocols are available on the IBM Z platform:

- **SMC - Remote Direct Memory Access (SMC-R)**
  
  SMC-R is a protocol for Remote Direct Memory Access (RDMA) communication between TCP socket endpoints in logical partitions (LPARs) in different systems. SMC-R runs over networks that support RDMA over Converged Ethernet (RoCE). It enables existing TCP applications to benefit from RDMA without requiring modifications. SMC-R provides dynamic discovery of the RDMA capabilities of TCP peers and automatic setup of RDMA connections that those peers can use.

- **SMC - Direct Memory Access (SMC-D)**
  
  SMC-D implements the same SMC protocol that is used with SMC-R to provide highly optimized intra-system communications. Where SMC-R uses RoCE for communicating between TCP socket endpoints in separate systems, SMC-D uses Internal Shared Memory (ISM) technology for communicating between TCP socket endpoints in the same system. ISM provides adapter virtualization (virtual functions (VFs)) to facilitate the intra-system communications. Hence, SMC-D does not require any additional physical hardware (no adapters, switches, fabric management, or PCIe infrastructure). Therefore, significant cost savings can be achieved when using the ISM for LPAR-to-LPAR communication within the same IBM Z platform.

**Important:** Both SMC-R and SMC-D require an existing TCP link between the images that are configured to use the SMC protocol.

Both SMC protocols use shared memory architectural concepts, eliminating TCP/IP processing in the data path, yet preserving TCP/IP quality of service (QoS) for connection management purposes.

7.1.1 Remote Direct Memory Access

RDMA is primarily based on InfiniBand (IFB) technology. It has been available in the industry for many years. In computing, RDMA is direct memory access from the memory of one computer to that of another without involving either one's operating system (kernel). This permits high-throughput, low-latency network transfer, which is commonly used in massively parallel computing clusters.

There are two key requirements for RDMA, as shown in Figure 7-1 on page 127:

- A reliable lossless network fabric (LAN for Layer 2 in data center network distance)
- An RDMA capable NIC and Ethernet fabric
RoCE uses existing Ethernet fabric (switches with Global Pause enabled as defined by the IEEE 802.3x port-based flow control) and requires RDMA-capable NICs (RNICs), such as 25GbE RoCE Express2, 10GbE RoCE Express2, and 10GbE RoCE Express features.

### 7.1.2 Direct Memory Access

Direct Memory Access (DMA) uses a virtual PCI adapter that is called ISM rather than an RNIC as with RDMA. The ISM interfaces are associated with IP interfaces (for example, HiperSockets or OSA-Express) and are dynamically created, automatically started and stopped, and are auto-discovered.

The ISM does not use queue pair (QP) technology like RDMA. Therefore, links and link groups based on QPs (or other hardware constructs) are not applicable to ISM. SMC-D protocol has a design concept of a logical point-to-point connection called an SMC-D link.

The ISM uses a virtual channel identifier (VCHID) similar to HiperSockets for addressing purposes. Figure 7-2 shows the SMC-D LPAR-to-LPAR communications concept.
7.2 SMC over Remote Direct Memory Access

SMC-R is a protocol that allows TCP sockets applications to transparently use RDMA. SMC-R defines the concept of the SMC-R link, which is a logical point-to-point link using reliably connected queue pairs between TCP/IP stack peers over a RoCE fabric. An SMC-R link is bound to a specific hardware path, meaning a specific RNIC on each peer.

SMC-R is a hybrid solution (see Figure 7-3) for the following reasons:

- It uses TCP connection to establish the SMC-R connection.
- Switching from TCP to “out of band” SMC-R is controlled by a TCP option.
- SMC-R information is exchanged within the TCP data stream.
- Socket application data is exchanged through RDMA (write operations).
- TCP connection remains to control the SMC-R connection.
- This model preserves many critical existing operational and network management features of an IP network.

![Figure 7-3 Dynamic transition from TCP to SMC-R](image)

The hybrid model of SMC-R uses the following key existing attributes:

- Follows standard IP network connection setup.
- Dynamically switches to RDMA.
- TCP connection remains active (idle) and is used to control the SMC-R connection.
- Preserves critical operational and network management IP network features:
  - Minimal (or zero) IP topology changes.
  - Compatibility with TCP connection-level load balancers.
- Preserves existing IP security model (for example, IP filters, policy, VLANs, or SSL).
- Minimal network administrative or management changes.

- Changing host application software is not required, so all host application workloads can benefit immediately.

7.2.1 SMC-R connectivity

IBM released with IBM z14 a new generation of RoCE, RoCE Express2. This generation includes 25GbE RoCE Express2 (FC 0430) and 10GbE RoCE Express2 (FC 0412) features. The previous generation, 10GbE RoCE Express (FC 0411), can be used as carry forward to the z14.

The key difference between the two features is the number of VFs that is supported per port:

- The 10GbE RoCE Express (FC 0411) has two ports. On z14, z14 ZR1, z13, and z13s, the feature supports up to 31 VFs by using both ports. On zEC12 and zBC12, the feature supports only one port and no virtualization.

- The 25GbE and 10GbE RoCE Express2 features have two ports and can support up to 63 VFs per port (126 per feature) on z14 and 31 VFs per port (62 per feature) on z14 ZR1.

25GbE RoCE Express2

IBM z14 GA2 introduces RoCE Express2 25GbE (FC 0430), which is based on the existing RoCE Express2 generation hardware and provides two 25GbE physical ports. The feature requires 25GbE optics. The following maximum number of features can be installed:

- Eight in z14
- Four in z14 ZR1

For 25GbE RoCE Express2, the PCI Function IDs (FIDs) are now associated with a specific (single) physical port (that is port 0 or port 1).

10GbE RoCE Express2

The 10GbE RoCE Express2 feature (FC 0412) is an RDMA-capable network interface card. It provides a technology refresh for RoCE on IBM Z. Most of the technology updates are related to internal aspects of the RoCE. It provides two physical 10 GbE ports (no change from the previous generation). The maximum number of features can be installed is:

- Eight in z14
- Four in z14 ZR1

On z14, both ports are supported. For 10GbE RoCE Express2, the PCI Function IDs (FIDs) are now associated with a specific (single) physical port (that is port 0 or port 1).

10GbE RoCE Express

The 10GbE RoCE Express feature (FC 0411) is an RDMA-capable network interface card. It is supported on the Z platform, and installs in the PCIe I/O drawer. Each feature has one PCIe adapter. The maximum number of features can be installed is:

- Eight in z14
- Four in z14 ZR1
- 16 features in the z13, z13s, zEC12, and zBC12
On z14, z14 ZR1, z13, and z13s, both ports are supported. Only one port per feature is supported on zEC12 and zBC12. The 10GbE RoCE Express feature uses a short reach (SR) laser as the optical transceiver and supports use of a multimode fiber optic cable that terminates with an LC duplex connector. Both point-to-point connection and switched connection with an enterprise-class 10 GbE switch are supported.

On z14, z14 ZR1, z13, and z13s, the 10GbE RoCE Express feature can be shared across 31 partitions (LPAR). However, if the 10GbE RoCE Express feature is installed in a zEC12 or zBC12, the feature cannot be shared across LPARs, but instead is defined as reconfigurable.

**Connectivity**

**25GbE RoCE Express2**

The 25GbE RoCE Express2 feature requires Ethernet switch with 25GbE support. Switch port must support 25GbE (negotiation down to 10GbE is not supported). The switches must meet the following requirements:

- Pause frame function enabled, as defined by the IEEE 802.3x standard
- Priority flow control (PFC) disabled
- No firewalls, no routing, and no intraensemble data network (IEDN)

The maximum supported unrepeated distance for 25GbE RoCE Express2 feature, point-to-point, is 100 meters (328 feet).

A customer-supplied cable is required. The following types of cables can be used for connecting the port to the selected 25GbE switch or to the 25GbE RoCE Express2 feature on the attached IBM Z platform:

- OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km, terminated with an LC duplex connector (up to 70 meters, or 229.6 feet)
- OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km, terminated with an LC duplex connector (up to 100 meters, or 328 feet)

**10GbE RoCE Express2 and 10GbE RoCE Express**

The 10GbE RoCE Express2 and 10GbE RoCE Express features are connected to 10 GbE switches, which must meet the following requirements:

- Pause frame function enabled, as defined by the IEEE 802.3x standard
- Priority flow control (PFC) disabled
- No firewalls, no routing, and no intraensemble data network (IEDN)

The maximum supported unrepeated distance, point-to-point, is 300 meters (984 feet).

A customer-supplied cable is required. Three types of cables can be used for connecting the port to the selected 10 GbE switch or to the RoCE Express feature on the attached IBM Z platform:

- OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km, terminated with an LC duplex connector (up to 300 meters, or 984 feet)
- OM2 50-micron multimode fiber optic cable that is rated at 500 MHz-km, terminated with an LC duplex connector (up to 82 meters, or 269 feet)
- OM1 62.5-micron multimode fiber optic cable that is rated at 200 MHz-km, terminated with an LC duplex connector (up to 33 meters, or 108 feet)
7.3 SMC over Direct Memory Access

SMC-D is a protocol that allows TCP socket applications to transparently use ISM. It is a hybrid solution (see Figure 7-4 on page 131) and has the following features:

- It uses a TCP connection to establish the SMC-D connection.
- The TCP connection can be either through OSA Adapter or IQD HiperSockets.
- A TCP option (SMCD) controls switching from TCP to out-of-band SMC-D.
- The SMC-D information is exchanged within the TCP data stream.
- Socket application data is exchanged through ISM (write operations).
- The TCP connection remains to control the SMC-D connection.
- This model preserves many critical existing operational and network management features of TCP/IP.

The hybrid model of SMC-D uses these key existing attributes:

- It follows the standard TCP/IP connection setup.
- The hybrid model switches to ISM (SMC-D) dynamically.
- The TCP connection remains active (idle) and is used to control the SMC-D connection.
- The hybrid model preserves the following critical operational and network management TCP/IP features:
  - Minimal (or zero) IP topology changes.
  - Compatibility with TCP connection-level load balancers.
  - Preservation of the existing IP security model, such as IP filters, policies, virtual LANs (VLANs), and Secure Sockets Layer (SSL).
  - Minimal network administration and management changes.
Host application software is not required to change, so all host application workloads can benefit immediately.

The TCP path can be either through an OSA-Express port or HiperSockets connection.

7.4 Software support

This section provides software support information for SMC-R and SMC-D.

Note: The SMC-R and SMC-D are hybrid protocols that have been designed and implemented to speed up TCP communication without making changes to applications.

SMC-R uses the RoCE Express features for providing (direct memory access) data transfer under the control of an established TCP/IP connection (OSA). SMC-D can use also an established TCP/IP connection over Hipersockets.

For Linux on Z, IBM is working with its Linux distribution partners to include support in future distribution releases. At the time of this writing (October 2018), SLES 12 SP3 includes support for Linux-to-Linux communication as “Tech Preview.”

For more information about RoCE features support, contact the distribution owners.

7.4.1 SMC-R

SMC-R with RoCE provides high-speed communications performance across physical processors. It helps all TCP-based communications across z/OS LPARs that are in different central processor complexes (CPCs).

These are some typical communication patterns:

- Optimized Sysplex Distributor intra-sysplex load balancing.
- IBM WebSphere Application Server Type 4 connections to remote IBM Db2, IBM Information Management System, and IBM CICS instances.
- IBM Cognos®-to-Db2 connectivity.
- CICS-to-CICS connectivity through Internet Protocol interconnectivity (IPIC).

Currently, IBM z/OS version 2.1 or later with program temporary fixes (PTFs) is the only OS that supports SMC-R protocol with RoCE. Also, consider the following factors:

- No rollback to previous z/OS releases
- Requires IOCP 3.4.0 or later input/output configuration program
- MOre PTF is required to support 25GbE RoCE Express2

z/VM V6.4 and later provide support for guest use of the RoCE Express feature of IBM Z. This feature allows guests to use RoCE for optimized networking. More PTFs are required to support 25GbE RoCE Express2.
7.4.2 SMC-D

SMC-D has the following prerequisites:

- z14, z14 ZR1, z13, or z13s with HMC/SE for ISM vPCI functions.
- At least two z/OS V2.2 or later LPARs in the same IBM Z platform with required service installed:
  - SMC-D can only communicate with another z/OS V2.2 or later instance, and peer hosts must be in the same ISM PNet.
  - SMC-D requires an IP Network with access that uses OSA-Express or HiperSockets, which has a defined PNet ID that matches the ISM PNet ID.
- If running as a z/OS guest under z/VM, z/VM 6.4 and later is required for guest access to RoCE (Guest Exploitation only).

Note: SMC (existing architecture) cannot be used in the following circumstances:

- Peer hosts are not within the same IP subnet and VLAN.
- TCP traffic requires IPSec or the server uses FRCA.

IBM is working with IBM Linux distribution partners to include support in future IBM Linux on IBM Z distribution releases.

7.5 Reference material

- Draft IETF RFC for SMC-R
- Performance information
- IBM z/OS V2R2 Communications Server TCP/IP Implementation Volume 1: Base Functions, Connectivity, and Routing, SG24-8360
- IBM z13 Technical Guide, SG24-8251
- IBM z13s Technical Guide, SG24-8294
- IBM z14 Technical Guide, SG24-8451
- IBM z14 ZR1 Technical Guide, SG24-8651
Chapter 8. HiperSockets technology

This chapter presents a high-level overview of the IBM HiperSockets capabilities as they pertain to the IBM Z platform.

This chapter includes the following sections:

- Description
- Connectivity
- Summary
- References
8.1 Description

The HiperSockets function, also known as internal queued direct input/output (iQDIO) or internal QDIO, is an integrated function of the Licensed Internal Code (LIC) of the Z platform. It provides an attachment to high-speed logical LANs with minimal system and network overhead.

HiperSockets provides internal virtual LANs that act like IP networks within the IBM Z platform. Therefore, HiperSockets provides the fastest IP network communication between consolidated Linux, IBM z/VM, IBM z/VSE, and IBM z/OS virtual servers on a Z platform.

The virtual servers form a virtual LAN. Using iQDIO, the communication between virtual servers is through I/O queues set up in the system memory of the Z platform. Traffic between the virtual servers is passed at memory speeds. For more information about the number of HiperSockets that are available for each Z platform, see 8.2, “Connectivity” on page 147.

This LIC function, coupled with supporting operating system device drivers, establishes a higher level of network availability, security, simplicity, performance, and cost effectiveness than is available when connecting single servers or logical partitions together by using an external IP network.

HiperSockets is supported by the following operating systems:

- All in-service z/OS releases
- All in-service z/VM releases
- All in service z/VSE releases
- Linux on IBM Z and the KVM hypervisor host

**Note:** Throughout this chapter, we provide Operating System (OS) support information for the functions described. Not every OS supports all features. The OS support information is provided for IBM z/OS, IBM z/VM, IBM z/VSE, IBM z/TPF, and Linux on IBM Z (supported distributions).

Regarding KVM hypervisor, KVM support is provided by Linux distribution partners. Refer to your distribution’s documentation regarding KVM support for the Z platform.

8.1.1 HiperSockets benefits

Using the HiperSockets function has several benefits:

- HiperSockets eliminates the need to use I/O subsystem operations and the need to traverse an external network connection to communicate between logical partitions in the same Z platform.
- HiperSockets offers significant value in server consolidation, connecting many virtual servers in the same Z platform. It can be used rather than certain coupling link configurations in a Parallel Sysplex cluster. All of the consolidated hardware servers can be eliminated, along with the cost, complexity, and maintenance of the networking components that connect them.
- Consolidated servers that must access data on the Z platform can do so at memory speeds, bypassing all of the network overhead and delays.
- HiperSockets can be customized to accommodate varying traffic sizes. In contrast, LANs such as Ethernet and Token Ring have a maximum frame size that is predefined by their architecture. With HiperSockets, a maximum frame size can be defined according to the traffic characteristics transported for each of the possible HiperSockets virtual LANs.
Because there is no server-to-server traffic outside of the Z platform, a much higher level of network availability, security, simplicity, performance, and cost-effectiveness is achieved, compared to servers that communicate across an external LAN. For example:

- Because the HiperSockets feature has no external components, it provides a secure connection. For security purposes, servers can be connected to different HiperSockets. All security features, such as firewall filtering, are available for HiperSockets interfaces, as they are for other IP network interfaces.

- HiperSockets looks like any other IP network interface. Therefore, it is apparent to applications and supported operating systems.

- HiperSockets can also improve IP network communications within a sysplex environment when the DYNAMICXCF facility is used.

- HiperSockets integration with the intraensemble data network (IEDN) function extends the reach of the HiperSockets network outside of the CPC to the entire ensemble, where it appears as a single Layer 2 network.

### 8.1.2 Server integration with HiperSockets

Many data center environments are multi-tiered server applications, with various middle-tier servers that surround the z14, z14 ZR1, z13, z13s, or zEnterprise data and transaction server. Interconnecting multiple servers affects the cost and complexity of many networking connections and components. The performance and availability of the interserver communication depends on the performance and stability of the set of connections. The more servers that are involved, the greater the number of network connections and complexity to install, administer, and maintain.

Figure 8-1 shows two configurations. The configuration on the left shows a server farm that surrounds an IBM Z platform, with its corporate data and transaction servers. This configuration has a great deal of complexity involved in the backup of the servers and network connections. This environment also results in high administration costs.

![Figure 8-1  Server consolidation](image-url)
Consolidating that mid-tier workload onto multiple Linux virtual servers that run on a Z platform requires a reliable, high-speed network for those servers to communicate over. HiperSockets provides that. In addition, those consolidated servers also have direct high-speed access to database and transaction servers that are running under z/OS on the same Z platform.

This configuration is shown on the right side in Figure 8-1 on page 137. Each consolidated server can communicate with others on the Z platform through HiperSockets. In addition, the external network connection for all servers is concentrated over a few high-speed OSA-Express interfaces.

### 8.1.3 HiperSockets function

HiperSockets implementation is based on the OSA-Express QDIO protocol. Therefore, HiperSockets is also called internal QDIO (iQDIO). The LIC emulates the link control layer of an OSA-Express QDIO interface.

Typically, before a packet can be transported on an external LAN, a LAN frame must be built. The MAC address of the destination host or router on that LAN must be then inserted into the frame. HiperSockets does not use LAN frames, destination hosts, or routers. IP network stacks are addressed by inbound data queue addresses rather than MAC addresses.

The z14, z14 ZR1, z13, z13s, or zEnterprise LIC maintains a lookup table of IP addresses for each HiperSockets function. This table represents a virtual LAN. At the time that an IP network stack starts a HiperSockets device, the device is registered in the IP address lookup table with its IP address and its input and output data queue pointers. If an IP network device is stopped, the entry for this device is deleted from the IP address lookup table.

HiperSockets copies data synchronously from the output queue of the sending IP network device to the input queue of the receiving IP network device by using the memory bus to copy the data through an I/O instruction.

The controlling operating system that performs I/O processing is identical to OSA-Express in QDIO mode. The data transfer time is similar to a cross-address space memory move, with hardware latency close to zero. For total elapsed time for a data move, the operating system I/O processing time must be added to the LIC data move time.

HiperSockets operations run on the processor where the I/O request is initiated by the operating system. HiperSockets starts write operations. The completion of a data move is indicated by the sending side to the receiving side with a signal adapter (SIGA) instruction. Optionally, the receiving side can use dispatcher polling rather than handling SIGA interrupts. The I/O processing is performed without using the system assist processor (SAP). This implementation is also called thin interrupt.
The data transfer is handled much like a cross-address space memory move that uses the memory bus, not the IBM Z I/O bus. Therefore, HiperSockets does not contend with other system I/O activity in the system. Figure 8-2 shows the basic operation of HiperSockets.

The HiperSockets operational flow consists of five steps:

1. Each IP network stack registers its IP addresses in a server-wide common address lookup table. There is one lookup table for each HiperSockets virtual LAN. The scope of each LAN is the logical partitions that are defined to share the HiperSockets IQD channel-path identifier (CHPID).

2. The address of the IP network stack’s receive buffers is appended to the HiperSockets queues.

3. When data is being transferred, the send operation of HiperSockets performs a table lookup for the addresses of the sending and receiving IP network stacks and their associated send and receive buffers.

4. The sending virtual server copies the data from its send buffers into the target virtual server’s receive buffers (in system memory).

5. The sending virtual server optionally delivers an interrupt to the target IP network stack. This optional interrupt uses the thin interrupt support function of the Z platform. This feature means that the receiving virtual server looks ahead, detecting and processing inbound data. This technique reduces the frequency of real I/O or external interrupts.

**Hardware assists**

A complementary virtualization technology that includes the following features is available for Z platform:

- QDIO Enhanced Buffer-State Management (QEBSM)
  
  Two hardware instructions that are designed to help eliminate the overhead of hypervisor interception.
Host Page-Management Assist (HPMA)

An interface to the z/VM main storage management function designed to allow the hardware to assign, lock, and unlock page frames without z/VM hypervisor assistance.

These hardware assists allow a cooperating guest operating system to start QDIO operations directly to the applicable channel, without interception by the z/VM operating system. This process improves performance. Support is integrated in IBM Z. However, always check the appropriate subsets of the 3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE Preventive Service Planning (PSP) buckets before implementation.

8.1.4 Supported functions

This section describes additional functions that are supported by HiperSockets technology.

**Broadcast support**

Broadcasts are supported across HiperSockets on Internet Protocol version 4 (IPv4) for applications. Applications that use the broadcast function can propagate the broadcast frames to all IP network applications that are using HiperSockets. This support is applicable to Linux, z/OS, and z/VM environments.

**VLAN support**

Virtual local area networks (VLANs), IEEE standard 802.1q, are supported by Linux on z Systems and z/OS version 1.8 or later for HiperSockets. VLANs can reduce overhead by allowing networks to be organized by traffic patterns rather than physical location. This enhancement permits traffic flow on a VLAN connection both over HiperSockets and between HiperSockets and OSA-Express Ethernet features.

**IPv6 support**

HiperSockets supports Internet Protocol version 6 (IPv6). IPv6 is the protocol that was designed by the Internet Engineering Task Force (IETF) to replace IPv4 to help satisfy the demand for additional IP addresses.

The support of IPv6 on HiperSockets (CHPID type IQD) is available on Z platform, and is supported by z/OS and z/VM. IPv6 support is available on the OSA-Express6S, OSA-Express5S, OSA-Express4S, and OSA-Express3 features in the z/OS, z/VM, and Linux on z System environments.

Support of guests is expected to be apparent to z/VM if the device is directly connected to the guest (pass through).

**HiperSockets Network Concentrator**

Traffic between HiperSockets and OSA-Express can be transparently bridged by using the HiperSockets Network Concentrator. This technique does not require intervening network routing overhead, thus increasing performance and simplifying the network configuration. This goal is achieved by configuring a connector Linux system that has HiperSockets and OSA-Express connections that are defined to it.

The HiperSockets Network Concentrator registers itself with HiperSockets as a special network entity to receive data packets that are destined for an IP address on the external LAN through an OSA port. The HiperSockets Network Concentrator also registers IP addresses to the OSA feature on behalf of the IP network stacks by using HiperSockets, thus providing inbound and outbound connectivity.
HiperSockets Network Concentrator support uses the next-hop IP address in the QDIO header, rather than a Media Access Control (MAC) address. Therefore, VLANs in a switched Ethernet fabric are not supported by this support. IP network stacks that use only HiperSockets to communicate with no external network connection see no difference, so the HiperSockets support and networking characteristics are unchanged.

To use HiperSockets Network Concentrator unicast and multicast support, a Linux distribution is required. You also need s390-tools. They are available at this page of the IBM developerWorks® website.

**HiperSockets Layer 2 support**
The IBM HiperSockets feature supports two transport modes on the Z platform:

- Layer 2 (link layer)
- Layer 3 (Network and IP layer)

HiperSockets is protocol-independent and supports the following traffic types:

- Internet Protocol (IP) version 4 or version 6
- Non-IP (such as AppleTalk, DECnet, IPCX, NetBIOS, SNA)

Each HiperSockets device has its own Layer 2 MAC address and allows the use of applications that depend on a Layer 2 address, such as DHCP servers and firewalls. LAN administrators can configure and maintain the mainframe environment in the same fashion as they do in other environments. This feature eases server consolidation and simplifies network configuration.

The HiperSockets device automatically generates a MAC address to ensure uniqueness within and across logical partitions and servers. MAC addresses can be locally administered, and the use of group MAC addresses for multicast and broadcasts to all other Layer 2 devices on the same HiperSockets network is supported. Datagrams are delivered only between HiperSockets devices that use the same transport mode (for example, Layer 2 with Layer 2 and Layer 3 with Layer 3).

A HiperSockets device can filter inbound datagrams by VLAN identification, the Ethernet destination MAC address, or both. This feature reduces the amount of inbound traffic, which leads to lower processor use by the operating system.

As with Layer 3 functions, HiperSockets Layer 2 devices can be configured as primary or secondary connectors or multicast routers that enable high-performance and highly available Link Layer switches between the HiperSockets network and an external Ethernet.

HiperSockets Layer 2 support is available on Z platform with Linux on IBM Z and by z/VM guest use.

**HiperSockets multiple write facility**
HiperSockets performance has been increased by allowing streaming of bulk data over a HiperSockets link between logical partitions. The receiving partition can process larger amounts of data per I/O interrupt. The improvement is apparent to the operating system in the receiving partition. Multiple writes with fewer I/O interrupts reduce processor use of both the sending and receiving logical partitions and is supported in z/OS.
**zIIP-Assisted HiperSockets for large messages**
In z/OS, HiperSockets is enhanced for IBM Z Integrated Information Processor (zIIP) use. Specifically, the z/OS Communications Server allows the HiperSockets Multiple Write Facility processing of large outbound messages that originate from z/OS to be performed on zIIP.

z/OS application workloads that are based on XML, HTTP, SOAP, Java, and traditional file transfer can benefit from zIIP enablement by lowering general-purpose processor use.

When the workload is eligible, the HiperSockets device driver layer processing (write command) is redirected to a zIIP, which unblocks the sending application.

**HiperSockets Network Traffic Analyzer**
HiperSockets Network Traffic Analyzer (HS NTA) is a function available in the z14, z14 ZR1, z13, z13s, zEC12, and zBC12 LIC. It can make problem isolation and resolution simpler by allowing Layer 2 and Layer 3 tracing of HiperSockets network traffic.

HiperSockets NTA allows Linux on IBM Z to control tracing of the internal virtual LAN. It captures records into host memory and storage (file systems) that can be analyzed by system programmers and network administrators by using Linux on IBM Z tools to format, edit, and process the trace records.

A customized HiperSockets NTA rule enables you to authorize an LPAR to trace messages only from LPARs that are eligible to be traced by the NTA on the selected IQD channel.

HS NTA rules can be set up on the Support Element (SE). There are four types of rules for the HS NTA:
- Tracing is disabled for all IQD channels in the system (the default rule).
- Tracing is disabled for a specific IQD channel.
- Tracing is allowed for a specific IQD channel. All LPARs can be set up for NTA, and all LPARs are eligible to be traced by an active Network Traffic Analyzer.
- Customized tracing is allowed for a specific IQD channel.

**HiperSockets Completion Queue**
The HiperSockets Completion Queue function enables HiperSockets to transfer data synchronously, if possible, and asynchronously if necessary. This process combines ultra-low latency with more tolerance for traffic peaks. With the asynchronous support, during high volume situations, data can be temporarily held until the receiver has buffers available in its inbound queue. This feature provides end-to-end performance improvement for LPAR to LPAR communication.

The HiperSockets Completion Queue function is supported on the Z platform, and requires at a minimum:
- z/OS version 1.13
- Linux on IBM Z distributions:
  - Red Hat Enterprise Linux (RHEL) version 6.2
  - SUSE Linux Enterprise Server version 11, SP2
  - Ubuntu: 16.04 LTS
- z/VSE version 5.1.1
- z/VM V6.4 provides guest exploitation support
HiperSockets Completion Queue is supported by Linux on IBM Z through AF_IUCV socket communication. Fast Path to Linux in a Linux LPAR requires the HiperSockets Completion Queue function of the IBM Z platform.

**HiperSockets integration with the IEDN**
The zEnterprise systems provide the capability to integrate HiperSockets connectivity to the IEDN. This feature extends the reach of the HiperSockets network outside the CPC to the entire ensemble, where it appears as a single Layer 2 network.

Within each CPC that is a member of an ensemble, a single IQDIO CHPID for HiperSockets can be defined to provide connectivity to the IEDN. The designated IQD CHPID is configured by using a channel parameter in the hardware configuration definition (HCD), which enables the internal Queued Direct I/O extensions (IQDX) function of HiperSockets. The IQDX function is a channel function and is not a new CHPID type. When the IQDX function is configured, the single IQD CHPID is integrated with the IEDN.

The support of HiperSockets integration with the IEDN function is available on z/OS Communications Server 1.13 and z/VM 6.2. z/VM V6.3 and later does not support the IEDN function.

In a z/OS environment, HiperSockets connectivity to the IEDN is referred to as the z/OS Communications Server IEDN-enabled HiperSockets function.

In an IBM z/OS environment, HiperSockets integration with the IEDN function provides the following benefits:

- Combines the existing high-performance attributes of HiperSockets for intra-CPC communications with the secure access control, virtualization, and management functions of the IEDN that are provided by the Unified Resource Manager.
- Converges OSA connectivity and HiperSockets connectivity into a single logical network interface.
- Eliminates the HiperSockets configuration tasks within z/OS Communications Server and the Unified Resource Manager.
- Simplifies z/OS movement by eliminating or minimizing reconfiguration tasks.
- Enables sysplex network traffic to use HiperSockets connectivity for VIPAROUTE processing over OSA-Express for IBM Z BladeCenter Extension (zBX) (CHPID type OSX) interfaces.
(CHPID type OSX) interfaces. Figure 8-3 shows a sample configuration of HiperSockets integration with IEDN.

**HiperSockets virtual switch bridge support**

The z/VM virtual switch is enhanced to transparently bridge a guest virtual machine network connection on a HiperSockets LAN segment. This bridge allows a single HiperSockets guest virtual machine network connection to also directly communicate with either of these points:

- Other guest virtual machines on the virtual switch
- External network hosts, through the virtual switch OSA UPLINK port

**Note:** IBM z/VM 6.2\(^a\) or later, IP network, and Performance Toolkit APARs are required for this support.

\(^a\) Earliest z/VM version in support at the time of this writing (October 2018) is 6.4

A HiperSockets channel alone can provide only intra-CPC communications. The HiperSockets bridge port allows a virtual switch to connect IBM z/VM guests by using real HiperSockets devices. This feature provides the ability to communicate with hosts that are external to the CPC. A single IP address and virtual machine network connection can be used to communicate over the internal and external segments of the LAN. The fact that any particular destination address might be on the local HiperSockets channel or outside of the CPC is apparent to the bridge-capable port.

Incorporating the HiperSockets channel into the flat Layer 2 broadcast domain through OSD or OSX adapters simplifies networking configuration and maintenance. The virtual switch HiperSockets bridge port eliminates the need to configure a separate next-hop router on the HiperSockets channel to provide connectivity to destinations that are outside of a HiperSockets channel. This configuration avoids the need to create routes for this internal route in all hosted servers and the extra hop of a router to provide the Layer 3 routing functions.
Figure 8-4 shows an example of a bridged HiperSockets configuration.

The virtual switch HiperSockets bridge support expands the use cases and capabilities of the HiperSockets channel to include the following items:

- Full-function, industry-standard robust L2 bridging technology.
- Single NIC configuration, which simplifies network connectivity and management.
- No guest configuration changes are required for use (apparent to guest OS).
- Live Guest Relocation (LGR) of guests with real HiperSockets bridge-capable IQD connections within and between bridged CPCs.
- Option of including a HiperSockets connected guest within the IEDN network to allow communications with other LPARs and zBXs that are part of an ensemble.
- Support for the network management attributes of the IEDN.
- No limit on the number of z/VM LPARs that can participate in a bridged HiperSockets LAN.
- Ability to create a single broadcast domain across multiple CPCs (Cross CPC bridged HiperSockets channel network).
- Highly available network connection to the external network, provided by the z/VM virtual switch by default.
Figure 8-5 shows a sample Cross CPC HiperSockets LAN configuration. This configuration enables the creation of a single broadcast domain across CPCs and HiperSockets channels at the same time delivering a highly available configuration on both CPCs.

In this configuration, VSwitch B in LPAR B and VSwitch D in LPAR D are the active bridge ports that provide external connectivity between the external bridged IQD channel in CPC OPS1 and the external IQD channel in CPC OPS2. This flat Layer 2 LAN essentially joins or extends the HiperSockets LAN between CPCs across the external Ethernet network VSwitch UPLINK port and VSwitch D UPLINK port.

**IBM z/VSE Fast Path to Linux**

Fast Path to Linux allows z/VSE IP network applications to communicate with an IP network stack on Linux without using an IP network stack on z/VSE. Fast Path to Linux in an LPAR requires that the HiperSockets Completion Queue function is available on Z platforms. The Fast Path to Linux function is supported starting with z/VSE 5.1.1 (version 5, release 1.1).
Figure 8-6 shows a sample configuration of z/VSE and the Fast Path to Linux function. z/VSE applications can directly communicate with Linux IP network through the HiperSockets without involving the IP network stack of z/VSE.

**z/VSE z/VM IP Assist (VIA)**

The Z platform processors provide an appliance called z/VSE z/VM IP Assist (VIA), that can be used by LFP. With VIA the processor provides the functionality to communicate to the TCP/IP network without having a Linux distribution installed. This appliance is used for z/VSE running under z/VM.

**z/VSE Network Appliance (VNA)**

As of June 30, 2017, for z13 and later servers, the z/VSE Network appliance (VNA) that can be used by LFP. VNA is deployed in a Secure Service Container LPAR and does not require any Linux on IBM Z to be installed. It provides TCP/IP communication for z/VSE running in an LPAR.

Both VIA and VNA are provided for z/VSE 5.1 and later.

### 8.2 Connectivity

HiperSockets has no external components or external network. There is no internal or external cabling. The HiperSockets data path does not go outside of the z14, z14 ZR1, z13, z13s, or zEnterprise platform. In addition, HiperSockets integration with the IEDN enables CPC to CPC communication in the same ensemble.

HiperSockets is not allocated a CHPID until it is defined. It does not occupy an I/O cage, an I/O drawer, or a PCIe I/O drawer slot. HiperSockets cannot be enabled if all of the available CHPIDs on the Z platform have been used. Therefore, HiperSockets must be included in the overall channel I/O planning.
HiperSockets IP network devices are configured similarly to OSA-Express QDIO devices. Each HiperSockets requires the definition of a CHPID like any other I/O interface. The CHPID type for HiperSockets is IQD, and the CHPID number must be in the range of hex 00 to hex FF. No other I/O interface can use a CHPID number that is defined for a HiperSockets, even though HiperSockets does not occupy any physical I/O connection position.

Real LANs have a maximum frame size limit that is defined by their architecture. The maximum frame size for Ethernet is 1492 bytes. Gigabit Ethernet has a jumbo frame option for a maximum frame size of 9 KB. The maximum frame size for a HiperSocket is assigned when the HiperSockets CHPID is defined. Frame sizes of 16 KB, 24 KB, 40 KB, and 64 KB can be selected. The default maximum frame size is 16 KB. The selection depends on the characteristics of the data that is transported over a HiperSockets and is also a trade-off between performance and storage allocation.

The MTU size that is used by the IP network stack for the HiperSockets interface is also determined by the maximum frame size. Table 8-1 lists these values.

Table 8-1 Maximum frame size and MTU size

<table>
<thead>
<tr>
<th>Maximum frame size</th>
<th>Maximum transmission unit size</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 KB</td>
<td>8 KB</td>
</tr>
<tr>
<td>24 KB</td>
<td>16 KB</td>
</tr>
<tr>
<td>40 KB</td>
<td>32 KB</td>
</tr>
<tr>
<td>64 KB</td>
<td>56 KB</td>
</tr>
</tbody>
</table>

The maximum frame size is defined in the hardware configuration (IOCP) by using the CHPARM parameter of the CHPID statement.

z/OS allows the operation of multiple IP network stacks within a single image. The read control and write control I/O devices are required only once per image, and are controlled by VTAM. Each IP network stack within the same z/OS image requires one I/O device for data exchange.

Running one IP network stack per logical partition requires three I/O devices for z/OS (the same requirement as for z/VM and Linux on IBM Z). Each additional IP network stack in a z/OS Logical Partition requires only one more I/O device for data exchange. The I/O device addresses can be shared between z/OS systems that are running in different logical partitions. Therefore, the number of I/O devices is not a limitation for z/OS.

An IP address is registered with its HiperSockets interface by the IP network stack when the IP network device is started. IP addresses are removed from an IP address lookup table when a HiperSockets device is stopped. Under operating system control, IP addresses can be reassigned to other HiperSockets interfaces on the same HiperSockets LAN. This feature allows flexible backup of IP network stacks.

Reassignment is only possible within the same HiperSockets LAN. A HiperSockets is one network or subnetwork. Reassignment is only possible for the same operating system type. For example, an IP address that is originally assigned to a Linux IP network stack can be reassigned only to another Linux IP network stack. A z/OS dynamic VIPA can be reassigned only to another z/OS IP network stack, and a z/VM IP network VIPA can be reassigned only to another z/VM IP network stack. The LIC forces the reassignment. It is up to the operating system’s IP network stack to control this change.
Enabling HiperSockets requires the CHPID to be defined as type=IQD by using HCD and IOCP. This CHPID is treated like any other CHPID and is counted as one of the available channels within the IBM Z platform.

**HiperSockets definition change:** A new parameter for HiperSockets IOCP definition was introduced with the z13 and z13s. As such, the z14, z13, and z13s IOCP definitions for HiperSockets devices require the keyword VCHID. VCHID specifies the virtual channel identification number that is associated with the channel path (type IQD). The valid range is 7E0 - 7FF.

The HiperSockets LIC on Z platform supports the following features:

- Up to 32 independent HiperSockets.
- For z/OS, z/VM, Linux, and z/VSE the maximum number of IP network stacks or HiperSockets communication queues that can concurrently connect on a single IBM Z platform is 4096.
- Maximum total of 12288 I/O devices (valid subchannels) across all HiperSockets.
- Maximum total of 12288 IP addresses across all HiperSockets. These IP addresses include the HiperSockets interface and virtual IP addresses (VIPA) and dynamic VIPA that are defined for the IP network stack.

With the introduction of the channel subsystem, sharing of HiperSockets is possible with the extension to the multiple image facility (MIF). HiperSockets channels can be configured to multiple channel subsystems (CSSes). They are transparently shared by any or all of the configured logical partitions without regard for the CSS to which the partition is configured.
Figure 8-7 shows spanned HiperSockets that are defined on an IBM Z platform. For more information about spanning, see 2.1.8, “Channel spanning” on page 24.

8.3 Summary

HiperSockets is part of IBM z/Architecture technology and includes QDIO and advanced adapter interrupt handling. The data transfer is handled much like a cross-address space memory move, by using the memory bus. Therefore, HiperSockets does not contend with other I/O activity in the system.

HiperSockets can be defined to separate traffic between specific virtual servers and logical partitions on one IBM Z system. Virtual private networks (VPNs) or network virtual LANs across HiperSockets are supported to further isolate traffic as required. With integrated HiperSockets networking, there are no server-to-server traffic flows outside the IBM Z system. The only way to probe these VLANs is by using the NTA function, and strict controls are required for that procedure.

The Z platform support up to 32 HiperSockets. Spanned channel support allows sharing of HiperSockets across multiple CSSes and LPARs.
8.4 References

For more information about the HiperSockets function and configuration, see *IBM HiperSockets Implementation Guide*, SG24-6816.

For more information about the HiperSockets virtual bridge support for z/VM, see *z/VM Connectivity*, SC24-6174.

For more information about the HiperSockets integration with IEDN and z/OS Communications Server, see *z/OS Communications Server: IP Configuration Guide*, SC31-8775.
Coupling links and common time

This chapter describes the connectivity options that support IBM Parallel Sysplex clustering technology and common time on IBM Z platforms. It covers the following topics:

- IBM Z cluster technology
- Connectivity options
- Time functions
- References
9.1 IBM Z cluster technology

Clustering technology brings the power of parallel processing to business-critical applications. A Parallel Sysplex cluster consists of up to 32 IBM z/OS images, connected to one or more coupling facilities (CFs), using high-speed specialized links, called coupling links, for communication and time-keeping. The coupling facilities at the heart of the cluster enable high-speed, record-level read/write data sharing among the images in a cluster.

The primary purpose of coupling links is to support communications between z/OS and coupling facilities. The CF images provide critical locking/serialization, data caching and coherency, and messaging and queueing capabilities, which allow the systems in the sysplex to cooperate and share data.

When configured properly, a cluster has no single point of failure and can provide users with near-continuous application availability over planned and unplanned outages.

Figure 9-1 shows a Parallel Sysplex configuration with the Server Time Protocol (STP) feature. STP provides time synchronization for multiple servers and coupling facilities. The use of Network Time Protocol (NTP) servers as an external time source is supported by STP.

**Note:** A time synchronization mechanism, STP is a mandatory hardware requirement for a Parallel Sysplex environment that consists of more than one IBM Z platform.

STP is a server-wide facility that is implemented in the Licensed Internal Code (LIC) presenting a single view of time to IBM Processor Resource/Systems Manager™ (IBM PR/SM™). It is a message-based protocol in which timekeeping information is passed over coupling links between CPCs.
The CPCs are configured with coupling peer mode links, such as the Integrated Coupling Adapter (ICA SR), Coupling Express Long Reach (CE LR), InfiniBand coupling links, InterSystem Channel peer (ISC-3) link, and the Internal Coupling peer (IC) link. All of these links except the IC link are supported by STP.

With these components, coupling performance is balanced with the performance of z/OS images that are running on IBM Z platforms in a common time cluster.

A CF runs the Coupling Facility Control Code (CFCC) that is loaded into main storage at the time the CF image is activated. CFCC can run on a stand-alone CF CPC or in a logical partition.

9.1.1 Coupling links and STP

STP is a message-based protocol in which STP timekeeping information is passed over externally defined coupling links. Coupling links that can be used to transport STP messages are the Coupling Express LR, ICA SR coupling links, InfiniBand (IFB) coupling links, and InterSystem Channel-3 (ISC-3) links in peer mode.

**Reminder:** ISC-3 is not supported on the IBM z13 and later platforms. The zEC12 and zBC12 were the last platforms to support ISC-3 features.

Also, IBM z14 Model ZR1 does not support InfiniBand coupling links (PSIFB).

**Timing-only coupling links**

For a CPC that is not part of a Parallel Sysplex, but required to be time-synchronized, coupling links must be configured for the CPC to be part of a Coordinated Timing Network (CTN). Hardware configuration definition (HCD) has been enhanced to support **timing-only** links to be defined when a CF does not exist at either end of the coupling link. The timing-only links can be of type CL5 for CE LR, CS5 for ICA SR coupling links, CIB for InfiniBand coupling links, or CFP for ISC-3 in peer mode. The control unit type is STP, and no devices are defined to it.

**Note:** CF messages cannot be transferred over timing-only links.

**Coupling link redundancy for STP**

Between any two CPCs that are intended to exchange STP messages, each CPC must be configured with at least two coupling links for redundancy reasons. This configuration prevents the loss of one link, causing the loss of STP communication between the CPCs. As mentioned, if a CPC does not have a CF logical partition, timing-only links can be used to provide STP connectivity.

The maximum number of attached CPCs that are supported by any STP-configured CPC in a CTN is equal to the maximum number of coupling links that are supported by the server in the configuration.

For more information, see the following publications:

9.1.2 Multi-site sysplex considerations

If a sysplex is configured across two or more sites, you must plan to extend the Coupling Express LR, 1x IFB, or ISC-3 links beyond the distance that is supported without repeaters. IBM supports Wavelength Division Multiplexing (WDM) products that are qualified by IBM for use in multisite sysplex solutions, such as IBM Geographically Dispersed Parallel Sysplex (GDPS).

If any messages are to be transmitted across multiple sites through WDM products, ensure that only qualified WDM products and supported configurations are used.

For a list of WDM vendor products that are qualified for coupling links (transporting CF or STP messages) in a multi-site sysplex, see 10.4.2, “IBM Z qualified WDM vendor products” on page 179.

Also, see Resource Link for a list of qualified WDM products (requires a Resource Link ID).

9.2 Connectivity options

From a hardware perspective, when configuring a cluster, connectivity is a primary area of focus, as are other basic hardware components, such as the CF, Server Time Protocol, and coupling links. In an IBM Parallel Sysplex, the objective is any-to-any connectivity and nondisruptive planned or unplanned outages. All channels and directors must be included in the design. The primary reasons for this inclusion are to ensure cross-connectivity to all devices in the cluster and to potentially eliminate the effects of any outages.

For availability purposes, use at least two coupling links to connect each server to each CF in a Parallel Sysplex cluster. The performance, availability, and distance requirements of a Parallel Sysplex environment are the key factors that identify the appropriate connectivity option for any particular configuration.

9.2.1 Coupling link options

To ensure that sysplex environments continue to scale with today’s necessary efficiency as CPC performance grows, coupling link design is continuously enhanced.

Attention: Parallel Sysplex technology supports connectivity between systems that differ by up to two generations (n-2). For example, an IBM z14 can participate in an IBM Parallel Sysplex cluster with z13, z13s, zEC12, and zBC12 platforms.

However, the IBM z14 ZR1 does not support InfiniBand connectivity, as such, an IBM z14 ZR1 does support connectivity to only one generation back (n-1): z14 ZR1 can connect to z14, z13, and z13s, using Integrated Coupling Adapter Short Reach (ICA SR) and Coupling Express Long Reach (CE LR) features.

Supported on the IBM z13 and newer Z platforms, the ICA SR, CS5 coupling links are available. CS5 coupling links support connectivity between any z14, z14 ZR1, z13, or z13s to another z14, z14 ZR1, z13, or z13s

Supported on the IBM z13\(^1\) and newer Z platforms, the Coupling Express LR, CL5 coupling links are available. CL5 coupling links support connectivity between any z14, z14 ZR1, z13, or z13s to another z14, z14 ZR1, z13, or z13s
On the z14, z13, z13s, and zEnterprise CPCs (zEC12, zBC12) platforms IFB\(^2\) coupling links are available. IFB coupling links support peer mode connectivity between z14, z13, zEC12, zBC12, to any z14, z13, zEC12, or zBC12 (12x IFB and 1x IFB)

As shown in Figure 9-2, the following supported connections are included:

- From IBM Z general-purpose platforms to IBM Z CF, whereby the CF can be hosted in a CF-only system or run in logical partition mode with other operating systems partitions on an IBM Z system.
- Within an IBM Z platform, an IC link uses memory-to-memory data transfers between logical partitions (LPARs).

Table 9-1 lists the coupling link support for each IBM Z platform. Restrictions on the maximum numbers can apply, depending on the configuration. Always check with your IBM support team.

Table 9-1  Coupling link support

<table>
<thead>
<tr>
<th>Feature</th>
<th>Maximum supported links(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>z14</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>32</td>
</tr>
<tr>
<td>IC</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Check the latest z13 HMC driver level information for feature support.

\(^2\) z14 is the last IBM Z server model to support InfiniBand coupling links.
For more information about distance support for coupling links, see Chapter 10, “Extended distance solutions” on page 169.

9.2.2 Internal Coupling link

IC links are Licensed Internal Code-defined links to connect a CF to a z/OS logical partition in the same CPC. These links are available on all IBM Z platforms. The IC link is an IBM Z coupling option that enables high-speed, efficient communication between a CF partition and one or more z/OS logical partitions that are running on the same CPC. The IC is a linkless connection (implemented in LIC) and does not require any hardware or cabling.

An IC link is a fast coupling link that uses memory-to-memory data transfers. IC links do not have PCHID numbers, but do require CHPIDs.

IC links have the following attributes:

- They provide the fastest connectivity that is significantly faster than any external link alternatives.
- They result in better coupling efficiency than with external links, effectively reducing the CPC cost that is associated with Parallel Sysplex technology.
- They can be used in test or production configurations, reduce the cost of moving into Parallel Sysplex technology, and enhance performance and reliability.
They can be defined as spanned channels across multiple CSSes.
They are available at no extra cost (no feature code). Employing ICFs with IC links results in considerable cost savings when configuring a cluster.

IC links are enabled by defining CHPID type ICP. A maximum of 32 IC links can be defined on an IBM Z platform.

9.2.3 InterSystem Channel-3 (ISC-3)

ISC-3 provides the connectivity that is required for data sharing between the CF and the systems. These links also provide connectivity for STP messaging between any zEnterprise (zEC12 and zBC12) system to any zEnterprise system.

**Note:** zEnterprise 196 and zEnterprise 114 were the last systems to offer ordering of ISC-3 connections. ISC-3 is not supported on IBM z13 or later CPCs.

ISC-3 links support point-to-point connections (directly connecting CFs and systems) and require a unique channel definition at each end of the link.

The ISC-3 feature consists of a mother card (ISC-M feature code 0217) and one or two ISC-3 daughter cards (ISC-D feature code 0218). Each ISC-3 daughter card has two independent ports with one CHPID associated with each active port. A port on an ISC-3 daughter card is activated by using Licensed Internal Code - Control Code (LIC-CC) with feature code 0219.

Each ISC-3 port operates in peer mode. The ISC-3 port mode is defined by the CHPID type (CFP) by using the HCD and the IOCP. ISC-3 feature ports in peer mode provide connection from general-purpose models to coupling facilities (Figure 9-2 on page 157).

An ISC-3 port is defined in peer mode when its associated CHPID is defined as CHPID type CFP. Each feature port supports connection to another CPCs ISC-3 port through a 9-micron single mode fiber optic cable that is terminated with an LC connector. An ISC-3 port in peer mode operates at 2 Gbps and supports a maximum unrepeated distance of 10 km. RPQ 8P2197 (InterSystem Channel 3, or ISC-3, Long Distance option) provides an ISC-3 daughter card with two active ports that clock at 1 Gbps and support a maximum unrepeated distance of 20 km.

Under certain conditions, RPQ 8P2263 might be required along with RPQ 8P2197. Check with your IBM representative.

For more information about unrepeated distance support, see Table 10-1 on page 170. ISC-3 in peer mode supports STP.

9.2.4 Integrated Coupling Adapter Short Range

The ICA SR was introduced with the IBM z13. It allows z14, z14 ZR1, z13, or z13s connectivity to another z14, z14 ZR1, z13, or z13s. The ICA SR (FC 0172) is a two-port, short-distance coupling feature that uses coupling channel type: CS5. The ICA SR uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

The ICA SR is designed to drive distances up to 150 m and supports a link data rate of 8 GBps. It is designed to support up to four CHPIDs per port and seven subchannels (devices) per CHPID. The maximum number of ICA SR features is 40 for the z14, 20 for the z13, and eight for the z14 ZR1 and z13s.
The ICA SR feature configuration for each Z platform is as follows:

- Up to 10 ICA SR features and up to 20 ICA SR ports are supported in a z14 CPC drawer. A maximum of 80 ports are supported in the z14.
- Up to eight ICA SR features (16 ports) are supported in a z14 ZR1 (Feature dependent, see Table 9-1 on page 157).
- Up to 10 ICA SR features and up to 20 ICA SR ports are supported in a z13 CPC drawer. A maximum of 40 ports are supported in a z13.
- Up to 8 ICA SR features and a maximum of 16 ICA SR ports are supported in a z13s.

The ICA SR cannot be connected to an IFB coupling feature such as HCA3-O or HCA3-O LR InfiniBand coupling features. For distances up to 100 m, OM3 fiber can be used, and for distances up to 150 m OM4 fiber must be used.

For more information, see IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, and Open System Adapters), GA23-1407. This publication is available in the Library section of Resource Link.

### 9.2.5 Coupling Express Long Reach

The Coupling Express LR occupies one slot in a PCIe I/O drawer\(^3\) or PCIe+ I/O drawer\(^4\). It allows z14, z14 ZR1, z13, or z13s connectivity only to another z14, z14 ZR1, z13, or z13s. The Coupling Express LR (FC 0433) is a two-port, long-distance coupling feature that uses a new coupling channel type: CL5.

The Coupling Express LR is a PCIe feature. It is designed to drive distances up to 10 km (unrepeated) or up to 100 km with a qualified DWDM, and supports a link data rate of 10 Gbps and connects using the same 9 µm, Single Mode fiber type as 1X IFB and ISC-3.

Coupling Express LR is used in a Point-to-Point topology (just like 1X IFB and ISC-3) and it cannot be used in a switched environment.

Coupling Express LR is designed to support up to four CHPIDs per port, 32 buffers (that is, 8 or 32 subchannels) per CHPID. The Coupling Express LR feature resides in the PCIe I/O drawer on IBM z14, z13, and z13s and in a PCIe+ I/O drawer in the z14 ZR1. Up to 32 Coupling Express LR features and up to 64 ports are supported on a z14 and z13. Up to 16 Coupling Express LR features and up to 32 ports are supported on a z14 ZR1 and z13s.

The Coupling Express LR cannot be connected to a 1 x IFB coupling feature such as HCA3-O LR InfiniBand coupling feature.

For more information, see IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express), GA23-1408. This publication is available in the Library section of Resource Link.

---

\(^3\) Check the latest z13 HMC driver level information for feature support.

\(^4\) PCIe+ I/O drawer (FC 4001) has been introduced with z14 ZR1, which is built in a 19“ rack. FC 4001 is the 19“ form factor PCIe Gen 3 - based I/O drawer. and can host up to 16 PCIe I/O features (adapters). It is not supported on z14 or older servers. Also, the PCIe I/O drawer cannot be carried forward during and MES upgrade to z14 ZR1. z14 ZR1 supports ONLY PCIe+ I/O drawers, up to four.
9.2.6 InfiniBand coupling links

Note: The IBM z14 (M/T 3906) is be the last IBM Z server to support any kind of InfiniBand coupling connectivity.

For more information about long distance coupling, see Hardware Announcement z Systems Long Distance Coupling, 117-031.

IFB coupling links are high-speed links on z14, z13, z13s, zEC12, and zBC12. The IFB coupling links originate from four types of fanouts:

- HCA3-O (FC 0171) for 12x IFB links on z14, z13, z13s, and zEnterprise CPCs
- HCA3-O LR (FC 0170) for 1x IFB links on z14, z13, z13s, and zEnterprise CPCs
- HCA2-O (FC 0163) for 12x IFB links on zEnterprise CPCs
- HCA2-O LR (FC 0168) for 1x IFB links on zEnterprise CPCs

Each fanout that is used for coupling links has an assigned adapter ID number (AID) that must be used for definitions in IOCDS to have a relation between the physical fanout location and the CHPID number.

12x IFB coupling links
The HCA2-O and HCA3-O fanout support IFB coupling links that operate at 6 GBps (12x IFB). IFB coupling links use a fiber optic cable that is connected to a HCA2-O or HCA3-O fanout. The maximum distance for an IFB link is 150 meters. The fiber cables are industry standard OM3 50/125 micrometer multimode optical cables with Multifiber Push-On (MPO) connectors. 12x IFB supports seven or 32 subchannels5 per CHPID.

12x IFB and 12x IFB3 protocols
Two protocols are supported by the HCA3-O for 12x IFB feature:

- 12x IFB3 protocol
  When HCA3-O fanouts are communicating with HCA3-O fanouts and have been defined with four or fewer CHPIDs per port, the 12x IFB3 protocol is used.

- 12x IFB protocol
  If more than four CHPIDs are defined per HCA3-O port or HCA3-O features are communicating with HCA2-O features on zEnterprise servers, links run with the 12x IFB protocol.

The HCA3-O feature that supports 12x InfiniBand coupling links improves service times. When no more than four CHPIDs are defined per HCA3-O port, the 12x IFB3 protocol is used. When using the 12x IFB3 protocol, synchronous service times for lock structures are up to 40% faster than when using the 12x IFB protocol.

1x IFB coupling links
The HCA2-O LR and HCA3-O LR fanout support 1x IFB coupling links that operate at up to 5.0 Gbps. 1x IFB coupling links use a fiber optic cable that is connected to a HCA2-O LR or HCA3-O LR fanout. The maximum unrepeated distance for a 1x IFB link is 10 km. When using repeaters, the maximum distance is up to 100 km. The fiber cables that are used for 1x IFB links are standard 9 µm single mode fiber optic cables with an LC duplex connector. 1x IFB supports seven subchannels per CHPID.

5 Depending on the version of the HCD, the default setting for the subchannels is set to 7 or 32.
Fanout adapter ID
Unlike channels that are installed in an I/O cage, I/O drawer, or PCIe I/O drawer, which are identified by a PCHID number that is related to their physical location, coupling link fanouts and ports are identified by an AID. The adapter ID value depends on its physical location. The AID must be used to assign a CHPID to the fanout in the IOCDS definition. The CHPID assignment is done by associating the CHPID to an AID and port. For more information, see 2.1.6, “Adapter ID” on page 21.

Note: The IBM zEnterprise EC12 and the IBM zEnterprise BC12 were the last IBM Z platforms to support the following features as carry forward on an upgrade:
- HCA2-O fanout for 12x IFB coupling links (FC 0163)
- HCA2-O LR fanout for 1x IFB coupling links (FC 0168)

9.3 Time functions
There is a long-standing requirement for accurate time and date information in data processing. A single operating system has been replaced by multiple, coupled operating systems on multiple CPCs. This need evolved to a requirement for both accurate and consistent clocks among these systems. Clocks are said to be consistent when the difference or offset between them is sufficiently small. An accurate clock is consistent with a standard time source.

The IBM z/Architecture, STP, facilitates the synchronization of CPC time-of-day (TOD) clocks to ensure consistent time stamp data across multiple CPCs and operating systems. STP provides a way to synchronize TOD clocks in different CPCs with a centralized time reference. This setup, in turn, can be set accurately based on an international time standard (external time source). The architecture defines a time-signal protocol and a distribution network, which allows accurate setting, maintenance, and consistency of TOD clocks.

STP facility
STP is a server-wide facility that is implemented in the Licensed Internal Code of IBM Z platforms and coupling facilities. STP presents a single view of time to PR/SM and provides the capability for multiple CPCs and coupling facilities to maintain time synchronization with each other. Any IBM Z platform can be enabled for STP by installing the STP feature. Each CPC that is planned to be configured in a CTN must be STP-enabled.

IBM z14 ZR1 STP Connectivity: If a z14 ZR1 plays a CTN role (PTS/BTS/Arbiter), then the other CTN role playing CPCs (BTS/PTS/Arbiter) must have coupling connectivity to the z14 ZR1 (N, N-1 connectivity only).

NTP client support is available to the STP code on the Z platform. With this function, the CPCs can be configured to use an NTP server as an external time source. NTP server with pulse per second (PPS) is supported.

Note: See the white paper Important Considerations for STP Server Role Assignments in the IBM Techdocs Library if you configured an STP CTN with three or more servers.
**STP recovery enhancements**

The following are the recovery enhancements to STP:

- New graphical interface for managing STP environment.
- Enhanced Console Assisted Recovery (z14, z14 ZR1, z13\(^6\), and z13s)

  Enhanced Console Assisted Recovery (ECAR) speeds up the process of BTS takeover:
  
  - When the Primary Time Server (PTS/CTS) encounters a checkstop condition, the CEC informs its Support Element (SE) and Hardware Management Console (HMC).
  - The PTS SE recognizes the checkstop pending condition, and the PTS SE STP code is called.
  - The PTS SE sends an ECAR request by using HMC to the BTS SE.
  - The BTS SE communicates with the BTS to start the takeover.

- Starting with IBM z14, an additional STP stratum level is supported (Stratum 4). This additional stratum level has been implemented to alleviate the additional complexity and expense of system reconfiguration using system upgrades. It should only be used as a temporary state during reconfiguration. Environments should not run with systems at Stratum level 4 for extended periods of time because of the lower quality of the time synchronization.

- “Going away” signal

  The HCA3 generation and ICA SR of host channel adapters (HCA3-O, HCA3-O LR, or ICA SR) send a reliable, unambiguous “going away” signal to indicate that the server on which the HCA3 is running is about to enter a Failed state (check stopped). When the going-away signal sent by the Current Time Server (CTS) in an STP-only CTN is received by the Backup Time Server (BTS), the BTS can safely take over as the CTS. It can do so without relying on the previous recovery methods of Offline Signal (OLS) in a two-server CTN or the Arbiter in a CTN with three or more servers.

  This enhanced recovery method is available on the Z platform. It is available only if you have an ICA SR, HCA3-O, or HCA3-O LR on the CTS, respectively, communicating with an ICA SR, HCA3-O, or HCA3-O LR on the BTS. The available STP recovery design is still available for the cases when a “going away” signal is not received or for failures other than a server failure.

---

### 9.3.1 Dynamic Split and Merge for Coordinated Timing Network

With HMC 2.14.1, is possible to dynamically split a CTN or merge two CTNs by using the STP GUI that is available on the HMC. STP understands the images that are present on each CEC in the source STP networks and the sysplex affiliation (more or less) of each of those z/OS images. STP can use this information for integrated checking, which operates with z/OS toleration for CTN change.

**CTN split**

When splitting a CTN, STP checks to ensure that it does result in a sysplex that “splits” across the two resultant CTNs. New STP roles are automatically assigned in the split-off CTN during the process. STP connectivity and max-stratum checking is performed in both split-off CTNs. Checking results are previewed and confirmed before proceeding.

---

\(^6\) Check the latest z13 HMC driver level information for feature support.
CTN Merge

When merging two CTNs, STP roles are automatically maintained in the merged-into CTN while STP connectivity and max-stratum checking is performed in the merged CTN. Preview and confirm the results before proceeding.

9.3.2 Dynamic I/O reconfiguration for standalone CF CPCs

With z14 Driver Level 36, support for dynamic activation of a new or changed IODF on a standalone CF is supported without requiring a POR/IML of the standalone CF CPC and without requiring the presence of any z/OS or z/VM image running an HCD instance on the same CPC. This solution does not require the use of Dynamic Partition Manager (DPM) mode.

A new LPAR, called the MCS LPAR, is a firmware based appliance version of the HCD instance that is deployed on the standalone CF CPC. This LPAR is driven by an updated HCD/HCM that is running in a z/OS LPAR on a remote z14 system (driver level 36) and managed by a HMC V2.14.1 connected to the standalone CF CPC.

A Power-on Reset with an IOCDS that includes and establishes the MCS LPAR on the standalone CF CPC is needed before this new capability can be used. After this “last” POR is done on the standalone CF CPC, all subsequent dynamic I/O changes can be done without subsequent POR/IMLs.

**HCD requirements:** Dynamic I/O for Standalone CF CPC requires that both systems (driver and standalone CF CPC) are at driver level 36 or later, HCD fixes for the driver system, and HMC 2.14.1.

The dynamic I/O for standalone CF CPC is shown in Figure 9-3.
9.3.3 NTP server with PPS support

STP can use an NTP server that has a PPS output signal as its external time source. This type of external time source device is available worldwide from several vendors that provide network timing solutions.

The NTP output of the NTP server must be connected to the SE LAN because the NTP client is running on the SE. Also, the PPS output of the same NTP server must be connected to the PPS input port on the External Clock Facility (ECF) card on zEC12 CPCs, or a port on the oscillator (OSC) card on Z platform. There is no automatic check for the correctness of this connection.

However, a manual approximation of this check can be achieved by having the user check the SE to confirm that the PPS is active, then unplugging the PPS cable from the NTP Server. The user could then check by using the SE to see that the PPS port is no longer receiving a signal. Then, the user could replug the cable to the NTP server and make sure that the SE indicates the PPS signal is now being received at the CPC. This process allows STP to confirm that the NTP server that is used for the source of time is the same one that the CPC is receiving the PPS signal from.

STP tracks to the PPS signal to maintain time accuracy for the CTN. STP maintains an accuracy of 10 microseconds to the PPS input signal. Several variables, such as the cable used to connect the PPS signal and the accuracy of the NTP server to its time source (GPS or radio signals), determine the ultimate accuracy of STP to Coordinated Universal Time.

zEC12 PPS support

Two ECF cards is a standard zEC12 systems feature. The ECF cards provide a dual-path interface for PPS support. A cable connection from the PPS port on the ECF card to the PPS output of the NTP server is required when the server is using STP and is configured in an STP-only CTN by using NTP with PPS as the external time source. Each of the two standard ECF cards has one PPS port for a coaxial cable connection to a PPS port on an NTP server.

z14, z14 ZR1, z13, z13s, and zBC12 PPS support

Each of the two oscillator cards in the IBM z14, z14 ZR1, z13, and z13s has a PPS port to provide connectivity to an NTP server for PPS support.

The two oscillator cards in the first processor drawer of the zBC12 provide a dual-path interface for PPS support. A cable connection from the PPS port on the oscillator card to the PPS output of the NTP server is required when the server is using STP and is configured in an STP-only CTN by using NTP with PPS as the external time source.

Each of the two standard oscillator cards has one PPS port for a coaxial cable connection to a PPS port on an NTP server.

Figure 9-4 on page 166 shows an expanded availability configuration with two ECF features in a zEC12. Each ECF card is connected to a different NTP server with PPS support:

- On z14 and z13, it is connected to the oscillator cards in the back of frame A.
- On z14 ZR1, the NTP server with PPS is connected to the oscillator cards (OSC) located in the CPC drawer (rear side of the CPC drawer).
- On z13s systems, the NTP server with PPS support is connected to the oscillator cards in the oscillator backplane located on the back of the frame.
- On zBC12 systems, it is connected to the oscillator cards in the first processor drawer.
For more information about NTP support, see the Parallel Sysplex advantages web page.

### 9.3.4 Operating system support

Software requirements vary with the design and use of common time. All current IBM z/OS versions support Server Time Protocol (STP).

For more information about STP operating system support, see the STP tab on the Parallel Sysplex web page.

Consult the appropriate Preventive Service Planning (PSP) buckets (3906DEVICE, 3907DEVICE, 2964DEVICE, 2965DEVICE, 2828DEVICE, and 2827DEVICE) before implementation.

### 9.4 References

See the Parallel Sysplex web page for help understanding, planning, and implementing a Parallel Sysplex cluster.

The following publications provide detailed information about Parallel Sysplex:

- *Planning for Fiber Optic Links*, GA23-0367
Getting the Most Out of a Parallel Sysplex, SG24-2073
IBM zEnterprise EC12 Technical Guide, SG24-8049
IBM zEnterprise BC12 Technical Guide, SG24-8138
IBM z13 Technical Guide, SG24-8251
IBM z13s Technical Guide, SG24-8294
IBM z14 Technical Guide, SG24-8451
IBM z14 ZR1 Technical Guide, SG24-8651
Extended distance solutions

This chapter describes architectural requirements and implementation for IBM Z platform connectivity over extended distances.

The following topics are covered in this chapter:

- Unrepeated distances
- Fibre Connection
- Coupling links
- Wavelength-division multiplexing
- References
10.1 Unrepeated distances

This section lists the maximum unrepeated distance and link budget for each type of IBM Z fiber optic link. Longer distances are possible by using repeaters, switches, channel extenders, and wavelength division multiplexing (WDM).

In Table 10-1, a link is a physical connection over a transmission medium (fiber) that is used between an optical transmitter and an optical receiver. The maximum allowable link loss, or link budget, is the maximum amount of link attenuation (loss of light), expressed in decibels (dB), that can occur without causing a possible failure condition (bit errors). When you use multimode fiber, as the link data rate increases, the unrepeated distance and link budget decreases.

The link budget is derived from combining the channel insertion loss budget with the unallocated link margin budget. The link budget numbers are rounded to the nearest tenth of a dB.

Table 10-1  Fiber optic connections: unrepeated distances

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Fiber type</th>
<th>Link data rate</th>
<th>Fiber bandwidth MHz-km</th>
<th>Maximum distance</th>
<th>Link budget dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICON LX</td>
<td>SM 9 µm</td>
<td>1 Gbps</td>
<td></td>
<td>10 km</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Gbps</td>
<td></td>
<td>4 km</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td></td>
<td>10 km</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td></td>
<td>4 km</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td></td>
<td>10 km</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td></td>
<td>4 km</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 Gbps</td>
<td></td>
<td>10 km</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16 Gbps</td>
<td></td>
<td>10 km</td>
<td>6.4</td>
</tr>
</tbody>
</table>
### Feature type

<table>
<thead>
<tr>
<th>Fiber type</th>
<th>Link data rate</th>
<th>Fiber bandwidth MHz-km</th>
<th>Maximum distance a</th>
<th>Link budget dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM 62.5 µm</td>
<td>1 Gbps</td>
<td>200</td>
<td>300 m</td>
<td>3.00</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>200</td>
<td>500</td>
<td>600 m</td>
<td>3.85</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>2 Gbps</td>
<td>200</td>
<td>150 m</td>
<td>2.10</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>300</td>
<td>300 m</td>
<td>2.62</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>4 Gbps</td>
<td>200</td>
<td>70 m</td>
<td>1.78</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>150</td>
<td>150 m</td>
<td>2.06</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>8 Gbps</td>
<td>200</td>
<td>21 m</td>
<td>1.58</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>50</td>
<td>50 m</td>
<td>1.68</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>16 Gbps b</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>35</td>
<td>35 m</td>
<td>1.63</td>
</tr>
<tr>
<td>SM 9 µm</td>
<td>1 Gbps</td>
<td>N/A</td>
<td>5 km</td>
<td>4.6</td>
</tr>
<tr>
<td>MM 62.5 µm c</td>
<td>500</td>
<td>500</td>
<td>500 m</td>
<td>2.4</td>
</tr>
<tr>
<td>MM 50 µm c</td>
<td>500</td>
<td>550</td>
<td>550 m</td>
<td>2.4</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>1 Gbps</td>
<td>200</td>
<td>275 m</td>
<td>2.6</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>550</td>
<td>550 m</td>
<td>3.6</td>
</tr>
<tr>
<td>SM 9 µm</td>
<td>10 Gbps</td>
<td>N/A</td>
<td>10 km</td>
<td>6</td>
</tr>
<tr>
<td>MM 62.5 µm</td>
<td>10 Gbps</td>
<td>200</td>
<td>33 m</td>
<td>2.5</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>500</td>
<td>82</td>
<td>82 m</td>
<td>2.3</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>2000</td>
<td>300</td>
<td>300 m</td>
<td>2.6</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>25 Gbps</td>
<td>2000</td>
<td>70 m</td>
<td>2.2</td>
</tr>
<tr>
<td>MM 50 µm</td>
<td>4700</td>
<td>100</td>
<td>100 m</td>
<td>2.5</td>
</tr>
</tbody>
</table>

a. Some types of features might have extended distance when using RPQs.
b. FICON Express16S features support 16 Gbps on z14, z14 ZR1, z13, and z13s only. A FICON Express16S+ feature has been released for z14 only.
c. Requires fiber optic mode conditioning patch (MCP) cables.

The following notes apply to Table 10-1 on page 170:

- Single-Byte Command Code Sets Connection (SBCON) is the American National Standards Institute (ANSI) standard for the command set that is used by FICON over a Fibre Channel physical interface. It is also known as FC-SB.
- All industry-standard links (FICON, Gigabit Ethernet) follow published industry standards. The minimum fiber bandwidth requirement to achieve the distances that are listed is applicable for multimode (MM) fiber only. There is no minimum bandwidth requirement for single mode (SM) fiber.
- The bit rates that are given might not correspond to the effective channel data rate in a particular application because of protocol overhead and other factors.
- LC duplex and SC duplex connectors are keyed per the ANSI Fibre Channel Standard specifications.
- Mode-conditioning patch (MCP) cable is required to operate certain links over multimode fiber.
- The ISC-3 feature supports an LC duplex connector. A conversion kit (SM SC duplex receptacle to LC duplex connector) can be used to enable existing single mode SC duplex cable infrastructure.
- The FICON Express2 and FICON Express features allow an auto-negotiated link speed of either 1 Gbps or 2 Gbps. The specifications in the table are for 2 Gbps link rate. MCP cables are not supported for 2 Gbps links.
- The FICON Express4 features allow an auto-negotiated link speed of either 1 Gbps, 2 Gbps, or 4 Gbps. MCP cables are not supported for 2 Gbps or 4 Gbps links.
- The FICON Express8 and FICON Express8S features allow an auto-negotiated link speed of either 2 Gbps, 4 Gbps, or 8 Gbps. MCP cables are not supported by FICON Express8 and FICON Express8S features.
- The FICON Express16S features, available on z14 and z14 ZR1 (carry forward only), and also on z13 and z13s, allow an autonegotiated link speed of either 4 Gbps, 8 Gbps, or 16 Gbps. MCP cables are not supported by FICON Express16S features.
- The FICON Express16S+ features, available on z14 only, allow an autonegotiated link speed of either 4 Gbps, 8 Gbps, or 16 Gbps. MCP cables are not supported by FICON Express16S+ features.
- As light signals traverse a fiber optic cable, the signal loses some of its strength. Decibels (dB) is the metric used to measure light power loss. The significant factors that contribute to light power loss are the length of the fiber, the number of splices, and the number of connections. The amount of light power loss (dB) across a link is known as the link budget. All links are rated for a maximum link budget (the sum of the applicable light power loss factors must be less than the link budget) and a maximum distance (exceeding the maximum distance causes undetectable data integrity exposures). Another factor that limits distance is jitter, but that is typically not a problem at these distances.
- Measure link budget and fiber bandwidth at the appropriate wavelength:
  - Long wavelength (1300 nm)
  - Short wavelength (850 nm)
For planning purposes, the following worst case values can be used to estimate the link budget. See the references listed and contact the fiber vendor for specific values, which might be different for your configuration:
  - Link loss at 1300 nm = 0.50 db/km
  - Link loss per splice = 0.15 db/splice (not dependent on wavelength)
  - Link loss per connection = 0.50 db/connection (not dependent on wavelength)
- Deviations from these specifications (longer distance or link budget) might be possible. These deviations are evaluated on an individual basis by submitting a request for price quotation (RPQ) to IBM.
10.2 Fibre Connection

This section describes architectural requirements and implementation solutions for Fibre Connection (FICON) channel connectivity over unrepeated and repeated distances. The term FICON represents the architecture as defined by the InterNational Committee of Information Technology Standards (INCITS) and published as ANSI standards. FICON also represents the names of the IBM Z platform feature types:

- FICON Express16S+
- FICON Express16S
- FICON Express8S
- FICON Express8
- FICON Express4
- FICON Express2
- FICON Express

All feature types support a long wavelength (LX) laser version and a short wavelength (SX) laser version. They support native FICON (FC, FCTC) and Fibre Channel Protocol (FCP) channel modes.

For more information, see 4.3, “Connectivity” on page 62.

10.2.1 FICON unrepeated distance

The unrepeated distance that is supported by IBM Z FICON features depends on these factors:

- The feature port transceiver type (LX or SX)
- The fiber type being used:
  - 9 µm single mode
  - 50 µm or 62.5 µm multimode
  Also, for multimode, the fiber bandwidth (MHz-km) of the fiber.
- The speed at which the feature port is operating
- Whether there are MCP cables in the fiber optic link

For more information, see Table 10-1 on page 170, and Planning for Fiber Optic Links, GA23-0367.

10.2.2 FICON repeated distance solutions

This section describes several extended distance connectivity solutions for FICON channel-attached I/O control units and devices.

The repeated distance for a FICON channel is IBM Z qualified to a maximum of 100 km. For all FICON features that use repeaters, the end-to-end distance between the FICON channel and the FICON Director port can be up to 100 km (62 miles).
The same end-to-end distance is also available between the FICON Director port and the control unit port. However, the overall end-to-end distance between the FICON channel and control unit is IBM Z qualified for up to 100 km (62 miles) only.

**FICON data rate droop**
For FICON channels, the channel-to-control unit end-to-end distance can be increased up to 100 km without data rate performance droop occurring if the FICON Director buffer credits are set appropriately. The number of buffer credits that are required depends on the link data rate and the maximum number of buffer credits that are supported by the FICON Director or control unit, and application and workload characteristics.

Although it is theoretically possible for FICON to maintain high bandwidth at distances greater than 100 km, these distances have not been qualified for use with IBM Z. They are achievable only if enough buffer credits exist to support the link speed.

The distance capability with FICON is becoming increasingly important as movement occurs towards remote I/O vaulting for disaster recovery, business continuity, and IBM Geographically Dispersed Parallel Sysplex (GDPS) for availability.

**FICON LX channel path with one FICON Director**
One FICON Director increases the maximum supported distance of a FICON LX channel path by using single-mode fiber optic cables (see Figure 10-1). For example, the maximum supported distance is 20 km (12.4 miles) with FICON Express8S LX (long wavelength) features for 2 Gbps\(^1\), 4 Gbps, and 8 Gbps LX links. The same is valid for the FICON Express16S features at 16 Gbps, available for z13 and the FICON Express16S+ features at 16 Gbps, and available for z14 only.

![FICON LX path with one Director](image)

**FICON LX channel path with two cascaded FICON Directors**
A FICON channel path can include a maximum of two cascaded FICON Directors. The maximum supported distance of the FICON channel path is FICON Director vendor specific. The use of extended distance long wavelength transceivers on the inter-switch links (ISL) between the FICON Directors might be required to achieve vendor-quoted distances. Each ISL requires one fiber trunk (two fibers) between the FICON Directors.

---

\(^1\) FICON Express16S+ features do not support 2 Gbps links.
For example, by using the configuration in Figure 10-2 and assuming that the distance between the two FICON Directors is 10 km (6.21 miles), the maximum supported distance is 30 km (18.64 miles) with FICON Express8S 10KM LX features for 2 Gbps, 4 Gbps, and 8 Gbps LX links. The example is also valid for the FICON Express16S features at 16 Gbps, available for z14 (carry forward only) and z13 servers and the FICON Express16S+ features at 16 Gbps, available for z14.

### FICON channel path: Transceiver intermix

A FICON channel path through one or two FICON Directors consists of multiple optical fiber links. Each link in the channel path can be either LX or SX, allowing the channel path to be made up of a mixture of link types. This configuration is possible because the FICON Director converts optical to electrical and back to optical (known as an OEO conversion, for optical-electrical-optical) of the channel path as it passes through the director.

**Note:** The transceiver type (LX or SX) at each end of a particular link must match.

### WDM technologies

Other extended distance connectivity technologies are available to extend FICON and other link types, for example, WDM. WDM technology also provides increased flexibility in that multiple links and protocol types can be transported over a single dark fiber trunk. For more information, see 10.4, “Wavelength-division multiplexing” on page 177.
10.3 Coupling links

This section describes architectural requirements and implementation solutions for coupling link connectivity over unrepeated and repeated distances.

Coupling link unrepeated distance

Table 10-2 lists the maximum unrepeated distances and link data rates that are supported for coupling links on IBM Z platforms. For more information, see Table 10-1 on page 170, and the accompanying notes in IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express), GA23-1408.

<table>
<thead>
<tr>
<th>Coupling link type</th>
<th>Maximum unrepeated distance</th>
<th>Link data rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC (internal)</td>
<td>N/A</td>
<td>Memory-to-memory (the highest bandwidth)</td>
</tr>
<tr>
<td>ISC-3 peer mode at 2 Gbps</td>
<td>10 km</td>
<td>2 Gbps</td>
</tr>
<tr>
<td></td>
<td>12 km&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>ISC-3 peer mode at 1 Gbps&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10 km</td>
<td>1 Gbps</td>
</tr>
<tr>
<td></td>
<td>20 km&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>ICA SR</td>
<td>150 meters&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8 Gbps</td>
</tr>
<tr>
<td>Coupling Express LR</td>
<td>10 km</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>12x IFB</td>
<td>150 meters</td>
<td>6 Gbps</td>
</tr>
<tr>
<td>1x IFB</td>
<td>10 km</td>
<td>2.5 Gbps or 5 Gbps</td>
</tr>
</tbody>
</table>

<sup>a</sup> Requires RPQ IBM Z Extended Distance; select RPQ based on the Z generation: zEC12: 8P2581; z13s and zBC12: 8P2781; z14 and z13: 8P2981.<sup>.</sup>

<sup>b</sup> Requires RPQ 8P2197. This RPQ provides an ISC-3 daughter card, which clocks at 1 Gbps in peer mode. This configuration allows the ISC-3 peer mode link to have an unrepeated distance extended up to 20 km. Under certain conditions, RPQ 8P2263 (or RPQ 8P2340 for the z10 BC) might be required with RPQ 8P2197. Check with your IBM representative.

<sup>c</sup> 150 meters distance is achieved by using OM4 fiber types only; with OM3 fiber, the distance is 100 meters maximum.
10.4 Wavelength-division multiplexing

WDM is a technique that is used to transmit several independent bit streams over a single fiber optic link (see Figure 10-3). It is an approach to opening up the conventional optical fiber bandwidth by breaking it up into many channels, each at a different optical wavelength (a different color of light). Each wavelength can carry a signal at any bit rate less than an upper limit defined by the electronics, typically up to several gigabits per second.

![WDM transmission technique](Figure 10-3)

The channels are protocol-independent, so a wide variety of protocols is supported, including FICON, FCP, coupling links, Server Time Protocol (STP), and Gigabit Ethernet.

The actual signal bandwidth that the electronics can handle over one wavelength is such a small fraction of the inter-channel spacing. Because of this feature, the signals do not interfere with one another and can therefore be multiplexed into a single fiber by using a passive grating multiplexer.

There are several extended distance uses of WDM technology:

- FICON/FCP channel connections to remote control units and devices
- LAN and network protocol connections to remote sites
- IBM System Storage® Metro Mirror (synchronous Peer-to-Peer Remote Copy (PPRC))
- IBM System Storage Global Mirror
- IBM System Storage z/OS Global Mirror (asynchronous Extended Remote Copy (XRC))
- Peer-to-Peer Virtual Tape Server (PnP VTS), a form of remote copying tape data

10.4.1 GDPS qualification

GDPS is an enterprise-wide continuous availability and disaster recovery automation solution that can manage recovery from planned and unplanned outages across distributed servers and IBM Z platforms. GDPS can be configured in either a single site or in a multisite configuration. It is designed to manage remote copy configuration between storage subsystems, automate Parallel Sysplex operational tasks, and perform failure recovery from a single point of control, improving application availability.

Historically, this solution was known as a Geographically Dispersed Parallel Sysplex. Today, GDPS continues to be applied as a general term for a suite of business continuity solutions, including those that do not require a dispersed or multisite sysplex environment.
GDPS supports the following forms of remote copy in multisite solutions:

- IBM Metro Mirror, synchronous Peer-to-Peer Remote Copy
- IBM Global Mirror, asynchronous Peer-to-Peer Remote Copy
- IBM z/OS Global Mirror, asynchronous XRC

The GDPS solution is also independent of disk vendors if the vendor meets the specific levels of IBM Metro Mirror, IBM Global Mirror, and IBM z/OS Global Mirror architectures. For more information, see the GDPS web page.

IBM supports only WDM products that are IBM Z qualified for use in GDPS solutions. To obtain this qualification, WDM vendors obtain licensed IBM patents, intellectual property, and know-how that are related to the GDPS architecture. This access allows vendors to use proprietary IBM protocols and applications that are used in a GDPS environment, including coupling and STP links, Metro Mirror, Global Mirror, and z/OS Global Mirror.

Licensing of IBM patents also provides the WDM vendor with technical information about future IBM releases. Qualified vendors typically license this information for an extended period, which allows them to subscribe to the latest GDPS architecture changes and to be among the first to market with offerings that support these features.

**Note:** Check with your WDM vendor for current licensing status.

In addition, these vendor products are tested and qualified by IBM technicians with the same test environment and procedures that are used to test the protocols that provide the required connectivity for a GDPS configuration. This testing includes functions, recovery, and, in certain cases, performance measurements. Having access to these test facilities allows IBM to configure a fully functional sysplex and to simulate failure and recovery actions that cannot be tested as part of a working customer environment.

IBM has the facilities to test and qualify these products with both current and previous generation equipment within the IBM Vendor Solutions Connectivity Lab in Poughkeepsie, New York, in the United States. This qualification testing allows IBM specialists to reproduce any concerns that might arise when using this equipment in a client’s application.

**Components**
The following GDPS components are used during the qualification process:

- IBM Parallel Sysplex
- IBM System Storage
- Optical Wavelength-Division Multiplexer (WDM)
- IBM System Storage Metro Mirror (PPRC), a synchronous form of remote copy
- IBM System Storage Global Mirror
- IBM System Storage z/OS Global Mirror (XRC), an asynchronous form of remote copy

**Protocols**
The following GDPS connectivity protocols are tested during the qualification process:

- Fibre Connection (FICON)
- Fibre Channel protocol (FCP)
- Fibre Channel InterSwitch Links (ISL)
- InterSystem Channel-3 (ISC-3) peer mode (2 Gbps)
- InterSystem Channel-3 (ISC-3) peer mode (1 Gbps) using RPQ 8P2197
- Server Time Protocol (STP)
10.4.2 IBM Z qualified WDM vendor products

The latest list of qualified WDM vendor products can be found through IBM Resource Link.

Select Hardware products for servers in the resource link, and then, select the page titled System z Qualified Wavelength Division Multiplexer (WDM) products for GDPS solutions.

Note: It is important to select the particular WDM vendor link in Resource Link and download the qualification letter to verify the details about the WDM product, model, firmware level, and the IBM Z server models for which it is qualified.

10.5 References

The following publications contain information that is related to fiber optic link distance:

- Coupling Facility Channel I/O Interface Physical Layer, SA23-0395
- Planning for Fiber Optic Links, GA23-0367
- Fiber Transport Services Direct Attach Planning, GA22-7234

For more information about IBM Z connectivity, see this web page.

For more information about GDPS solutions, see these resources:

- GDPS home page
- Parallel Sysplex home page
- IBM GDPS Family: An introduction to Concepts and Capabilities, SG24-6374

For more information about IBM Z qualified WDM vendor products, use this IBM Redbooks publications search result.
Appendix A. Cryptographic features

This appendix briefly describes the optional Peripheral Component Interconnect Express (PCIe) cryptographic features of IBM Z platforms. It includes the following topics:

- Overview
- Crypto Express6S
- Crypto Express5S
- Crypto Express4S
- Crypto Express3
- References
Overview

Public Key Cryptography Standards (PKCS) #11\(^1\) and the IBM Common Cryptographic Architecture (CCA) define various cryptographic functions, external interfaces, and a set of key cryptographic algorithms. These specifications provide a consistent, end-to-end cryptographic architecture across IBM z/OS, IBM AIX®, and IBM i (formerly i5/OS) operating systems and other platforms, including Linux and Microsoft Windows.

The following cryptographic external interfaces are part of the IBM Z environment:

- Central Processor Assist for Cryptographic Function (CPACF-CP Assist). CPACF-CP Assist offers a set of symmetric cryptographic functions for high encrypting and decrypting performance of clear key operations. This interface is for Secure Sockets Layer/Transport Layer Security (SSL/TLS), VPN, and data-storing applications that do not require US Federal Information Processing Standard (FIPS\(^2\)) 140-2 Level 4 security. The CP Assist for Cryptographic Function is integrated with the compression unit in the coprocessor in the core of the IBM Z.

Each coprocessor in an IBM Z platform consists of one compression (and expansion) unit, one cryptographic cipher engine, and one cryptographic hash engine. The cryptographic engine is embedded in each processing unit (PU core) of the IBM Z PU chip. On IBM Z before z EC12, this duty was shared between two processing units (cores).

For z13 and z13s, the COP has been redesigned to support SMT operation. For throughput increase, the z14 COP was further developed so that the coprocessor results are now stored directly in the L1 cache of the core. Figure A-1 compares the logical flow of the coprocessor on the chip for z13, z13s, z14 and z14 ZR1.

Crypto Express features are optional and available in different generations. All Crypto Express features can be configured during installation as either a secure coprocessor or an accelerator. The type of the Crypto Express feature can be changed after installation from the initial mode in a disruptive process.

---

\(^1\) One of the industry-accepted Public Key Cryptographic Standards (PKCS) provided by RSA Laboratories from RSA, the security division of Dell EMC Corporation.

The support of the different generations of Crypto Express features depends on the IBM Z server generation.

Crypto Express features provide a secure hardware and programming environment for cryptographic processes. Each cryptographic coprocessor includes a general-purpose processor, non-volatile storage, and specialized cryptographic electronics.

- The cryptographic coprocessor functions are supported by the Integrated Cryptographic Service Facility, a component of the z/OS operating system, and by the IBM Common Cryptographic Architecture Support Program for Linux on IBM Z.

**Crypto Express6S**

The Crypto Express6S feature is supported only on the z14. It has the following characteristics:

- It occupies one I/O slot in the PCIe I/O drawer.
- It has one PCIe adapter, with one PCHID assigned to it according to its physical location in the PCIe I/O drawer.
- On a z14 each Crypto Express6S PCIe adapter can be configured in one of the following modes:
  - Secure IBM CCA coprocessor (CEX6C) for FIPS 140-2 Level 4 certification. This mode includes secure key functions. The Crypto Express6s supports user-defined extensions (UDX), which you can use to define and load customized cryptographic functions.
  - Secure IBM Enterprise PKCS #11 (EP11) coprocessor (CEX6P) implements an industry-standardized set of services that adhere to the PKCS #11 specification v2.20. This cryptographic coprocessor mode introduced the PKCS #11 secure key function. A Trusted Key Entry (TKE) workstation is required to support the administration of the Crypto Express4S when it is configured in EP11 mode.
  - Accelerator (CEX6A) for acceleration of public key and private key cryptographic operations that are used with SSL/TLS processing.

These modes can be configured by using the Support Element. The PCIe adapter must be configured offline to change the mode.

**Note:** When the Crypto Express6S PCIe adapter is configured as a secure IBM CCA coprocessor, it still provides accelerator functions. However, you can achieve up to three times better performance for those functions if the Crypto Express6S PCIe adapter is configured as an accelerator.

- Up to 16 Crypto Express6S features are supported (16 PCIe adapters per z14).
- The Crypto Express6S adapters support up to 85 domains on z14 and up to 40 domains on z14 ZR1 for logical partitions.
- In z/VM, up to 85 domains on z14 and up to 40 domains on z14 ZR1 for logical partitions or z/VM guests are supported. This enhancement is based on the new Adjunct Processor Extended Addressing (APXA), which enables the z/Architecture to support up to 256 domains in an Adjunct Processor (AP).
Crypto Express5S

The Crypto Express5S feature is supported on the z14 and z14 ZR1 (carry forward only), and also z13 and z13s. It has the following characteristics:

- It occupies one I/O slot in the PCIe I/O drawer.
- It has one PCIe adapter, with one PCHID assigned to it according to its physical location in the PCIe I/O drawer.
- On a z14 and z14 ZR1 (carry forward only), and also z13 or z13s, each Crypto Express5S PCIe adapter can be configured in one of the following modes:
  - Secure IBM CCA coprocessor (CEX5C) for FIPS 140-2 Level 4 certification. This mode includes secure key functions. The Crypto Express5S supports UDX, which you can use to define and load customized cryptographic functions.
  - Secure IBM Enterprise PKCS #11 (EP11) coprocessor (CEX5P) implements an industry-standardized set of services that adhere to the PKCS #11 specification v2.20. This cryptographic coprocessor mode introduced the PKCS #11 secure key function. A TKE workstation is required to support the administration of the Crypto Express5S when it is configured in EP11 mode.
  - Accelerator (CEX5A) for acceleration of public key and private key cryptographic operations that are used with SSL/TLS processing.

These modes can be configured by using the Support Element. The PCIe adapter must be configured offline to change the mode.

Note: When the Crypto Express5S PCIe adapter is configured as a secure IBM CCA coprocessor, it still provides accelerator functions. However, you can achieve up to three times better performance for those functions if the Crypto Express5S PCIe adapter is configured as an accelerator.

- Up to 16 Crypto Express5S features are supported per z14, z14 ZR1, z13, or z13s.
- The Crypto Express5S adapters support up to 85 domains on z14 and z13, and 40 domains on z14 ZR1 and z13s.
- Started with z13, up to 85 domains (up to 40 for z13s) for logical partitions or z/VM guests is supported. This enhancement is based on the new APXA, which enables the z/Architecture to support up to 256 domains in an AP.

Crypto Express4S

The Crypto Express4S feature is supported on the zEC12 and zBC12. It has the following characteristics:

- It occupies one I/O slot in the PCIe I/O drawer.
- It has one PCIe adapter, with one PCHID assigned to it according to its physical location in the PCIe I/O drawer.
- On the zEC12 and zBC12, each Crypto Express4S PCIe adapter can be configured in one of the following modes:
  - Secure IBM CCA coprocessor (CEX4C) for FIPS 140-2 Level 4 certification. This mode includes secure key functions. The Crypto Express4S supports UDX, which you can use to define and load customized cryptographic functions.
– Secure IBM Enterprise PKCS #11 (EP11) coprocessor (CEX4P) implements an industry-standardized set of services that adhere to the PKCS #11 specification v2.20 and more recent amendments. It was designed for extended FIPS and Common Criteria evaluations to meet public sector requirements.

This new cryptographic coprocessor mode introduced the PKCS #11 secure key function.

A TKE workstation is required to support the administration of the Crypto Express4S when it is configured in EP11 mode.

– Accelerator (CEX4A) for acceleration of public key and private key cryptographic operations that are used with SSL/TLS processing.

These modes can be configured by using the Support Element. The PCIe adapter must be configured offline to change the mode.

**Note:** When the Crypto Express4S PCIe adapter is configured as a secure IBM CCA coprocessor, it still provides accelerator functions. However, you can achieve up to three times better performance for those functions if the Crypto Express4S PCIe adapter is configured as an accelerator.

▶ Up to 16 Crypto Express4S features are supported.

### Crypto Express3

The Crypto Express3 feature is supported on zEC12 and zBC12 servers. It is available on a carry-forward only basis when you are upgrading from earlier generations to zEC12 and zBC12. Crypto Express3 has the following characteristics:

▶ The Crypto Express3 feature occupies one I/O slot in an I/O cage or in an I/O drawer.

▶ The Crypto Express3 feature has two PCIe adapters\(^3\), with two PCHIDs assigned to it according to its physical location in the I/O cage or I/O drawer.

▶ There is no need to define a CHPID for the Crypto Express3 feature in the HCD/IOCP. Do not allow a Crypto Express3-associated PCHID to be used by another device in the HCD/IOCP definition.

▶ On Z platforms, each Crypto Express3 PCIe adapter can be configured in one of the following modes:

  – Secure coprocessor for FIPS 140-2 Level 4 certification. This mode includes secure key functions. You have the option to program it to deploy additional functions and algorithms by using a UDX.

  – Accelerator for public key and private key cryptographic operations that are used with SSL/TLS processing.

  – These modes can be configured by using the Support Element. The PCIe adapter must be configured offline to change the mode.

**Note:** When the Crypto Express3 PCIe adapter is configured as a secure coprocessor, it still provides accelerator functions. However, you can achieve up to three times better performance for those functions if the Crypto Express3 PCIe adapter is configured as an accelerator.

\(^3\) The Crypto Express3-1P feature has one PCIe adapter and just one associated physical channel ID (PCHID).
Up to eight Crypto Express3 features (16 PCIe adapters) for zEC12 or z BC12 CPCs.

Up to eight Crypto Express3-1P features (eight PCIe adapters) for each IBM zEnterprise BC12 (zBC12).

**Note:** The Crypto Express3-1P feature is exclusive to the zBC12 and the features can be ordered just in an MES process.

### References

For more information, see the following publications:

- IBM zNextTechnical Introduction, SG24-8450
- IBM zNext Configuration Setup, SG24-8460
- IBM z13 and IBM z13s Technical Introduction, SG24-8250
- IBM z13 Technical Guide, SG24-8251
- IBM z13s Technical Guide, SG24-8294
- IBM z13 Configuration Setup, SG24-8260
- IBM zEnterprise System Technical Introduction, SG24-8050
- IBM zEnterprise EC12 Technical Guide, SG24-8049
- IBM zEnterprise BC12 Technical Guide, SG24-8138
- IBM zEnterprise EC12 Configuration Setup, SG24-8034
zEDC Express feature

This chapter describes the optional IBM zEnterprise Data Compression (zEDC) Express feature of the Z platform. The information covers the following topics:

- Benefits of using the zEDC Express feature
- zEDC Express operation
- Software support
Benefits of using the zEDC Express feature

The growth of data that needs to be captured, transferred, and stored for long periods of time is unrelenting. Software-implemented compression algorithms are costly in terms of processor resources. Storage costs are not negligible, either.

The optional zEDC Express feature (FC 0420) addresses those requirements by providing hardware-based acceleration for data compression and decompression. When used with IBM z/OS version 2.1, the zEnterprise data compression acceleration capability of the zEDC Express feature reduces processor use, optimizes performance of compression-related tasks, and enables more efficient use of storage resources. This integrated solution also lowers the cost of computing and optimizes the cross-platform exchange of data.

zEDC Express operation

The feature installs exclusively in the PCIe I/O drawer. Up to two zEDC Express features can be installed per PCIe I/O drawer domain. However, if the domain contains a Flash Express or 10GbE RoCE feature, then only one zEDC Express feature can be installed on that domain. For more information, see Chapter 7, “Shared Memory Communications” on page 125.

Between one and eight\(^1\) features can be installed on the system. There is one PCIe adapter/compression coprocessor per feature, which implements compression as defined by RFC1951 (DEFLATE).

A zEDC Express feature can be shared by up to 15 logical partitions (LPARs).

**zEDC RAS considerations for zEC12 and zBC12**

Adapter support for zEDC is provided by Resource Group code that is running on the system’s integrated firmware processor (IFP). For resilience, on zEC12 and zBC12 there are always two independent Resource Groups on the system, and they share the firmware processor. Therefore, a minimum of two zEDC Express features should be installed, with one for each group.

Also, consider the total data throughput that is required. If one feature becomes unavailable, the others must be able to absorb the load. For best data throughput and availability, install at least two features per Resource Group.

**zEDC RAS considerations for z14, z14 ZR1, z13, and z13s**

Availability rules have been relaxed for z14, z14 ZR1, z13, and z13s servers with the introduction of four Resource Groups. With four features installed, if one resource Group becomes unavailable due to firmware maintenance, only one zEDC feature is taken offline, the remaining three are capable of handling the workload with less performance impact.

For more information on resource group allocation for PCIe I/O drawers and PCIe+ I/O drawers (FC #4001) see the following publications:

- IBM zEC12 Technical Guide, SG24-8049
- IBM z13 Technical Guide, SG24-8251
- IBM z14 (M/T 3906) Technical Guide, SG24-8451
- IBM z14 ZR1 Technical Guide, SG24-8651

\(^1\) A maximum of 16 zEDC Express features can be installed in the z14, and a maximum of 8 can be installed on z14 ZR1.
Software support

Support of zEDC Express functions is provided by IBM z/OS 2.1 and later for data compression and decompression.

More and more programs are enabled to use zEDC. At the time of writing, the following programs and subsets of z/OS are supported:

- IBM Sterling Connect. Direct for z/OS V5.2 can automatically use zEDC Express Accelerator for file compression and decompression as files are transferred cross-platform.
- Sequential data that uses BSAM/QSAM extended format.
- IBM Data Facility Storage Management Subsystem (DFSM), Data Set Services (DFSMsDss), and DFSM Hierarchical Storage Management (DFSMShsm).

z/VM V6.3 with PTFs and later provide guest exploitation support of the zEDC Express feature (0420) on the Z platform.

IBM Z Batch Network Analyzer

The IBM Z Batch Network Analyzer (zBNA) is a PC-based productivity tool for estimating the elapsed time for batch jobs. This estimation is based on the differences in processor speeds of a base processor and a target processor, the number of engines on each system, and system capacities. Data sharing is not factored into the estimate. It is available as-is, without charge.

zBNA replaces the BWATOOL. It works with the Microsoft Windows operating system and provides graphs and text reports, including Gantt charts, and support for alternative processors.

zBNA can be used to analyze your SMF records to identify jobs and data sets that are candidates for zEDC compression, across a specified time (typically, in batch mode). zBNA generates lists of data sets by job:

- Those that already use hardware compression and might be candidates for zEDC
- Those that might be zEDC candidates but are not in extended format

You can use the zBNA tool to determine which zEDC features that you need to use.

IBM Business Partners can download zBNA and related CPS tools from this IBM PartnerWorld® web page, which requires registration (at no charge).

IBM employees can download zBNA and related CPS tools from the IBM Z Batch Network Analyzer (zBNA) Tool web page.
Flash Express feature

This appendix briefly describes the optional Flash Express feature of the z13, z13s, zEC12, and zBC12 platforms. The following topics are covered in this appendix:

- Overview
- Flash Express

Replacement of Flash Express

Note: Starting with the IBM z14, Flash Express (FC #0402 and FC 0403) was replaced by Virtual Flash memory (VFM). VFM implements Extended Asynchronous Data Mover (EADM) architecture by using HSA-like memory instead of Flash card pairs.

The VFM features are available in the following sizes:

- For z14, one VFM feature is 1.5 TB (FC #0604). Up to four features per system can be ordered.
- For z14 ZR1, one VFM feature is 512 GB (FC #0614). Up to four features per system can be ordered.
Overview

Flash Express is an innovative, optional feature that is available on z13, z13s, zEC12, and zBC12 servers. It improves performance and availability for critical business workloads that cannot afford any reduction in service levels. Flash Express is easy to configure, requires no special skills, and provides rapid time to value.

Flash Express is internal storage that is implemented with solid-state drives (SSDs) mounted in PCIe Flash Express features. Flash Express features are ordered in pairs for availability. A pair provides 1.4 TB of storage. A maximum of four pairs are supported in z13, z13s, zEC12, and zBC12. Flash Express storage is allocated to each partition and is similar to main memory allocation. The allocation is specified at the Support Elements (SE) or Hardware Management Console (HMC).

The IBM z/OS operating system uses the Flash Express feature as storage-class memory (SCM). The Coupling Facility Control Code (CFCC) can also use Flash Express as SCM. For the CF, flash memory provides a standby capacity for use by IBM MQ shared queues structures. This technique provides protection against transient backups in the shared queue from causing the structure to fill up.

Flash Express provides immediate benefits. It is used by the real storage manager (RSM) for z/OS paging and supervisor call (SVC) memory dumps. It reduces processor cycles from paging and improves availability by reducing SVC memory dump duration due to page-in of AUX storage. The value of Flash Express is not in replacing memory with flash, but replacing disk with it. It provides support for pageable large pages (1 MB). If contiguous flash space is available, pageable large pages are preferentially written to flash. If available space exists on both flash and disk, makes a selection that is based on response time.

The Flash Express feature (0403) on z13 (new build), z13s, and (0402) on zEC12, and zBC12 with CFCC level 19 or later is used as a resilient solution for overflow of IBM MQ shared queues in the coupling facility. This new function allows list structure data to be migrated to Flash Express memory as needed. It is needed when the consumers of data do not keep pace with data creators for some reason, and then migrate it back to real memory to be processed.

Flash Express

Flash Express is supported on z13, z13s, zEC12, and zBC12 servers:

- The Flash Express feature occupies one I/O slot in a PCIe I/O drawer.
- The Flash Express feature is always installed in pairs. The minimum order is two features and the maximum order is eight features, in increments of two features. A maximum of four pairs (4 x 1.4 = 5.6 TB) are supported in a system.
- One PCIe I/O drawer can support a maximum of two pairs with each Flash Express feature installed in a unique domain.
- A second PCIe I/O drawer is required to support more than two Flash Express pairs.
- One Flash Express pair is installed in the front slots, 1 and 14. The other pair is installed in the rear slots, 25 and 33.
- Flash Express slot positions are reserved and used only when there are no spare slots available.
- The Flash Express feature has no CHPID assigned. A Flash Express feature has a PCHID assigned to it according to its physical location in the drawer.
There is no need to define a CHPID for the Flash Express feature in the configuration definition (HCD) or input/output configuration program (IOCP). Take care to avoid the use of Flash Express-associated PCHIDs by another device in the HCD or IOCP definitions.

Flash Express subchannels are predefined and allocated from the 250 addresses that are reserved in subchannel set 0.

Flash Express is accessed by using the Z Extended Asynchronous Data Mover (EADM) function. Access to Flash Express is initiated with a Start Subchannel instruction.

Data is encrypted on the Flash Express adapter with AES key encryption. Encryption keys are stored on smart cards that are plugged into the server’s Support Element. Removing the smart cards renders the data on the Flash Express adapter inaccessible.

Flash Express features are connected to each other by two external cables to form a RAID 10 mirror for redundancy.
Channel conversion solutions

This appendix describes the possibilities of conversion from FICON channel connectivity to ESCON or connectivity to Bus and Tag (B/T) devices (parallel channel). An ESCON channel feature is not supported on any marketable IBM mainframes.
Conversion solutions

The following are the available solutions for channel conversion.

**FICON to ESCON conversion**

Prizm Protocol Convert from Optica Technologies Inc. provides a FICON-to-ESCON conversion function that has been qualified for use with IBM Z. To view qualification letters, see the IBM Z I/O Connectivity web page.

Select the Products tab, then **FICON / FCP Connectivity**. Scroll down to the “other supported devices” area on the web page.

The Optica Prizm converter can be implemented in point-to-point, switched, and cascaded FICON configurations as shown in Figure D-1.

![Figure D-1 Optica Prizm possible configurations](image-url)
FICON to Bus and Tag conversion

Also, for IBM Z platforms that still require connectivity to B/T devices, a combination of two Optica converters can be used: The Optica Prizm to convert FICON to ESCON and the Optica ESBT to convert ESCON to B/T that is operating for devices in block mode only, as shown in the Figure D-2.

For more information about Prizm Protocol Convert, see the Optica website.

Note: IBM cannot confirm the accuracy of compatibility, performance, or any other claims by vendors for products that have not been qualified for use with IBM Z. Address questions regarding these capabilities and device support to the suppliers of those products.

IBM Facilities Cabling Services ESCON-to-FICON migration services can help you take advantage of high-speed FICON to support an upgraded IBM Z environment. Also, it is possible to keep using ESCON-attached devices to reduce migration costs.
Channel feature attributes

This appendix lists the cable types and attributes of channel features that are supported on the Z platform. Not all channel features can be ordered for all server families. Certain features are available only when carried forward on a server upgrade.
Cable types and attributes

Table E-1 lists the cable types and attributes of the channel types that are supported on the Z platform.

**Note:** Table 1-1 on page 10 lists the maximum number of channels or ports that are supported per server.

The connector type for most fiber optic cable types is LC duplex, with the following exceptions:

- **zHyperLink Express**  MTP connector
- **12x IFB**  MPO connector
- **ICA SR**  MTP connector
- **1000BASE-T Ethernet**  RJ-45 connector (UTP copper Ethernet cable)

For more information about fiber optic cables and connector types, see Appendix F, “Fiber optic cables” on page 207.

The footnotes in Table E-1 reflect special considerations for certain features. Check the referenced chapters for in-depth information.

The Carry forward only withdrawn from marketing (WDFM) column indicates special conditions for each feature. WDFM means that the feature is supported on the platform. However, because the feature at this platform is withdrawn from marketing, no ordering is possible. Carry forward means that the feature can be transferred to the new platform only by using the miscellaneous equipment specification (MES) process. The feature cannot be ordered for a new build system. For more information about support of extended distances, see “Extended distance solutions” on page 169.

<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distance(^a)</th>
<th>Platform</th>
<th>NB(^c), or Carry forward</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Storage Connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zHyperLink</td>
<td>0431</td>
<td>8 Gbps</td>
<td>MM 50 µm OM3 OM4</td>
<td>100 m (2000) 150 m (4700)</td>
<td>z14</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S+ 10 KM LX</td>
<td>0427</td>
<td>4, 8, or 16 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14</td>
<td></td>
</tr>
<tr>
<td>FICON Express16S+ SX</td>
<td>0428</td>
<td>16 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>15 m (200) 35 m (500) 100 m (2000) 125 m (4700)</td>
<td>z14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>21 m (200) 50 m (500) 150 m (2000) 190 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200) 150 m (500) 380 m (2000) 400 m (4700)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Table E-1 on page 10 lists the maximum number of channels or ports that are supported per server.

---

200  IBM Z Connectivity Handbook
<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distance&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Platform</th>
<th>NB&lt;sup&gt;2&lt;/sup&gt;, or Carry forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICON Express16S</td>
<td>0418</td>
<td>4, 8, or 16 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14*, z14 ZR1* z13, z13s</td>
<td>*Carry forward</td>
</tr>
<tr>
<td>10 KM LX</td>
<td></td>
<td></td>
<td>MM 62.5 µm</td>
<td>15 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>35 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>21 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>190 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>380 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express16S SX</td>
<td>0419</td>
<td>16 Gbps</td>
<td>MM 62.5 µm</td>
<td>z14*, z14 ZR1* z13, z13s</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>z14*, z14 ZR1* z13, z13s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td>FICON Express8S</td>
<td>0409</td>
<td>2, 4, or 8 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14*, z14 ZR1* z13, z13s*</td>
<td>*Carry forward</td>
</tr>
<tr>
<td>10 KM LX</td>
<td></td>
<td></td>
<td>MM 62.5 µm</td>
<td>21 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>50 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>380 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>150 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express8S SX</td>
<td>0410</td>
<td>8 Gbps</td>
<td>MM 62.5 µm</td>
<td>z14*, z14 ZR1* z13, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>z14*, z14 ZR1* z13, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express8</td>
<td>3325</td>
<td>2, 4, or 8 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
</tr>
<tr>
<td>10 KM LX</td>
<td></td>
<td></td>
<td>MM 62.5 µm</td>
<td>21 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>50 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>380 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400 m (4700)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>150 m (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express8 SX</td>
<td>3326</td>
<td>8 Gbps</td>
<td>MM 62.5 µm</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3 OM4</td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>z13*, z13s*</td>
<td>*Carry forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel feature</td>
<td>Feature codes</td>
<td>Bit rate</td>
<td>Cable type</td>
<td>Maximum unrepeated distance&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Platform</td>
<td>NB&lt;sup&gt;b&lt;/sup&gt;, or Carry forward</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>----------</td>
<td>------------</td>
<td>----------------------------------------</td>
<td>----------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>FICON Express4-2C SX</td>
<td>3318</td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200) 150 m (500) 380 m (2000) 400 m (4700)</td>
<td>zBC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>300 m (200) 500 m (500) 860 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICON Express4 10 KM LX</td>
<td>3321</td>
<td>1, 2, or 4 Gbps</td>
<td>SM 9 µm</td>
<td>10 km/20 km&lt;sup&gt;c&lt;/sup&gt;</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td>FICON Express4 SX</td>
<td>3322</td>
<td>4 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3 OM4</td>
<td>70 m (200) 150 m (500) 380 m (2000) 400 m (4700)</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>150 m (200) 300 m (500) 500 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>300 m (200) 500 m (500) 860 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OSA-Express</strong></td>
<td>Chapter 5, “Open Systems Adapter-Express” on page 71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE LR</td>
<td>0424</td>
<td>10 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S 10 GbE SR</td>
<td>0425</td>
<td>10 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>33 m 82 m (500) 300 m (2000)</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S GbE LX</td>
<td>0422</td>
<td>1 Gbps</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCP 50 µm</td>
<td>550 m (500)</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S GbE SX</td>
<td>0423</td>
<td>1 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM3</td>
<td>275 m (200) 550 m (500)</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td>OSA-Express6S 1000BASE-T Ethernet</td>
<td>0426</td>
<td>100/1000 Mbps</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14, z14 ZR1*</td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE LR</td>
<td>0415</td>
<td>10 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>zEC12, zBC12, z13, z13s, z14*, z14 ZR1*</td>
<td>Carry forward</td>
</tr>
<tr>
<td>OSA-Express5S 10 GbE SR</td>
<td>0416</td>
<td>10 Gbps</td>
<td>MM 62.5 µm MM 50 µm OM2 OM3</td>
<td>33 m 82 m (500) 300 m (2000)</td>
<td>zEC12, zBC12, z13, z13s, z14*, z14 ZR1*</td>
<td>Carry forward</td>
</tr>
</tbody>
</table>

<sup>a</sup>Maximum unrepeated distance varies depending on cable type.

<sup>b</sup>NB: New.

<sup>c</sup>Distance range for SM 9 µm.

<sup>d</sup>Cable type and distance range vary depending on platform.
<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distance</th>
<th>Platform</th>
<th>NB², or Carry forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA-Express5S GbE LX</td>
<td>0413</td>
<td>1 Gbps</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>zEC12, zBC12, z13, z13s, z14*, z14 ZR1*</td>
<td>Carry forward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCP 50 µm</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S GbE SX</td>
<td>0414</td>
<td>1 Gbps</td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td>z14*, z14 ZR1*, z13, z13s, zEC12, zBC12, z14 ZR1*</td>
<td>Carry forward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM3</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express5S 1000BASE-T Ethernet</td>
<td>0417</td>
<td>100/1000 Mbps</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14*, z14 ZR1*, z13, z13s, zEC12, zBC12</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE LR</td>
<td>0406</td>
<td>10 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>z14 ZR1*, z13, z13s, zEC12, zBC12, z14 ZR1*</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td>OSA-Express4S 10 GbE SR</td>
<td>0407</td>
<td>10 Gbps</td>
<td>MM 62.5 µm</td>
<td>33 m</td>
<td>z14 ZR1*, z13, z13s, zEC12, zBC12</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3</td>
<td>82 m (500) 300 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S GbE LX</td>
<td>0404</td>
<td>1 Gbps</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>z14 ZR1*, z13, z13s, zEC12, zBC12</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCP 50 µm</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S GbE SX</td>
<td>0405</td>
<td>1 Gbps</td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td>z14 ZR1*, z13, z13s, zEC12, zBC12</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM3</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express4S 1000BASE-T Ethernet</td>
<td>0408</td>
<td>10/100/1000 Mbps</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>z14*, z13*, z13s* zEC12, zBC12</td>
<td>Carry forward, WDFM</td>
</tr>
<tr>
<td>OSA-Express3 10 GbE LR</td>
<td>3370</td>
<td>10 Gbps</td>
<td>SM 9 µm</td>
<td>10 km</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td>OSA-Express3 10 GbE SR</td>
<td>3371</td>
<td>10 Gbps</td>
<td>MM 62.5 µm</td>
<td>33 m</td>
<td>zEC12, zBC12</td>
<td>Carry forward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3</td>
<td>82 m (500) 300 m (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express3 GbE LX</td>
<td>3362</td>
<td>1 Gbps</td>
<td>SM 9 µm</td>
<td>5 km</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCP 50 µm</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express3 GbE SX</td>
<td>3363</td>
<td>1 Gbps</td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM3</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express3-2P GbE SX</td>
<td>3373</td>
<td>1 Gbps</td>
<td>MM 62.5 µm</td>
<td>275 m (200)</td>
<td>zBC12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM3</td>
<td>550 m (500)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSA-Express3 1000BASE-T Ethernet</td>
<td>3367</td>
<td>10/100/1000 Mbps</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td>Channel feature</td>
<td>Feature codes</td>
<td>Bit rate</td>
<td>Cable type</td>
<td>Maximum unrepeated distancea</td>
<td>Platform</td>
<td>NBc, or Carry forward</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>----------</td>
<td>------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>----------------------</td>
</tr>
<tr>
<td>OSA-Express3-2P 1000BASE-T Ethernet</td>
<td>3369</td>
<td>10/100/ 1000 Mbps</td>
<td>UTP Cat5 or 6</td>
<td>100 m</td>
<td>zBC12</td>
<td></td>
</tr>
</tbody>
</table>

**RoCE and SMC-Direct**  
Chapter 7, “Shared Memory Communications” on page 125

<table>
<thead>
<tr>
<th>10GbE RoCE Express2</th>
<th>0412</th>
<th>10 Gbps</th>
<th>MM 62.5 µm</th>
<th>33 m</th>
<th>z14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3</td>
<td>82 m (500) 300 m (2000)</td>
<td></td>
</tr>
<tr>
<td>10GbE RoCE Express</td>
<td>0411</td>
<td>10 Gbps</td>
<td>MM 62.5 µm</td>
<td>33 m</td>
<td>z14*, z14 ZR1* z13, z13s zEC12, zBC12,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MM 50 µm OM2 OM3</td>
<td>82 m (500) 300 m (2000)</td>
<td>*Carry forward</td>
</tr>
<tr>
<td>SMC-Direct</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>z14, z14 ZR1* z13b, z13s</td>
</tr>
</tbody>
</table>

**HiperSockets**  
Chapter 8, “HiperSockets technology” on page 135

| HiperSockets        | n/a  | n/a     | n/a        | n/a  | z14, z14 ZR1* z13, z13s zEC12, zBC12, |

**Coupling links**  
Chapter 9, “Coupling links and common time” on page 153

| IC                 | n/a  | n/a     | n/a        | n/a  | z14, z14 ZR1* z13, z13s, zEC12, zBC12, |
|                    |      |         |            |      | |
| CE LR              | 0433 | 10 Gbps | SM 9 µm    | 10 km | z14, z14 ZR1* z13b, z13s  |
|                    |      |         |            |      | NBc, Carry forward |
| ICA SR             | 0172 | 8 GBps  | MM 50 µm OM3 OM4 | 100 m (2000) 150 m (4700) | z14, z14 ZR1* z13, z13s |
|                    |      |         |            |      | NBc, Carry forward |
| HCA3-O LR (1x IFB) | 0170 | 5 Gbps or 2.5 Gbps | SM 9 µm | 10 km | z14, z13, z13s, zEC12, zBC12, |
|                    |      |         |            |      | NBc, Carry forward |
| HCA3-O (12x IFB)   | 0171 | 6 GBps  | MM 50 µm OM3 | 150 m (2000) | z14, z13, z13s, zEC12, zBC12, |
|                    |      |         |            |      | NBc, Carry forward |
| HCA2-O LR (1x IFB) | 0168 | 5 Gbps or 2.5 Gbps | SM 9 µm | 10 km | zEC12, zBC12 |
| HCA2-O (12x IFB)   | 0163 | 6 GBps or 3 GBps | MM 50 µm OM3 | 150 m (2000) | zEC12, zBC12 |
| ISC-3              | 0217d 0218 0219 | 2 Gbps | SM 9 µm MCP 50 µm | 10 km/20 kmc 550 m (500) | zEC12, zBC12 |
| ISC-3 (RPQ 8P2197 )|      | 1 Gbps  | SM 9 µm    | 20 kmc | |

a Maximum unrepeated distance refers to the maximum distance over which the signal can travel without degradation. b Carry forward for previous model compatibility. c NBc: Not Bifurcated.
### Channel feature attributes

<table>
<thead>
<tr>
<th>Channel feature</th>
<th>Feature codes</th>
<th>Bit rate</th>
<th>Cable type</th>
<th>Maximum unrepeated distance&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Platform</th>
<th>NB&lt;sup&gt;b&lt;/sup&gt;, or Carry forward</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crypto</strong></td>
<td>Appendix A, “Cryptographic features” on page 181</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crypto Express6S</td>
<td>0893</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>z14, z14 ZR1</td>
<td></td>
</tr>
<tr>
<td>Crypto Express5S</td>
<td>0890</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>z14*, z14 ZR1&lt;sup&gt;*&lt;/sup&gt;</td>
<td>*Carry forward</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>z13, z13s</td>
<td></td>
</tr>
<tr>
<td>Crypto Express4S</td>
<td>0865</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td>Crypto Express3</td>
<td>0864</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
<tr>
<td>Crypto Express3-1P</td>
<td>0871</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>zBC12</td>
<td>Carry forward</td>
</tr>
<tr>
<td><strong>zEDC Express</strong></td>
<td>Appendix B, “zEDC Express feature” on page 187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zEDC Express</td>
<td>0420</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>z14, z14 ZR1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>z13, z13s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>zEC12, zBC12</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses where applicable.

<sup>b</sup> See the HMC driver level for feature support.

<sup>c</sup> NB - new build. Features can be added for a new build system or as carry forward during an MES / Upgrade.

<sup>d</sup> There are two feature codes for the ISC-3:
- Feature code 0217 is for the ISC Mother card (ISC-M).
- Feature code 0218 is for the ISC Daughter card (ISC-D).
Fiber optic cables

This appendix describes the physical attributes of optical fiber technologies that are supported on Z platforms. The following topics are covered in this appendix:

- Description
- Connector types for fiber cables
- Mode-conditioning patch cables
- InfiniBand cables
- zHyperLink Express and ICA SR cables
- Conversion kits
Description

Fiber optic cables use light for data transmission, rather than electrical current on copper cables. Fiber optic cables have many advantages:

- Many times lighter and have substantially less bulk
- No pins
- A smaller and more reliable connector
- Reduced loss and distortion
- Free from signal skew and the effects of electro-magnetic interference

Figure F-1 illustrates the two types of optical fiber that are used in a data center environment with Z platforms:

- Multimode
- Single mode

The difference between them is the way that light travels along the fiber. Multimode has multiple light paths, but single mode has only one light path.

Each fiber type consists of three parts:

- The core can be 50 or 62.5 µm in diameter for multimode or 9 µm in diameter for single mode.
- The cladding that surrounds the core is 125 µm in diameter.
- The outer coating is 250 µm in diameter.

Figure F-1  Fiber optic cable types

Note: To keep the data flowing, thorough cleaning of fiber optic connectors is critical. Make sure that you have the necessary fiber optic cleaning procedures in place.
Connector types for fiber cables

For all optical links, the connector type is LC duplex, except the ESCON connector, which has an MT-RJ type connector, and 12x IFB, which has an MPO connector.

Figure F-2 shows the most common fiber cable connectors that are used in data center environments.

![Connector types for fiber cables](image)

**ESCON Duplex**  
**MPO connector**  
**MT-RJ**  
**LC Duplex**

*Figure F-2 The connectors that are commonly used for optical cables*

Mode-conditioning patch cables

In certain situations, for reusing existing multimode fiber optic cabling infrastructure, it is possible to connect a long wavelength (1300 nm) single-mode transceiver with multimode fiber by placing a special device called a mode conditioning patch (MCP) cable. The MCP cables are 2 meters long and have a link loss of up to 5.0 dB.

The MCP cable *must* be installed on both ends of a link and occupy the same space as a standard 2-meter jumper cable. Adapter kits containing the MCPs are available with either SC Duplex connectors (to support coupling links) or ESCON connectors (to support ESCON-to-FICON migration). MCP adapters differ for 50 or 62.5 μm fiber. MCPs reduce the maximum link distance to 550 meters for Gigabit links.

Optical mode conditioners are supported for FICON, coupling links, and OSA. For more information, see *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express)*, GA23-1408.

**Important:** One MCP cable must be plugged into the long wavelength transceiver at each end of the link.

Fiber optic MCP cables cannot be ordered as product feature codes for IBM Z.
Fiber optic Mode Conditioning Patch cables (listed in Table F-1) can be ordered through the IBM Networking Services fiber cabling services options.

**Table F-1  MPC cables**

<table>
<thead>
<tr>
<th>MCP cable description</th>
<th>MCP cable connector/receptacle description</th>
<th>MCP cable connector/receptacle illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 µm single mode to 50 µm multimode</td>
<td>SC duplex connector to ESCON duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>9 µm single mode to 50 µm multimode</td>
<td>SC duplex connector to SC duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>9 µm single mode to 62.5 µm multimode</td>
<td>SC duplex connector to SC duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>9 µm single mode to 62.5 µm multimode</td>
<td>SC duplex connector to ESCON duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>ISC-3 compatibility 9 µm single mode to 50 µm multimode</td>
<td>LC duplex connector to SC duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>9 µm single mode to 62.5 µm multimode</td>
<td>LC duplex connector to SC duplex receptacle</td>
<td></td>
</tr>
<tr>
<td>9 µm single mode to 62.5 µm multimode</td>
<td>LC duplex connector to ESCON duplex receptacle</td>
<td></td>
</tr>
</tbody>
</table>
InfiniBand cables

An OM3 50/125 µm (2000 MHz-km @ 850 nm) multimode fiber optic cable with Multifiber Push-On (MPO) connectors is used for 12x IFB-DDR connections.

The cable is an InfiniBand Trade Association (IBTA) industry standard cable. It contains one pair of fibers per lane for the 12x IFB-DDR connection (resulting in 24 duplex fibers).

The sender and receiver connectors are clearly marked with either RX or TX and the connectors are keyed (Figure F-3).

![OM3 50/125 µm multimode fiber cable with MPO connectors](image)

Figure F-3  OM3 50/125 µm multimode fiber cable with MPO connectors

The following standard cable lengths are available from IBM Global Technology Services® (GTS), or cables with individual length can be ordered:

- 10 meters (32.8 feet)
- 13 meters (42.7 feet)
- 15 meters (49.2 feet)
- 20 meters (65.6 feet)
- 40 meters (131.2 feet)
- 80 meters (262.4 feet)
- 120 meters (393.7 feet)
- 150 meters (492.1 feet)
zHyperLink Express and ICA SR cables

The HyperLink Express and ICA SR features are designed to drive distances up to 150 meters and support a link data rate of 8 GBps using customer supplied OM4 (4.7 GHz-Km @ 850 nm) fiber optic cables. With OM3 (2.0 GHz-Km @ 850 nm) fiber optic cables, the zHyperLink Express and ICA SR distance drops to 100 meters. Figure F-4 shows the OM4 fiber cable with 24-fibers (12 transmit plus 12 receive fibers) and Multi-fiber Terminated Push-on (MTP) connectors.

Custom cable lengths or standard cable lengths shown in Table F-2 are available from IBM GTS or through other vendors, such as Anixter Inc.

Table F-2  ICA-SR cables - standard lengths

<table>
<thead>
<tr>
<th>Part Number</th>
<th>Length meters (feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00JA683</td>
<td>1 m (3.28')</td>
</tr>
<tr>
<td>00JA684</td>
<td>2 m (6.56')</td>
</tr>
<tr>
<td>00JA685</td>
<td>3 m (9.84')</td>
</tr>
<tr>
<td>00JA686</td>
<td>5 m (16.40')</td>
</tr>
<tr>
<td>00JA687</td>
<td>8 m (26.24')</td>
</tr>
<tr>
<td>00LU282</td>
<td>10 m (32.80')</td>
</tr>
<tr>
<td>00LU283</td>
<td>13 m (42.65')</td>
</tr>
<tr>
<td>00JA688</td>
<td>15 m (49.21')</td>
</tr>
<tr>
<td>00LU669</td>
<td>20 m (65.61')</td>
</tr>
<tr>
<td>00LU284</td>
<td>40 m (131.23')</td>
</tr>
<tr>
<td>00LU285</td>
<td>80 m (262.36')</td>
</tr>
<tr>
<td>00LU286</td>
<td>120 m (393.78')</td>
</tr>
<tr>
<td>00LU287</td>
<td>150 m (492.12')</td>
</tr>
</tbody>
</table>

For more information, see IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express), GA23-1408. This publication is available in the Library section of Resource Link.
Conversion kits

Conversion kits allow for the reuse of already installed cables that are the same fiber optic mode but have different connectors than the ones required (see Table F-3).

<table>
<thead>
<tr>
<th>Conversion kit cable description</th>
<th>Conversion kit cable connector/receptacle description</th>
<th>Conversion kit cable connector/receptacle illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 µm single mode</td>
<td>LC Duplex Connector to SC Duplex Receptacle</td>
<td><img src="image1.png" alt="Illustration" /></td>
</tr>
<tr>
<td>62.5 µm multimode</td>
<td>MT-RJ Connector to ESCON Duplex Receptacle</td>
<td><img src="image2.png" alt="Illustration" /></td>
</tr>
<tr>
<td>50 µm Multimode</td>
<td>LC Duplex Connector to SC Duplex Receptacle</td>
<td><img src="image3.png" alt="Illustration" /></td>
</tr>
<tr>
<td>62.5 µm Multimode</td>
<td>LC Duplex Connector to SC Duplex Receptacle</td>
<td><img src="image4.png" alt="Illustration" /></td>
</tr>
<tr>
<td>62.5 µm Multimode</td>
<td>LC Duplex Connector to ESCON Duplex Receptacle</td>
<td><img src="image5.png" alt="Illustration" /></td>
</tr>
<tr>
<td>62.5 µm Multimode</td>
<td>LC Duplex Connector to MT-RJ Connector with Coupler</td>
<td><img src="image6.png" alt="Illustration" /></td>
</tr>
<tr>
<td>62.5 µm Multimode</td>
<td>SC Duplex Connector to LC Duplex Connector with Coupler</td>
<td><img src="image7.png" alt="Illustration" /></td>
</tr>
</tbody>
</table>
Conversion kit cable description | Conversion kit cable connector/receptacle description | Conversion kit cable connector/receptacle illustration
--- | --- | ---
9 µm Single Mode | SC Duplex Connector to LC Duplex Connector with Coupler | ![Connector Illustration]

**Note:** Fiber optic conversion kits are not orderable as product feature codes for Z.

Fiber optic conversion kits can be ordered by using the IBM Networking Services fiber cabling services options. Each conversion kit contains one cable.

**References**

The following publications include information that is related to the topics in this appendix:

- *ESCON I/O Interface Physical Layer Document*, SA23-0394
- *Fiber Channel Connection for S/390 I/O Interface Physical Layer*, SA24-7172
- *Coupling Facility Channel I/O Interface Physical Layer*, SA23-0395
- *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express)*, GA23-1408
- *S/390 Fiber Optic Link (ESCON, FICON, Coupling Links and OSA) Maintenance Information*, SY27-2597
- *Fiber Transport Services Direct Attach Planning*, GA22-7234
Related publications

The publications listed in this section are particularly suitable for a more detailed description of the topics covered in this book. There are also “References” sections in some chapters in this book that cite publications that are helpful for the topics of those chapters.

IBM Redbooks publications

For more information about ordering the following publications, see “How to get IBM Redbooks publications” on page 217. Certain documents that are cited here might be available in softcopy only:

- *IBM z14 (M/T 3907) Technical Guide*, SG24-8451
- *IBM z14 ZR1 Technical Guide*, SG24-8651
- *IBM z14 (M/T 3906) Technical Introduction*, SG24-8450
- *IBM z14 ZR1 Technical Introduction*, SG24-8550
- *IBM z13 Technical Guide*, SG24-8251
- *IBM z13 and IBM z13s Technical Introduction*, SG24-8250
- *IBM zEnterprise BC12 Technical Guide*, SG24-8138
- *IBM zEnterprise System Technical Introduction*, SG24-8050
- *IBM zEnterprise 196 Technical Guide*, SG24-7833
- *IBM zEnterprise System Technical Introduction*, SG24-7832
- *IBM System z10 Enterprise Class Technical Guide*, SG24-7516
- *IBM System z10 Enterprise Class Technical Introduction*, SG24-7515
- *IBM System z10 Business Class Technical Overview*, SG24-7632
- *IBM HiperSockets Implementation Guide*, SG24-6816
- *OSA-Express Implementation Guide*, SG24-5948
- *FICON Planning and Implementation Guide*, SG24-6497
- *IBM Communication Controller Migration Guide*, SG24-6298
- *IBM Communication Controller for Linux on System z V1.2.1 Implementation Guide*, SG24-7223
- *Enterprise Extender Implementation Guide*, SG24-7359

Other publications

These publications are also relevant as further information sources:

- *IBM zNext Installation Manual for Physical Planning*, GC28-6965
- *zEnterprise EC12 System Overview*, SA22-1088
• Communications Server: IP Configuration, SC31-8513
• Communications Server: SNA Network Implementation Guide, SC31-8777
• Communications Server: SNA Resource Definition Reference, SC31-8565
• Fiber Optic Link Planning, GA23-0367
• Enterprise Systems Architecture/390 Principles of Operation, SA22-7201
• FICON I/O Interface Physical Layer, SA24-7172
• Fiber Optic Links (ESCON, FICON, Coupling Links and OSA) Maintenance Information, SY27-2597
• Hardware Configuration Definition: User’s Guide, SC28-1848
• Hardware Configuration Definition (HCD) Planning, GA32-0907
• Processor Resource/Systems Manager Planning Guide, SB10-7033
• IBM Z Stand-Alone Input/Output Configuration Program User’s Guide, SB10-7173
• z/Architecture Principles of Operation, SA22-7832
• System z ESCON and FICON Channel-to-channel Reference, SB10-7034
• OSA-Express Customer’s Guide and Reference for z990, SA22-7935

Online resources

These resources are also relevant for further information:

• Fibre Channel standards
  – http://www.t10.org
  – http://www.t11.org

• IBM Z I/O connectivity

• IBM Parallel Sysplex
  http://www.ibm.com/servers/eserver/zseries/pso

• IBM Z networking
  http://www.ibm.com/systems/z/hardware/networking/

• FICON Director vendors

• IBM zEC12: zEC12, HMC V2.12.1, and SE V2.12.1 documentation in the IBM Knowledge Center
  http://ibm.co/1vE4Ju6

• IBM Resource Link for documentation and tools
  http://www.ibm.com/servers/resourcelink
How to get IBM Redbooks publications

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy books or CD-ROMs, at this website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services