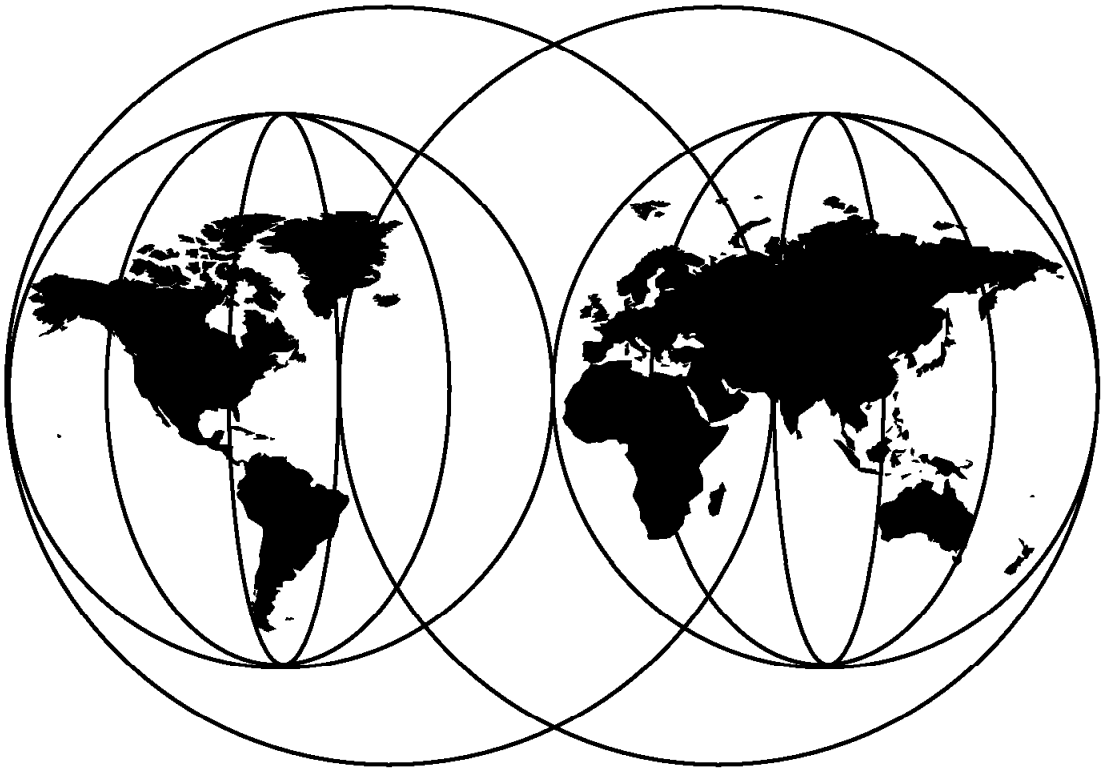
IBM 9729
Optical Wavelength Division Multiplexer

Tim Kearby
Robert Beiderbeck



International Technical Support Organization

http://www.redbooks.ibm.com

SG24-2138-00

International Technical Support Organization

SG24-2138-00

IBM 9729
Optical Wavelength Division Multiplexer

June 1998

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, “Special Notices” on page 109.

First Edition (June 1998)

This edition applies to the IBM 9729 Optical Wavelength Division Multiplexer Models 001 and 041 in a Metropolitan Area Network (MAN) Environment.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. HZ8 Building 678
P.O. Box 12195
Research Triangle Park, NC 27709-2195

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Preface vii
The Team That Wrote This Redbook vii
Comments Welcome viii

Chapter 1. Introduction 1
1.1 Applications of the IBM 9729 2
1.1.1 Mainframe Interconnection 2
1.1.2 LAN Interconnection 3
1.2 Wavelength Division Multiplexing (WDM) Technology 4
1.3 Cost Analysis 5
1.4 Current Models 6
1.4.1 Devices Supported 6

Chapter 2. 9729 Technology and Implementation 9
2.1 IBM 9729 Operation 9
2.1.1 Grating Assembly 11
2.1.2 Laser/Receiver Cards (LRC) 13
2.1.3 Channel Wavelength Allocation 14
2.2 Physical Configuration 15
2.2.1 I/O Cards (IOCs) 16
2.3 Availability Features 17
2.3.1 Dual Fiber Switching Feature 17
2.3.2 Component Redundancy 17
2.3.3 Other High Availability Features 17
2.4 System Engineering Considerations 17
2.4.1 Dispersion 18
2.4.2 Jitter 19
2.4.3 Attenuation 20
2.4.4 Crosstalk and Noise 20
2.4.5 Other Distance Limitations 22
2.5 Network Management 22

Chapter 3. IBM 9729 in a Large System Environment 25
3.1 Parallel Sysplex Overview 26
3.2 Availability Considerations 27
3.2.1 Using Two Pairs to Achieve Continuous Availability 28
3.2.2 Cable Routing Considerations 29
3.2.3 Power Considerations 29
3.2.4 Single Failure Scenario 30
3.3 The 9729 in an ESCON Environment 31
3.3.1 ESCON Operation 31
3.3.2 Using 9729s with ESCON 34
3.3.3 ESCON Configurations 35
3.3.4 ESCON Distance Considerations 38
3.3.5 ESCON Link Recovery 43
3.3.6 Problem Determination on ESCON Links 47
3.4 The 9729 in an External Time Reference (ETR) Network 48
3.4.1 ETR Network 48
3.4.2 9037 Sysplex Timer 50
3.4.3 SysPlex Timer Unit Configurations 52
3.4.4 S/390 Server Sysplex Timer Attachment Features 55

3.4.5 Propagation Delays and the 9037 . . . 57
3.4.6 Extending Existing ETR and CLO Links with 9729s . . . 58
3.4.7 Error Recovery on ETR Links . . . 60
3.4.8 Problem Determination on ETR Links . . . 64
3.5 Using the 9729 with Coupling Facility Links . . . 64
3.5.1 Coupling Facilities . . . 64
3.5.2 Coupling Facility Configurations . . . 68
3.5.3 How Many CF Links Do I Need? . . . 69
3.5.4 9729 CF Link Limitations . . . 71
3.5.5 Effect of Distance on CF Links . . . 71
3.5.6 Extending Existing CF Links Using 9729s . . . 72
3.5.7 Error Recovery on CF Links . . . 77
3.5.8 Problem Determination on CF Links . . . 84
3.6 The 9729 in a Remote Copy Environment . . . 86
3.6.1 Remote Copy Functional Overview . . . 86
3.6.2 Remote Copy Components . . . 87
3.6.3 PPRC Data Flow . . . 89
3.6.4 PPRC Availability Configuration . . . 91
3.6.5 Establishing and Monitoring PPRC Links . . . 94
3.6.6 Link Error Reporting and Recovery . . . 97

Appendix A. Fiber Optic Technology . . . 101
A.1 Optical Fiber Advantages . . . 101
A.2 Optical Fibers . . . 102
A.3 Propagation of Light in an Optical Fiber . . . 103
A.4 Bandwidth . . . 103
A.5 Transmission Modes . . . 103
A.5.1 Single-Mode Fiber . . . 104
A.5.2 Multimode Fiber . . . 104
A.6 Refractive Index Profile . . . 105
A.6.1 Step Index . . . 105
A.6.2 Graded Index . . . 105
A.7 Dispersion . . . 106
A.8 Light Sources and Detectors . . . 106
A.8.1 Transmitters . . . 106
A.8.2 Receivers . . . 106
A.9 Fiber Optic Standards . . . 107
A.9.1 Fiber Channel Standard . . . 107
A.9.2 Synchronous Optical Network (SONET) . . . 107
A.9.3 Integrated Services Digital Network (ISDN) . . . 107
A.9.4 Asynchronous Transfer Mode (ATM) . . . 108
A.9.5 Fiber Distributed Date Interface (FDDI) . . . 108

Appendix B. Special Notices . . . 109

Appendix C. Related Publications . . . 111
C.1 International Technical Support Organization Publications . . . 111
C.2 Redbooks on CD-ROMs . . . 111
C.3 Other Publications . . . 111

How to Get ITSO Redbooks . . . 113
How IBM Employees Can Get ITSO Redbooks . . . 113
How Customers Can Get ITSO Redbooks . . . 114
IBM Redbook Order Form . . . 115

Index 117

ITSO Redbook Evaluation 119

Preface

Nothing compares with optical networking for speed, flexibility, and reliability. But until now, the advantages of optical networking carried a high price: a separate leased optical fiber for each communications link would be required.

Using revolutionary new technology that divides a single leased fiber optic line into as many as 20 separate channels, the IBM 9729 Optical Wavelength Division Multiplexer helps make optical networking affordable by providing the same capacity supplied from a bundle of fibers while paying only for a single strand.

This redbook is a practical guide that shows you how to incorporate the 9729 into your host environment. It shows you how to use the 9729 to interconnect Enterprise System/9000 processors, ESCON directors, and I/O controllers across extended distances up to 50 Km.

A focus of the redbook is the use of the 9729 in the Parallel Sysplex environment. SysPlex Timer links, coupling facility links, ESCON Channel-to-Channel links, and DASD Remote Dual Copy are covered in detail. Applications include remote site backup, running dim or lights-out mode data centers, remote tape archiving and DASD farms, mirroring data from multiple client/server installations, and improving operations across multiple data centers.

The 9729 Multiplexer's pioneering optical wavelength division technology -- the most significant breakthrough in optical networking since the field began more than 10 years ago -- lets you take advantage of the immense, untapped bandwidth in today's fiber optic networks and replaces multiple, high-speed serial links between sites with a single fiber.

The Team That Wrote This Redbook

Tim Kearby is an Advisory ITSO Specialist for Networking at the Systems Management and Networking ITSO Center, Raleigh. He writes redbooks and teaches workshops on local and wide area networking. Tim has held various positions in his IBM career including assignments in product development, systems engineering, and consulting. He holds a Bachelors of Science degree in Electrical Engineering from Purdue University.

Robert Beiderbeck is a Specialist for Large Systems in the IBM Germany Hardware Support Center. He has three years of experience supporting ES/9000s, S/390 CMOS Servers and Enterprise System Connection Networks. Robert has worked for IBM for eight years. His areas of expertise include all the elements of a Parallel Sysplex environment including Coupling Facilities, Sysplex Timers, ESCON Directors and Optical Wavelength Division Multiplexers.

This redbook is the result of work conducted at the International Technical Support Organization (ITSO) Systems Management and Networking Center, Raleigh and the ITSO S/390 Center, Poughkeepsie.

In addition, many people from various organizations across IBM contributed to this project both in terms of time and assistance and content of this book. The authors acknowledge the following people for their invaluable contributions to this project:

Harry Dutton
International Technical Support Organization
IBM Sydney, Australia

Ken Trowell
International Technical Support Organization
IBM Poughkeepsie, NY

Henrik Thorsen
International Technical Support Organization
IBM Poughkeepsie, NY

John Kuras
NHD Development
IBM Research Triangle Park, NC

Noshir Dhondy
S/390 Parallel Center
IBM Poughkeepsie, NY

Jim Murray
S/390 Product Engineering
IBM Poughkeepsie, NY

Mario Borelli
S/390 Testing
IBM Poughkeepsie, NY

Dennis Kekas
NHD Development
IBM Research Triangle Park, NC

Ernie Swanson
NHD Marketing
IBM Research Triangle Park, NC

George Middlebrook
NHD Product Line Management
IBM Research Triangle Park, NC

Dave Petersen
Washington Systems Center
IBM Gaithersburg, MD

Comments Welcome

Your comments are important to us!

We want our redbooks to be as helpful as possible. Please send us your comments about this or other redbooks in one of the following ways:

- Fax the evaluation form found in “ITSO Redbook Evaluation” on page 119 to the fax number shown on the form.
- Use the electronic evaluation form found on the Redbooks Web sites:

  For Internet users http://www.redbooks.ibm.com/

For IBM Intranet users http://w3.itso.ibm.com/

- Send us a note at the following address:

  redbook@us.ibm.com

Chapter 1. Introduction

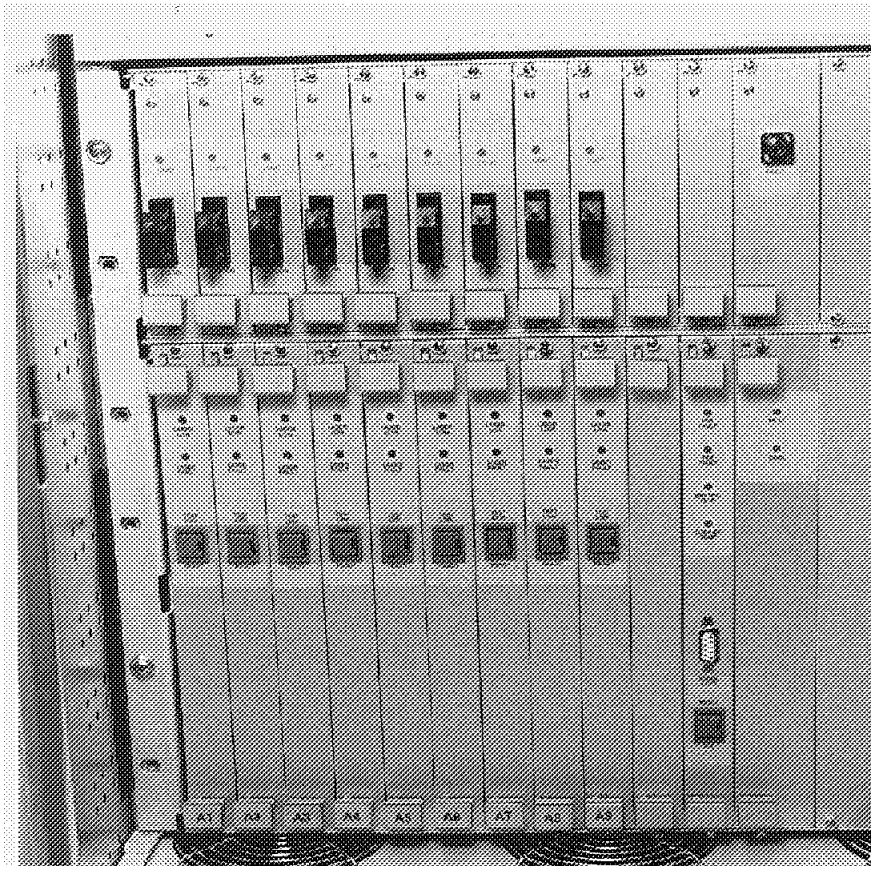


Figure 1. IBM 9729 Optical Wavelength Division Multiplexer

The IBM 9729 is an optical multiplexer for high-speed serial data links. It enables economical transmission of multiple bidirectional bit streams over a single fiber optic line between geographically separate locations. It does this by dividing a single fiber trunk into as many as 20 separate channels. Each channel carries a separate bit stream, with each bit stream possibly using a different communication protocol, bit rate, and frame format. The channels can carry such protocols as ESCON, FDDI, Sysplex Timer, Coupling Links, OC-3, and Fast Ethernet. The two sites can be separated by as much as 50 kilometers.

With the 9729, you avoid the high cost of obtaining additional fiber pairs, thus dramatically reducing your line costs while delivering fast, reliable fiber optic connections between the two sites. In addition, the 9729 has high reliability, very low preventative maintenance requirements, and virtually no limitations to providing new protocol support.

1.1 Applications of the IBM 9729

The primary uses of the IBM 9729 Optical Wavelength Division Multiplexer are:

- Mainframe interconnection
- LAN interconnection

Each of these are discussed in the sections that follow.

1.1.1 Mainframe Interconnection

One of the primary applications of the IBM 9729 is for interconnecting two or more large mainframe computer sites. The 9729 plays a vital role in providing cost-effective communication solutions for enterprises with multiple data centers, especially in today's environment with an increased emphasis on implementing backup/recovery solutions. By coupling computer centers together, corporate users are able to achieve excellent backup and recovery while minimizing their operational costs.

A typical application of the IBM 9729 is shown in Figure 2. This interconnection typically requires several parallel high bandwidth connections. Without the 9729, each of these would require a separate fiber between the sites.

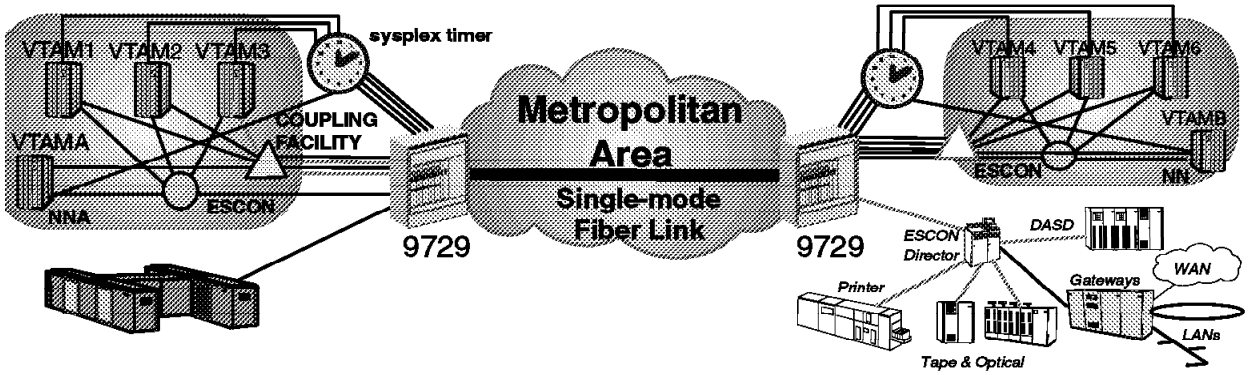


Figure 2. Connecting Multiple Data Centers with the IBM 9729

The 9729 meets a wide spectrum of connectivity needs in the large system environment. It can be used to interconnect S/390 Parallel Enterprise Servers, S/390 Multiprise 2000 Servers, and ES/9000 processors as well as ESCON Directors, Sysplex Timers, and ESCON-capable I/O control units. It can also be used between Sysplex Timers and attached servers and processors. These capabilities make it an extremely useful solution for applications like:

- Remote site backup
- Dim or lights-out mode data centers

- Remote tape archiving and DASD farms
- Improving operations across multiple data centers
- Data center consolidations and reorganizations

1.1.2 LAN Interconnection

LAN-to-LAN traffic continues to increase as more client/server applications are being rolled out and corporate intranets are being implemented. In addition corporate establishments (buildings or campuses) often have other inter-site communication requirements such as a need to interconnect PBX equipment.

LAN-to-LAN connectivity over traditional WAN links using remote bridges or multi-protocol routers has historically been expensive, with the resulting performance still being rather poor. This has even led some organizations to choose not to extend the LAN protocols off the campus at all.

The advent of LAN switching with ATM uplinks and FDDI and SONET-based Metropolitan Area Networks (MANs) have radically improved the performance of interconnected LANs albeit at a still relatively high price for connecting sites separated by any significant distance.

The 9729 can help to make LAN-to-LAN connectivity more cost effective. With the support of FDDI and 155 Mbps OC-3 interfaces, it can be used to meet a variety of connectivity needs. Figure 3 shows some of the possibilities for interconnecting LANs using 9729s.

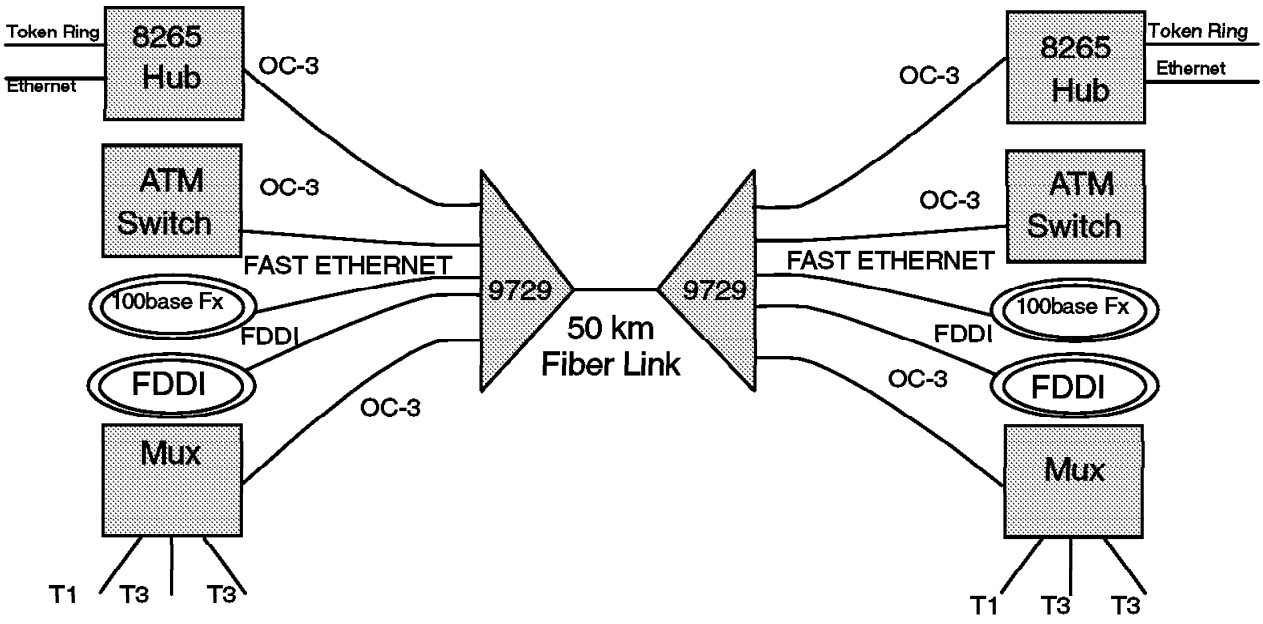


Figure 3. LAN-to-LAN Interconnection Using the IBM 9729

1.2 Wavelength Division Multiplexing (WDM) Technology

Over the past 10 years or so we have witnessed a major transformation in long-haul communication networks. Optical fiber technology has all but replaced copper wire in these networks except for the *last mile* from the end user to the local exchange. This is good: fiber allows the signal to go faster (about 100 times), further (about 20 times), and with fewer errors (about a million times) than copper wire transmission.

However until very recently we have still been using fiber in the most basic possible way: As a better electric wire. That is, a fiber connection consists of a single optical channel on a single strand of fiber connecting two points. All the routing and processing of the signal takes place electronically.

Driven by the ever-increasing demand for more bandwidth and the high cost of installing new fiber strands, engineers have been working hard to make more efficient use of the tremendous bandwidth (many terabits per second)[1] available in a single fiber strand.

There are a number of potential ways to do this. We could increase the speed of a single optical channel on a fiber. However, at a current speed of 2.4 Gbps we have almost reached the practical limit of electronic single channel communication systems. Commercial systems operating at 10 Gbps are available but are problematic in many ways. This is caused both by limitations in the speed of the driving electronics and by dispersion of the optical signal.

We could, if we wanted to, build totally optical networks where the *processing* of the signals in addition to their *transmission* would be handled optically as opposed to electronically. A wide variety of purely optical devices exist to do the job and they do not have the same limitations as do their electronic counterparts. The capacities of such networks would be ten thousand times greater than the best current networks. However, the cost of such a network is still too high to implement in the practical sense.

We could send hyper-fast optical signals (say 100 Giga pulses per second) and use optical time-division multiplexing (OTDM) to separate the optical signal into many (slower) electronic signals. This is the focus of much current research but is a long way from being a commercial technique today.

Another technique called Code Division Multiple Access (CDMA) has been demonstrated in laboratories but is also a long way from being commercial.

The technique of choice for increasing the capacity of an optical fiber is to place many optical signals on the same fiber - each at a different wavelength. This is called Wavelength Division Multiplexing (WDM). The principle is identical with the one we use when we tune a radio to a different station or the TV set to a different channel. In the electronic world we call it Frequency Division Multiplexing (FDM).

FDM and WDM are virtually the same since frequency and wavelength are just two different ways of looking at the same thing. Another way of thinking about it in the optical context is to think of each optical channel as being a different color

[1] One terabit equals one trillion (10**+12) bits.

with each color being carried independently down the fiber. In fact, all the optical channels are in the infrared range of the electromagnetic spectrum so they are invisible, but the principle holds true.

The key to getting greater fiber capacity with WDM is that rather than trying to run a single opto-electronic channel at a very high speed, we run many channels at much slower speeds. This solves many of the electronic (circuit limitations) and optical (dispersion) problems immediately.

In addition, the use of WDM offers a number of significant advantages:

1. Fiber pair gain

   In the Telco voice world, the term *pair gain*: is used to describe analog carrier transmission that increases the utilization of cables by enabling more than one customer to share each physical wire pair.

   In the case of the IBM 9729-001 a single strand of fiber replaces up to 20 separate fibers. This can mean a very large cost saving in rental of fiber.

2. *Protection Switching*

   If you have a large number of fibers connecting the same two endpoints, usually they will be in the same cable. If you want to provide a backup path between the two systems, you need more fibers on the other route. Switching perhaps 100 fibers from one path to another in a failure situation is a major problem! If many channels are carried on the same fiber then it is only necessary to switch a single fiber (or rather, a small number of fibers) from the primary to the backup path. This makes automatic switchover and protection switching very much easier and a lot lower in cost.

3. Protocol Independence

   A major attraction of using WDM to share the fiber rather than other available techniques such as Time Division Multiplexing (TDM) is that each channel can be completely independent of each other channel. That is, the protocols can be totally different. You might have a few channels of ESCON, some ATM, an FDDI connection as well as a few fast Ethernet connections - all sharing the same strand of fiber. Other multiplexing techniques (such as SDH/Sonet) are TDM based and require that all users conform to the same protocol *and* that all connections be synchronized to the same clock! The requirement for clock synchronization is difficult and expensive to meet. With WDM you have no such requirement.

1.3 Cost Analysis

Currently, the cost of renting a single fiber is at least $150-300 per mile per month. Even at $150 per mile per month, a full duplex link over 10 miles costs $36,000 per year. This prices optical networking out of reach for many applications. The 9729 Multiplexer helps make optical networking affordable by providing the same capacity supplied from a bundle of fibers while paying only for a single strand.

For example, using these same cost figures, the cost savings for a 10-channel, full-duplex link over 10 miles would be $342,000 per year. Even a four-channel system would allow you to cut your fiber costs by 75%.

Figure 4 on page 6 shows the cost comparison of using 9729s versus traditional channel extenders. As you can see, the break-even point is only 10 months given the stated assumptions.

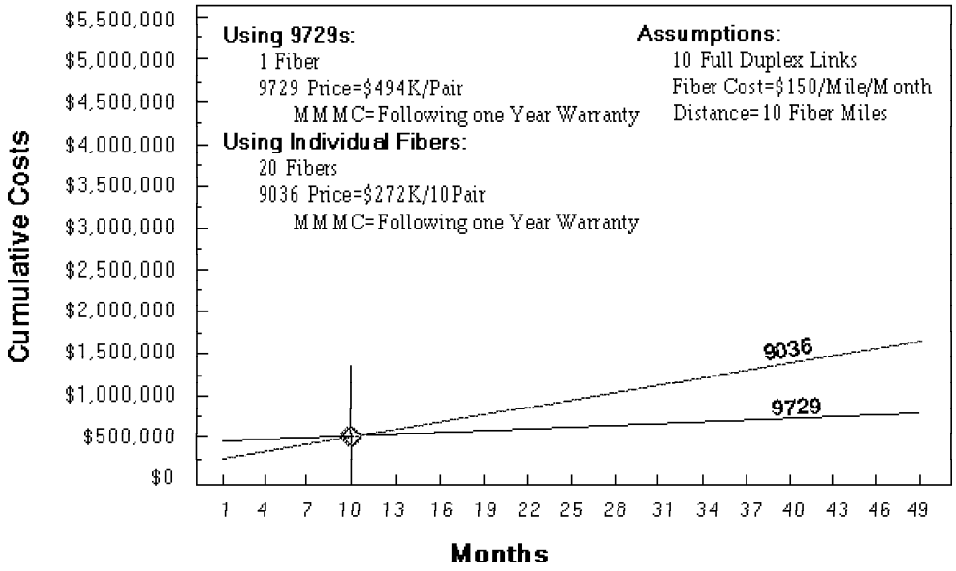


Figure 4. Cost Comparison of 9729s versus 9036s Using Individual Fibers

1.4 Current Models

The 9729 is currently available in two models:

- 9729-001
- 9729-041

The models differ mainly in the number of channels that can be multiplexed into a single fiber trunk. The 9729-041 can multiplex four full-duplex bit streams over a single fiber. The 9729-001 can multiplex 10 full-duplex bit streams over a single fiber. Both models provide channels that can carry signals up to 1 Gbps.

Please Note

The maximum distance between the two locations (the length of the fiber trunk) can be up to 50 km (31 miles). The total end-to-end link distance between the attached devices may be further limited by the attaching device specifications. See 2.4, “System Engineering Considerations” on page 17 and Chapter 3, “IBM 9729 in a Large System Environment” on page 25 for more information on distance limitations.

1.4.1 Devices Supported

Both models support the following fiber I/O interfaces:

- Enterprise Systems Connection (ESCON)

  The 9729 can attach to the following systems and devices using the External Time Reference/ESCON (ETR/ESCON) Input/Output Card (IOC):

  - S/390 Parallel Enterprise Servers (All Models)
  - S/390 Multiprise 2000 Server (All Models)
  - ES/9000 Processors
  - ESCON Directors (All Models)
  - 9037 Sysplex Timer (Model 2)
  - ESCON-capable Control Units

  The ETR/ESCON IOC has the following characteristics:

  - Multimode input and output at 1300nm
  - IBM duplex connector, part # 5605519
  - Supports a maximum of 3 km (1.9 miles) multimode fiber between the 9729 and the attached ESCON device
  - Supports 200 Mbps ESCON, 16 Mbps Sysplex Timer or 155 Mbps OC-3

- Fiber Distributed Data Interface (FDDI)
  - Multimode input and output at 1300nm
  - ANSI FDDI duplex plug: 125 micron fiber
  - 125 Mbps line rate
  - Supports a maximum of 2 km (1.2 miles) distance between the 9729 and the attached FDDI device
- Inter-System Coupling (ISC) Channel

  The 9729 can attach to both single mode ISC and HiPerLinks using the ISC for Coupling Links IOC card.

  The ISC IOC has the following characteristics:

  - Single mode input and output at 1300nm
  - Duplex SC connector
  - 1.062 Gbps line rate
  - Supports a maximum of 3 km (1.9 miles) distance between the 9729 and the attached device
- Synchronous Optical Networks (SONET) Optical Carrier Level 3 (OC-3)

  The ETR/ESCON IOC is also capable of supporting OC-3 signals. This is a physical media attachment only. The 9729 does not provide clocking, jitter removal, or OC-3 framing interpretation. The following characteristics apply to attachment of OC-3 equipment:

  - Multimode input and output at 1300 nm
  - Supports a maximum of 100m (325 feet) multimode fiber between the 9729 ETR/ESCON adapter and the OC-3 device
  - Supports 155 Mbps ATM
- 100 Mbps Ethernet-F

  The FDDI IOC is also capable of supporting Fast Ethernet signals. This is a physical media attachment only. The following characteristics apply to the attachment of Fast Ethernet equipment:

  - Multimode input and output at 1300 nm
  - 62.5/125 micron multimode fiber
  - Simplex ST, FDDI MIC, or Duplex SC connector for ENET-F attachment

- Supports a maximum of 10m (33 feet) multimode fiber between the 9729 and the Fast Ethernet device

Chapter 2. 9729 Technology and Implementation

A conventional optical fiber has an accessible bandwidth of 25,000 GHz. Wavelength Division Multiplexing (WDM) is an approach to opening up as much of this bandwidth as possible by breaking it up into many channels, each at a different optical wavelength (in effect, a different color of light).

The bandwidth of each optical channel is still much larger than the capability of the electronic-based attaching equipment. For example, 1 nm of bandwidth at 1550 nm wavelength equals approximately 125 Gigahertz.[2] The maximum bandwidth that even the most advanced electronic equipment can handle today is only a few gigabits per second.

The 9729 Model 001 uses 20 optical wavelengths in the 1540 to 1559 nm wavelength range. The 9729 Model 041 uses eight optical wavelengths in the 1540 to 1547 nm wavelength range. In each case, the channels are spaced about 1 nm apart. Half of the channels are used for transmission in one direction and half in the opposite direction on the same fiber. Each channel is received as a separate I/O signal, then multiplexed with the others, and then sent out on the fiber trunk. At the receiver end the channels have to be split out from one another (de-multiplexed) and directed to separate receivers.

The 9729 Multiplexer does not touch the bits, and they pass through the unit as though it consisted of individual fiber paths.

2.1 IBM 9729 Operation

The 9729 system structure is shown in Figure 5.

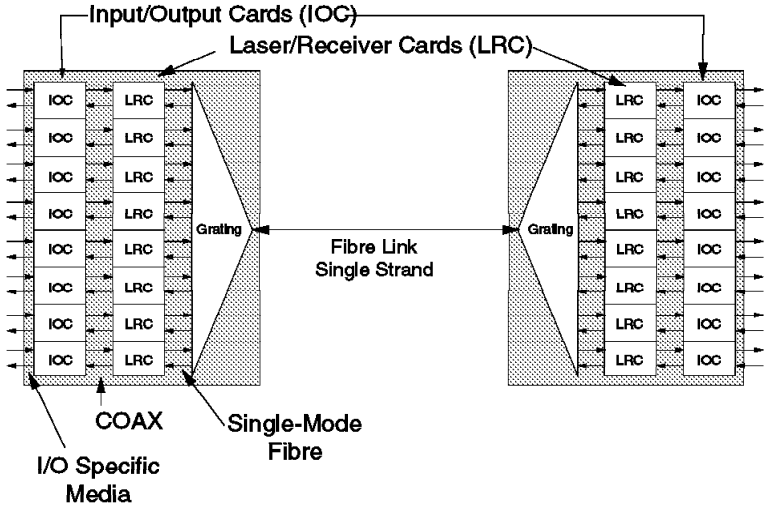


Figure 5. IBM 9729 System Structure

[2] At one bit per baud, this equates to 125 Gbps.

There are two types of plug-in modules: *input/output cards* (IOCs) whose pin connections and voltages are specific to the data stream type, and wavelength-specific *laser/receiver cards* (LRCs). Each full-duplex link requires an LRC and an IOC at each end inside the 9729 Optical WDM units.

For example, in order to provide a full-duplex link for an Enterprise Systems Connection (ESCON) channel, you would use an ESCON-specific IOC. This card provides a plug-compatible ESCON interface that receives the multimode ESCON signal, converts it into an electronic bit stream, and sends it to the corresponding LRC to be retransmitted on a unique optical wavelength. At the other end, this unique wavelength is directed to an LRC by the grating, where it is received, converted to electronics and transferred to the corresponding IOC card. The bit stream is then transmitted out as a multimode ESCON signal again.

The same process happens in the other direction as well.

The attached ESCON devices do not see the 9729 Optical WDM; it is as if they were provided with two independent fiber cables.

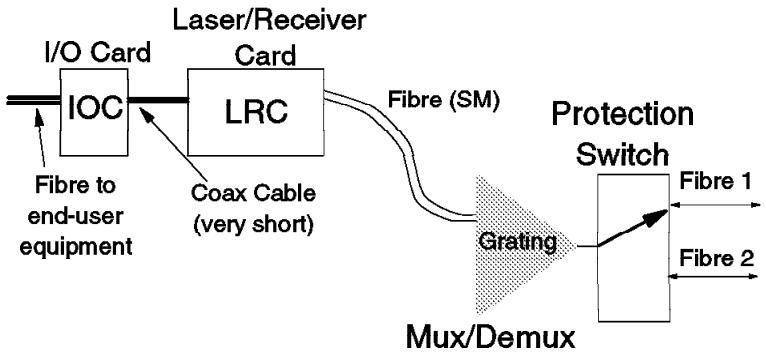


Figure 6. IBM 9729 Major Components and Data Flow

The major components of the IBM 9729 are shown in Figure 6. Overall operation is as follows:

- An optical signal from the end-user equipment is received and converted to electronic form on an IOC adapter card.

  There are a number of different IOCs that apply to different end-user protocols. (See 2.2.1, “I/O Cards (IOCs)” on page 16 for the different IOCs available.)

- The signal is sent from the IOC over a pair of very short coaxial cables to the LRC. Each LRC card in a system operates at a different wavelength. The laser has been selected and tuned to operate on one and only one specific channel. The received electronic square wave is used directly to modulate the drive current of the transmitting laser.

- The output from the transmit laser is connected on a short single mode fiber to a specific port on the grating multiplexer. The grating is used bi-directionally both to mix the output of the transmit lasers before sending on the single fiber and to split the wavelengths up and direct them to different receivers when receiving.

- When the signal is received by the partner 9729, the grating assembly in that unit receives the mixed signal and directs each channel to a different outgoing fiber depending on the signal wavelength.
- The signal is then received on the LRC card and converted to an electronic signal. This signal is then re-shaped (made into a square wave) just as happened in the transmit direction.

  The signal is then passed from the LRC card to the IOC card and it is used to control the transmitter (laser or LED) for forwarding on to the attaching device.

2.1.1 Grating Assembly

The grating assembly is the heart of the IBM 9729 system. It performs the multiplexing of outgoing signals and the de-multiplexing of incoming ones as shown in Figure 7.

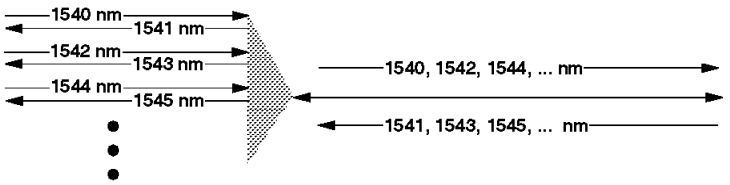


Figure 7. Grating Function

In physical construction it is a Littrow reflective grating combined with a concave mirror. It is a totally passive element.

A schematic of the grating is illustrated in Figure 8 on page 12. As light reaches the end of one of the input fibers (the signals labeled as λ_1 through λ_4), it does not do what we would like it to do, which is just continue in a straight line. Instead, it scatters out in all directions.

The concave mirror is used to focus it. The mirror focuses the light arriving on the end of each of the fibers and re-directs it to a point on the grating.

The grating *bends* the signal according to its wavelength, such that signals come off the grating headed for the same point on the mirror. This works using the same refraction properties as a prism.

The light comes off the grating and again scatters. The mirror is used again to re-focus what is now the combined signal and directs it out onto the fiber trunk (labeled as $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$).

The process works exactly the same in the opposite direction. Thus we are able to achieve full-duplex channels.

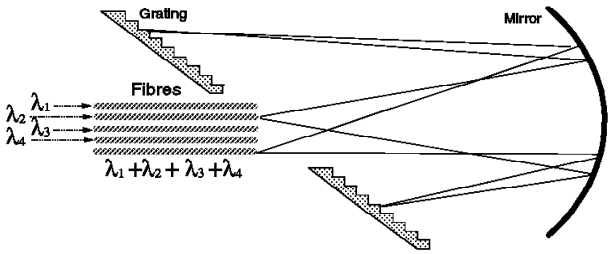


Figure 8. Grating Assembly Schematic

Note: While in the 9729 it is used in bi-directional mode, there is nothing at all that restricts the directionality at the grating level. There would be nothing to prevent the grating being used to transmit 20 channels simultaneously or to receive 20. Provided it's at the correct wavelength (different for each fiber) a signal arriving on the fiber will be merged onto the common input/output fiber.

The grating is able to select an individual channel from among all other received channels with crosstalk and noise (picked up from the other channels) of less than -30 dB.

The reason a grating is used to mix the outgoing laser signals is to minimize the amount of optical loss. If you used a fused fiber coupler or a star coupler, you would lose so much light an amplifier would be needed to boost the signal again before transmission.

The nominal insertion loss of the grating including its connectors is 6 dB in each direction.[3] Thus a signal originating on one LRC in one 9729 and received in another 9729 will suffer a maximum of 12 dB of attenuation due to the gratings.

A very important part of the grating function is the temperature control system. While the grating doesn't produce any heat itself, it expands and contracts with variations in ambient temperature. This causes a shift in the center wavelength of the optical channel.[4] For this reason, the grating in each 9729 Optical WDM is housed in a temperature-controlled compartment. The temperature within the compartment is maintained by a grating controller card that plugs into a card slot.

The IBM 9729 grating is operated at a temperature of 29° C. The grating controller card has six LEDs that indicate:

HEAT The grating is being heated.

COOL The grating is being cooled.

OPEN One or more of the three thermal units are open.

SHORT One or more of the three thermal units are shorted.

TEMP The grating temperature is out of range.

FAN The grating fan is not operating.

---

[3] The measured figure is somewhat better than this but the 6 dB is the design assumption based on worst-case conditions.

[4] The center wavelength of each channel changes by .01 nm per degree C of change in the grating temperature.

2.1.2 Laser/Receiver Cards (LRC)

The LRC cards are the transceivers that send and receive optical signals on the common trunk. They are completely protocol independent. They have both a transmitter and a receiver on board.

2.1.2.1 The Transmitters

The 9729 uses a very high precision Distributed Feedback (DFB) laser for the transmitters in the LRC cards. These lasers have a narrow *linewidth*.[5] The nominal linewidth of these lasers is .3 nm at 20 dB below the peak of the pulse. They are built with a grating in (really adjacent to) the cavity to stabilize the wavelength and narrow the linewidth and reduce the *chirp*.[6]

The 9729 employs an additional measure that almost completely eliminates chirp. The lasers are operated so that a logical zero level is still *above* the lasing threshold. Thus the laser is transmitting a low-level signal even in the zero state. This has the additional benefit of reducing the linewidth thus reducing dispersion.

All laser output has a tendency to drift over time. This occurs as the laser ages and the wavelengths produced changes slightly. The transmitter inside the LRC must not be allowed to drift outside its allocated band. In the 9729, the lasers can be tuned over a range of about 1 nm. This is used to fine-tune the lasers to the exact channel wavelength. This procedure needs to be performed about once a year.

The ability to tune the lasers also solves a cost problem in the manufacturing process of the lasers. It is very difficult and expensive to make lasers for exactly the wavelength you want. Instead, manufacturers produce a large number of lasers and then measure the wavelength of each one, choosing the ones that happen to fit what they need! Electronic tuning relaxes the demands on manufacturing tolerances and increases the usable yield of lasers.

Laser output can also drift with temperature changes. The lasers in the 9729 have their own individual temperature control that effectively eliminates drift due to temperature changes.

The lasers also have their own feedback control of power level.

2.1.2.2 Receiver

Because we are operating in the 1550 nm band, we need a receiver that can detect the relatively low energy of these signals. The LRC card uses an Indium Phosphide Avalanche PhotoDiode (APD).

These detectors are sensitive to signals at any and all wavelengths used by the system. They have no way of distinguishing between signals by wavelength. Thus it is up to the grating to discriminate and select an individual signal and route it to the correct receiver.

---

[5] Contrary to popular belief, lasers don't produce a single wavelength signal. Rather they produce a band of wavelengths referred to as the linewidth of the laser. Since each channel is a wavelength band within which the signal must stay if the system is to work, the laser used to transmit the individual channel must have a linewidth that fits comfortably within the channel.

[6] Chirping is the undesirable transient shifts in wavelengths produced from the laser when it is turned on (begins to transmit a "1" bit). It is due to both the change in carrier concentrations within the laser cavity and to temperature change.

The received signal is re-shaped before sending to the IOC card but it is not re-clocked. (See 2.4.2, “Jitter” on page 19 for the implications of jitter.)

2.1.3 Channel Wavelength Allocation

Figure 9 shows the channel allocation in the 9729 Model 001.

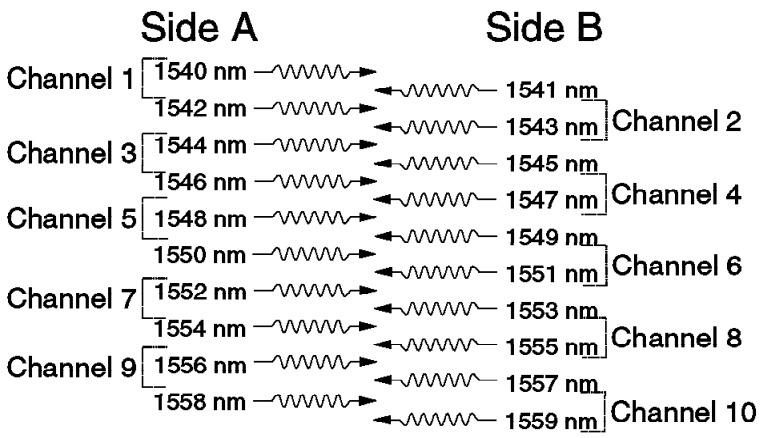


Figure 9. IBM 9729 Wavelength Allocation

As can be seen in the figure, the channels are spaced 1 nm apart. At a wavelength 1550 nm, this inter-channel spacing is equivalent to approximately 125 Gbps of bandwidth. (In this application, one baud equals one bit.)

Considering that the maximum bit rate that the 9729 is designed to handle on any one channel is 1 Gbps, this seems very conservative. However, as discussed above, we have to allow room for component tolerances, temperature variations, and wavelength drift with age.

As also can be seen in the figure, the signal direction is alternated so that signals traveling in the same direction are separated by 2 nm.

2.1.3.1 Other LRC Characteristics

There is only one type of LRC card and it is used for all protocols. LRCs are completely protocol independent. However, since each channel operates at a specific wavelength, the LRCs must be manufactured to the center wavelength of the channel in which they will be operating.

Thus, for example, in an installation using 9729 Model 001s, there are 20 different LRC cards representing the 20 different wavelengths used. They are shipped in matched pairs as an A side and a B side and each is identified as to which channel on the multiplexer it is to be used. For example, LRC A1 will be used on the A side in channel 1 with a corresponding LRC B1 on the B side.

2.1.3.2 Laser Safety

The IBM 9729 is designed as a class 1 laser product. This means that the total power transmitted from all lasers in the box is less than the class 1 safety limits as specified by the relevant regulations. Thus, the unit has been designated as inherently safe. This means that no special provisions need to be made for the fact that lasers are used.

The 9729 employs additional measures to increase the level of safety. The LRC cards are controlled from a central diagnostic and control processor card. At one-second intervals each LRC card is checked to determine if light is being received. If there is no light at the receiver then the LRC turns off its laser. This version of the Open Fiber Control system is used so that no laser is allowed to transmit for very long unless a signal is being received from the other direction.

Intermittently, when light is not being received, the LRC sends a pulse of light to the LRC at the other end. This is a signal for the other LRC to turn on its laser.

These features are important as the product is designed to be operated by people without special training in fiber optics. In contrast to equipment operated by telephone companies that is operated and installed by engineers and trained technicians, the 9729 can be installed in both a computer room or a regular office environment.

2.2 Physical Configuration

Each 9729 Optical WDM unit is housed in an enclosure that contains the grating assembly, IOCs, LRCs, the diagnostic card, and the grating controller card, as well as the power supplies and fans. Figure 10 shows the front panel of the 9729.

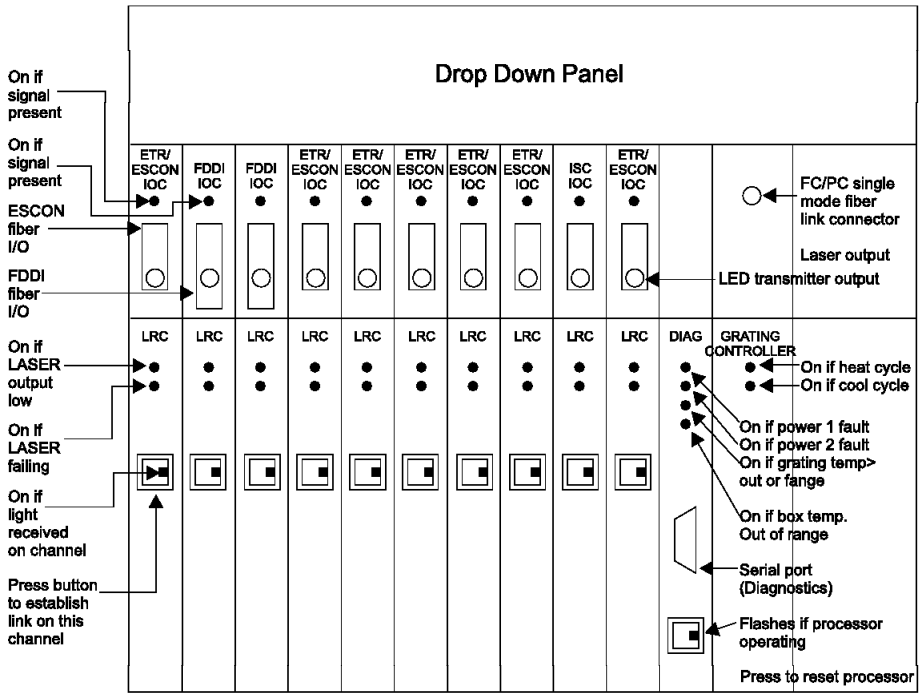


Figure 10. 9729 Optical WDM Components

The IOCs, LRCs, the diagnostic card, and the grating controller card are all housed in a standard 19-inch rack-mountable card cage which measures 15.5 in. (394 mm) high and 11.0 in. (279 mm) deep. The 9729 Optical WDM has two rows of up to 10 pluggable modules on the front panel. The lower row contains the LRCs and the upper row contains the IOCs.

The LRCs are 6U (10.5 in. [267 mm]) high. The IOCs are 3U (5.25 in. [133 mm]) high. Both plug into the backplane from the front.

Every board has a front panel including LED status indicators. Figure 10 shows the status indicators and gives an overview of their function.

The 12th slot contains the single- or dual-fiber connector in the upper row and the grating temperature controller board in the lower row. The eleventh slot in the lower row contains the Diagnostics board.

The grating in each 9729 Optical WDM is housed in a temperature-controlled compartment mounted on the card cage above the IOC boards. The fibers between the individual channels and the grating are routed in back of the card cage to the backplane. The common fiber is routed to the backplane slot that contains either the single or dual fiber I/O card. The fiber I/O card routes the common fiber to the appropriate I/O connector.

The unit can be configured with a minimum of one full-duplex channel and a maximum of 10 (9729-001). Each channel requires a pair of IOC and LRC cards at each end of the link. An installed link consists of two 9729 Optical WDM units A and B, one at each end of the link. To add channels, simply plug in pairs of IOCs and LRCs.

2.2.1 I/O Cards (IOCs)

There are currently three different types of IOC cards. These are:

- FDDI Card

  In addition to the Fiber Distributed Data Interface (FDDI), this adapter can also be used for 100 Mbps Fast Ethernet as these interfaces use the same optical specification as FDDI (namely an LED transmitter in the 1280-1320 nm wavelength band over multimode fiber).

  Processing performed by the FDDI card is very simple indeed. The received optical signal is re-shaped (made into a square pulse again) and converted into electronic form. However, there is *no* clock recovery or re-timing of the signal. This is an important feature as it allows the system to handle any protocol using compatible wavelengths and power levels and does not exceed the speed handling capabilities of the IOC′s electronics.

- ESCON Card

  This card may be used for ESCON, 9037 Sysplex-Timer Model 002 External Time Reference (ETR) and Control Link Oscillator (CLO) links as well as 155 Mbps OC-3 signals such as SONET/SDH and 155 Mbps ATM. The attachment interface uses an LED transceiver in the 1300 nm band over multimode fiber.

- Inter-System Coupling (ISC) Card

  This card may only be used for Parallel Sysplex Coupling Facility links (includes S/390 High Performance Coupling Links (HiPerLinks). It uses the IBM ISC 1.0625 Gbps interface, which specifies a laser transmitter at 1310 nm over single-mode fiber.

  This IOC is unique in that it re-shapes *and re-clocks* the input signal before passing it on to the LRC card. This removes any jitter in the signal. However, the frame structure is not inspected and the code structure is not decoded. The optical signal is unchanged except for the re-timing to remove the jitter.

  The card also uses the Open Fiber Control (OFC) safety interlock as described in 2.1.3.2, “Laser Safety” on page 14.

2.3 Availability Features

As the 9729 is often employed in mission-critical environments, the reliability and availability features of the product are very important. This section highlights several of these features.

2.3.1 Dual Fiber Switching Feature

As an option the user may choose to utilize two independent fiber trunks between the IBM 9729s that have different physical routings. This creates a backup path in case the main path is cut.

The dual fiber switching feature is a physical switch on the trunk side of the grating. It is electronically activated to switch from one trunk fiber to another when the operational trunk fails. If the active trunk fails for any reason (for example the cable is cut) then both 9729s (at each end) automatically switch to the backup path.

It works as follows: at one-second intervals the control module monitors the LRC cards to determine if any are receiving light. If none of the LRCs is detecting light, then the unit switches to the backup fiber. Once switched to the backup fiber the lasers are switched on in sequence over a period of about one second for safety control reasons. This switching can be completed in less than two seconds and hence the higher-layer protocols can recover from the failure transparently.

If light is not detected on the backup fiber then a switch is made back to the primary fiber.

Note: The 9729 maintains a protocol that ensures both sides of the link will synchronize with each other quickly. This prevents the units from switching back and forth from one fiber to the other even if both were operational.

The system also has a manual switchover that works via SNMP so that the customer can initiate a switchover or a switchback.

2.3.2 Component Redundancy

Power supplies and cooling are duplexed such that the system will operate normally in the case of a failure of a power supply and/or a cooling fan.

2.3.3 Other High Availability Features

All cards (IOC, LRC, diagnostic card and grating controller) are hot-pluggable and hence may be inserted and removed while the 9729 is operating.

The lasers used are telecom grade with mean time between failures (MTBFs) of over 10 years.

2.4 System Engineering Considerations

There are many important aspects to the design of an operational WDM system. A few of the more important issues are:

1. Dispersion
2. Jitter

3. Attenuation
4. Crosstalk and Noise

These all contribute to the signal quality available at the receiver and hence to the maximum possible link distance. Because of this important connection, each is given special focus in the sections below.

2.4.1 Dispersion

Dispersion is the distortion of a pulse of light as it travels down an optical fiber. As can be seen in Figure 11, a *square* pulse will look quite different at the receiver than it did when it left the laser transmitter at the other end of the fiber. Dispersion of the pulse limits the distance of the link because the receiver must be able to accurately recover the signal.

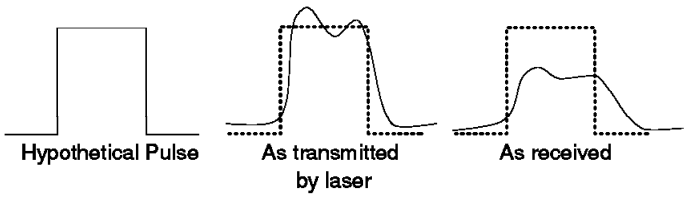


Figure 11. Distortion of an Optical Pulse

The amount of dispersion that a system can tolerate is a function of both the speed and the distance of the link. For example, at 1 Gbps we can afford a maximum dispersion of 200 ps. This is 20% of the 1 ns pulse. Using standard fiber, we can get to around 50 km before dispersion reaches this level. However, if we slow the link to 200 Mbps, we can go approximately five times this distance or about 250 km!

Note: These figures are very rough guidelines only. Each proposed configuration needs to be tested if it in any way approaches these limits.

While the laws of physics tell us that we will encounter dispersion, there are techniques that can be used to minimize its effect in an optical communication system. Some of these are:

- Dispersion compensating fiber

  Standard fiber has a dispersion of 17 picoseconds per nanometer per kilometer (ps.nm.km.).[7] You can install specially designed fiber called dispersion compensating or *shifted* fiber that has significantly less dispersion than standard fiber but just enough to minimize crosstalk. Shifted fiber has about 4 ps.nm.km dispersion, which allows the signal to travel about four times farther.

  Note: You can also get fiber that has zero dispersion. However, this is not recommended in WDM systems because of the possibility of interference (crosstalk) between optical channels. This is much more appropriate in single channel communication systems.

  Of course, dispersion compensating fiber only applies if you are laying new fiber. Most installed fiber is standard (non-shifted).

[7] Dispersion is a function of the wavelength. This figure is based on wavelengths in the 1540 to 1560 range, which is the range used by the 9729.

The 9729 is designed for distances of less than 100 km and link speeds less than 1 Gbps. You can use standard fiber without dispersion compensation. For speeds greater than 1 Gbps or distances above about 50 km, you will probably need compensation.

- Narrow linewidth lasers

  Dispersion can be significantly reduced by using lasers with a narrower linewidth and external modulators. The nominal linewidth of the lasers used in the 9729 is .3 nm at 20 dB below the peak of the pulse. This is narrowed some by operating the laser just above threshold and therefore minimizing the effect of chirp. Further, the gratings function as wavelength selective filters and tend to narrow the linewidth.

- Chirped In-Fiber Bragg Gratings

  This is a very new device still undergoing field trials. The principle is that a wavelength-selective mirror is written into the core of a fiber. The wavelength at which the mirror reflects is changed uniformly down the fiber.

  When the dispersed signal arrives, it is directed into the mirrored section of fiber. Since a dispersed pulse has its high frequency (short wavelength) components first and the low frequency ones last, the grating allows the short wavelengths to travel further than the longer ones before it is reflected.

  The grating length is arranged such that the dispersion from the travel down the fiber is exactly balanced by dispersion in the opposite direction within the grating.

  You have to do special things to direct the signal into the mirrored section of fiber and get it back after it is reflected. But the principle works very well and is extremely low in cost.

- Signal re-shaping

  The 9729 re-shapes each received signal into an electronic square wave. This is done by coupling the APD output through an a/c transformer and through filtering of the signal. This re-shaping is done both at the IOC interface as the signal comes into the 9729 and also at the LRC at the other end of the trunk. By doing this, the effects of dispersion are limited to the individual fiber segment and generally do not affect the total end-to-end distance limitations of the connection.

2.4.2 Jitter

Jitter is the shift in the timing in the received bit stream due to the uncertainty of the receiver as to when a bit transition has taken place. The amount of jitter increases with distance due primarily to the effects of dispersion. If bad enough, the receiver can drop bits and even completely lose the signal.

Many protocols have strict jitter requirements. For example, SONET/SDH protocol has very strict jitter requirements that are hard to meet. Other protocols, such as 100 Mbps ATM, have much less stringent jitter requirements.

The effects of jitter can be cumulative over successive links. This can impact the overall system design of an IBM 9729 deployment. For example, an FDDI device may be located up to 2 km away from the 9729. An FDDI pulse arriving at the 9729 IOC after 2 km of travel over a multimode fiber will be significantly distorted. The receiver in the IOC will re-shape the signal, but it will not re-clock it. Thus there is jitter in the regenerated square wave pulse as it transits the 9729. When received at the destination 9729, the dispersion of the signal over

the trunk will create more jitter, which is added to jitter already in the signal from the last hop. If the attaching device is another 2 km from the destination 9729, even more jitter is added to the signal.

Thus, while the 9729 removes much of the effect of dispersion and distortion from the pulse during its travel, it cannot remove the jitter without re-clocking the signal. Unless the signal timing is regenerated when the pulse is reshaped, jitter is not removed.

Therefore before using 9729s with attaching equipment at a significant distance (more than 100 meters) away, the user should evaluate whether the introduced jitter will be within acceptable limits.

2.4.3 Attenuation

Attenuation is the reduction in intensity of the light pulse as it travels down the optical fiber. It limits the maximum link distance for the same reason as dispersion. The receiver has to be able to detect the incoming signal.

The term *link budget* is used to specify how much loss in signal strength can be tolerated and still allow the receiver to interpret an accurate signal. It is simply the difference between transmitter power and receiver sensitivity adjusted to allow for other components that introduce attenuation such as the gratings and the connectors with a margin for component tolerances and aging.

The IBM 9729 has a link budget of 15 dB at 200 Mbps.[8] With respect to the link budget, the maximum distance is really the length of fiber you can get to with no more than 15 dB attenuation. Since modern fiber has a nominal attenuation of .24 dB/km[9] in the 1550 nm band, this equates to about 62.5 km with a 15 dB link budget.

Note: It is possible to buy very special fiber with an attenuation of only .18 dB/km. However, the 9729 specifications are based on standard fiber.

2.4.4 Crosstalk and Noise

In WDM systems, crosstalk (interference between channels) and noise are critical issues. There are many effects that can contribute to the generation of crosstalk and noise and in order for the system to operate correctly these effects need to be considered in the basic system design.

The major cause of crosstalk is called *Four-Wave Mixing* (FWM). This is mixing of signals to produce sum and difference signals. There are two problems here. One, the effect removes power from the signal channels in a random way, thus creating noise. Second, if the newly created harmonic signals coincide in wavelength with other signal channels, this adds noise to those channels.

The 9729 has the following design characteristics that help minimize crosstalk and noise problems:

Maximum Power Level

Light in fiber exhibits many strange phenomena especially at higher power levels. These effects cause the light to behave in a way that the

---

8 Receiver sensitivity is a function of line speed and reduces by about 3 dB whenever you double the line speed.

9 Installed, a good rule of thumb is .26 dB/km.

signal is distorted so badly that it cannot be recovered at the other end of the link. Two such phenomenon are Stimulated Brillouin Scattering (SBS) and Stimulated Raman Scattering (SRS).

In the 9729, the maximum power level employed both for individual optical channels and for the total is much lower than required to produce most of these non-linear effects. Each channel is transmitted at a maximum power level of -9 dBm and the total of 10 channels can never exceed 0 dBm. SBS and SRS effects require much higher power levels to become a problem.

Wavelength Selection

The 9729 wavelengths are spaced at intervals of 1 nm (2 nm for channels travelling in the same direction). When you use channel spacings expressed as a constant in nanometers, the generated harmonics if any fall outside the center wavelengths of the other channels.[10] This prevents Four Wave Mixing (FWM) which is another non-linear effect that occurs at the sum and difference frequencies of the channels involved. This mitigates some of the effects of Four Wave Mixing (FWM).

Use of Standard Fiber

The 9729 system design limits are based on the assumption that standard, or *non dispersion-shifted* fiber will be used. Standard fiber is recommended. Operation of the 9729 over fiber with a dispersion minimum at 1550 n (called dispersion-shifted fiber) is not recommended by IBM.

FWM is the major potential problem in WDM systems. In standard (unshifted) fiber the signal channels in the 1540-1560 nm band travel at different speeds. Thus the signals do not stay in phase for long enough for FWM effects to build up.

Of course the use of unshifted fiber brings with it an increased concern about dispersion. However, the 9729 has been designed to operate at its announced speeds at its supported distances in the presence of normal dispersion from standard fiber.

Of course in practical terms most installed fiber is unshifted and cannot be economically changed. So we need to use it for practical reasons anyhow.

If you are installing new fiber, there is a special fiber available which has enough dispersion to prevent FWM but much less dispersion than traditional standard fiber. This fiber is designed for WDM systems. If you have the opportunity, install this type of fiber. Most major fiber manufacturers have their own versions of this fiber but the AT&T version is called Tru-Wave fiber.

Use of the grating de-multiplexer

The particular type of Littrow grating de-multiplexer used is extremely high in cost but offers a very high selectivity between channels. Crosstalk introduced by the grating is less than 30 dB.

---

[10] This is because a constant spacing in nanometers implies a *different* spacing in frequency terms. For example the 1 nm gap between the wavelengths of 1550 nm and 1551 nm represents a frequency bandwidth of 124.789 Ghz. The same 1 nm gap between the wavelengths 1560 nm and 1561 nm equals 123.195 Ghz.

2.4.5 Other Distance Limitations

The preceding section discussed several important factors of distance limits imposed by the physical layer. There are many other factors introduced by higher layer protocols that can further limit the maximum link distance.

An example is propagation delay. Light in an optical fiber travels at about 2/3rds of the speed as it does in a vacuum.[11] This equates to about 5 microseconds per kilometer of travel. While this seems very fast, it can provide a significant limitation for some devices and systems. For example:

- Device timing constraints limit the distance on some ESCON devices to distances significantly less than the capability of the IBM 9729. (See Chapter 3, “IBM 9729 in a Large System Environment” on page 25 for more information on distance limitations in the ESCON environment.)
- Ethernet-100 can be used as a point-to-point protocol connecting LAN switches or routers. In this application use of Ethernet-100 over a long link with WDM can be highly practical and cost-effective. However, if an Ethernet-100 connection is used as a part of an Ethernet collision domain (that is, if you operate the connection as part of a regular Ethernet LAN) then the maximum distance between any two devices in that domain is 200 meters!

The important point here is that the maximum distance for any particular proposed application *could* be significantly less than the distance allowed by the IBM 9729.

2.5 Network Management

A system overview of IBM 9729 network management is shown in Figure 12.

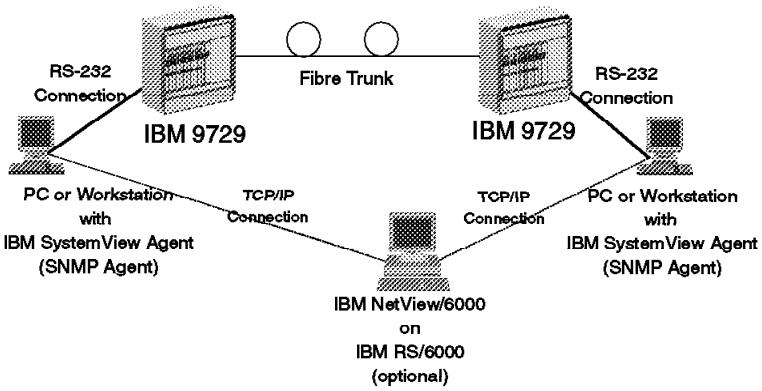


Figure 12. IBM 9729 Network Management Configuration

Inside each 9729 unit, there is a diagnostic card that monitors many aspects of the internal functioning of the 9729. The card’s on-board processor maintains information about status and accepts some control commands from the management system through Simple Network Management Protocol (SNMP) Management Information Base (MIB) entries.

[11] The speed of light in a vacuum is 3.0 x 10**8 meters per second.

A PC or workstation directly attached to the diagnostic card via an RS-232 interface runs SNMP agent and subagent software. This workstation can monitor up to five 9729s in one location. The subagent software interacts with the diagnostic card to gather status information and to relay control commands.

The SNMP agent software is a TCP/IP application that interrogates the subagent and interfaces with the end user via a screen and keyboard.

Optionally, all the IBM 9729s in an organization may be managed together with the rest of the organization's network equipment through a centrally located workstation running IBM NetView/6000 software.

2.5.1.1 Hardware and Software Requirements

The workstation that executes the SNMP subagent must include the following components:

- One or more EIA 232 ports to attach to the 9729 to be managed
- An Ethernet or token-ring network adapter card for communication between the subagent and the Network Management Station
- At least 8 MB of memory
- Available hard disk space as follows:

  For OS/2 - 140 KB
  For AIX - 170 KB
  For Windows 95 - 240 KB
  For Windows NT - 240 KB

The NetView for AIX application requires:

- An IBM RS/6000 running AIX V3.2.5. For the purpose of installing NetView for AIX V3, the RS/6000 should have the following:

  A minimum of 150 MB or hard disk space that will be associated with directory /usr/OV

  A minimum of 64 MB of memory

  A minimum of 200 MB of swap/paging space

  A tape drive, CD-ROM, or installable file image on a central server for installing NetView for AIX

  AIXwindows Environment/6000

  X Window System Version 11 Release 5

  X11 fonts: X11fnt.ibm850.pc.fnt, X11fnt.core.X.fnt

  OSF/Motif Version 1 Release 2

  SNMP Agent

  TCP/IP

- NetView for AIX V3 installed on the IBM RS/6000.
- RS/6000 networking capabilities as well as a network route to the LAN on which the PC or laptop running the SNMP-DPI subagent is located.
- Extra disk space of 3.5 MB in /usr/OV to install the 9729 NetView applications. Root permissions will be required to install the 9729 NetView applications.

Chapter 3. IBM 9729 in a Large System Environment

A primary application of the IBM 9729 is for interconnecting two or more large mainframe computer sites. For example, this might be a large mainframe complex and a backup site. By coupling computer centers together in this manner, corporations are able to achieve excellent backup and recovery capabilities while minimizing their operational costs.

Figure 13 shows a typical configuration for connecting two mainframe data centers. The interconnection of mainframes in a configuration such as this requires several parallel high-bandwidth connections that are usually optical in today's environment. The connections are either dedicated or switched via IBM 9032 ESCON Directors (ESCDs).

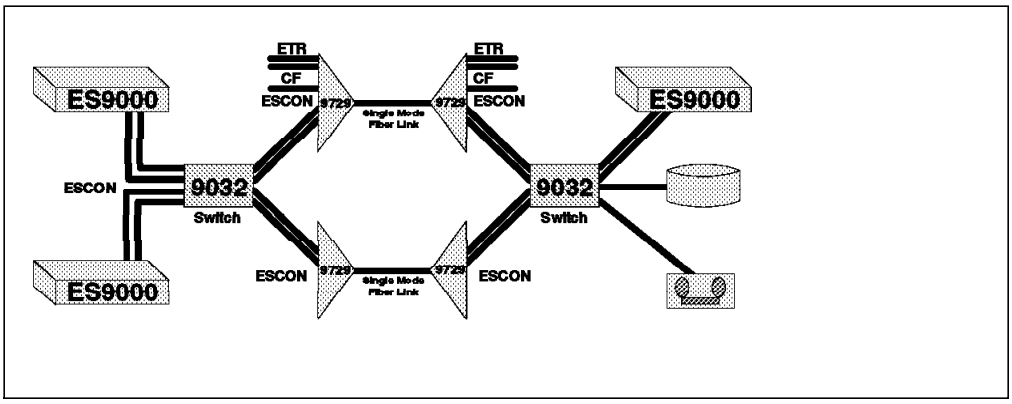


Figure 13. Illustration of Two Remotely Connected Data Centers

When used in a large system environment, the IBM 9729 can extend the following types of links:

- **Enterprise Systems Connection Architecture (ESCON)**

  The 9729 can attach ES/9000 processors, S/390 Parallel Servers, S/390 G3 Enterprise Servers, S/390 Multiprise 2000 servers, ESCON Directors, and the 9037 Sysplex Timer Model 2 using ESCON interfaces with 62.5/125 micron or 50/125 micron fiber and an IBM duplex plug.

- **ETR**

  The 9729 is able to extend the External Timer Reference (ETR) links from an IBM 9037 to all Central Processor Complex (CPC) attachment cards using the same type of fiber as ESCON channels. The 9729 can also extend the Control Link Oscillator (CLO) links between the 9037 Model 002s in an expanded availability configuration.

- **Coupling Facility Channels**

  The 9729 can attach to both single mode ISC and HiPerLinks using the ISC for Coupling Links IOC card. These links use 9/125 micron fiber with a 1300 nm optical signal and a data rate of 1.0625 Gbps.

- **DASD Remote Copy**

  The 9729 can be used to extend the ESCON links between two 3990 Model 006 DASD control units in a Peer-to-Peer Remote Copy (PPRC) configuration.

This redbook discusses each of these applications in succeeding chapters. However, we first present an overview of the Parallel Sysplex environment.

3.1 Parallel Sysplex Overview

Most of the 9729s that are installed today in the large system environment are used in the IBM Parallel Sysplex environment. Because of this important connection, we provide a brief overview of this environment in the context of installing and implementing the IBM 9729 in this environment. For more information on Parallel Sysplex, an excellent source is *OS/390 MVS Parallel Sysplex Configuration, Volume 3:Connectivity*, SG24-2077. This redbook is one in the three-volume series on IBM Parallel Sysplex.

For our purposes, the term *system* is used to describe the operating system. In addition, the following terms may also be used to indicate the operating system:

- MVS
- OS/390

The SYStems comPLEX (sysplex) is a set of MVS systems communicating and cooperating with each other through certain multi-system hardware components and software services to process common workloads. The sysplex addresses customer requirements for growth, continuous availability, enterprise-wide access to data and applications, as well as efficient systems management and operations.

IBM has introduced two kinds of MVS sysplexes:

- Base sysplex
- Parallel Sysplex

Both approaches link individual mainframes into a single-system image. However, they differ in the means they use to communicate across systems. In a base sysplex, central processing complexes (CPCs)[12] connect using channel-to-channel communications and a shared data set to support the communication.

In a Parallel Sysplex, high-performance data sharing across multiple systems is implemented through a coupling facility (CF). The coupling facility is basically a separate mainframe that all the other CPCs link to in order to make high-performance sysplex data sharing possible. It is the coupling facility that is the key differentiator between the base sysplex and the Parallel Sysplex

The IBM 9729 is capable of extending these CF links between each CPC and the coupling facility. These links use the 9729′s ISC IOC interface to the 9729 since it is Inter-System Coupling (ISC) Channel that is the physical interface used by CF links. The details of using the 9729 to extend CF links can be found in 3.5, “Using the 9729 with Coupling Facility Links” on page 64.

Anytime there is more than one CPC involved, an IBM 9037 Sysplex Timer is used to synchronize the time on all systems. This is the case whether it is a

[12] In the context of the IBM ESA/390 architecture, a Central Processing Complex (CPC) consists of one or more Central Processing Units (CPUs) and associated hardware units (such as main and expanded storage, TOD clocks, and channels) that can be configured to operate under the control of a single operating system.

base sysplex or a Parallel Sysplex. The 9037 is a key element in the sysplex. It is a centralized External Time Reference (ETR) designed to distribute Time-of-Day (TOD) information to all attached CPCs in the sysplex, thus maintaining TOD clock synchronization between the supported CPCs. Attachment by the CPCs to the External Time Reference (ETR) provided by an IBM 9037 Sysplex Timer permits events initiated by different processors to be time stamped in sequence. This ensures that multiple OS/390 and MVS/ESA systems can appear as a single system image, delivering the flexibility of running applications simultaneously on multiple systems.

The IBM 9729 can be used with the IBM 9037 Sysplex Timer in two different ways. First, it can be used to extend the links between the CPC and the 9037. These links are called ETR links and they use the ETR/ESCON IOC of the 9729.

The 9729 can also be used to extend the links between 9037s when running an ETR network in the so called expanded availability configuration. The links between 9037s are called Control Link Oscillator (CLO) links and they also use the ETR/ESCON IOC to interface to the 9729.

For the details of using the 9729 in an ETR network, please see 3.4, “The 9729 in an External Time Reference (ETR) Network” on page 48.

The system software in a sysplex environment allows the CPCs to be linked together into a single system image. The base sysplex uses the Cross-system Coupling Facility (XCF), which is a component of OS/390 or MVS, to provide functions that support cooperation between authorized applications, either on the same or different systems. In addition to XCF, a Parallel Sysplex uses Cross-system Extended Services (XES) which is a set of services that manage data within the coupling facility and allow authorized applications or subsystems running in the Parallel Sysplex to share data.

In short, Parallel Sysplex builds on the base sysplex capability and allows you to increase the number of CPCs and MVS images that can directly share the work. Up to 32 MVS systems may be coupled to appear as a single image to the user.

The capability of linking many systems and providing multi-system data sharing makes the sysplex ideal for parallel processing, particularly for online transaction processing (OLTP) and decision support.

3.2 Availability Considerations

As discussed in 2.3.1, “Dual Fiber Switching Feature” on page 17, the IBM 9729 supports two fibers between a single pair of 9729s where one fiber is the primary and used for active transmission, and the other fiber is the backup. In case of a failure of the primary trunk, the 9729s automatically switch to the backup fiber. The time for the 9729 to detect the loss of light, switch from the primary to the backup fiber, and then re-establish all the channels on this backup trunk is a maximum of 2.3 seconds.

However, in the sysplex environment, the fact that the IBM 9729 can use a backup fiber does not guarantee a high availability configuration. Many of the links that are carried over the 9729 trunks in a scenario such as this are critical to the operation of the sysplex. They can not tolerate a two-second period of downtime.

For this reason most of these links, for example, ETR and CLO links, are designed to use redundant connections. However, if both these redundant connections are carried across the same 9729 trunk and you lose the trunk (for example, the fiber gets cut) you lose both links! This is true despite the availability of the backup fiber.

The following text describes what happens to each of the link types in a sysplex environment if the 9729 trunk fails. For this example, assume that all the cross-site links in a multi-site Parallel Sysplex go through one pair of 9729s. The 9729s are equipped with the Dual Fiber Switching Feature and use a pair of fibers, each taking a different route.

ESCON links

The channel subsystem detects the loss of light and attempts to redrive the I/O through other non-functional paths resulting in the I/O being unsuccessful.

PPRC links

The primary DASD control unit detects the loss of light within 1 second (ESCON architected channel timeout) and attempts to redrive the I/O through other paths within 10′s of milliseconds. Since these too are unsuccessful, the PPRC volume pairs are suspended.

The primary DASD CU detects when the links are repaired and automatically reestablishes the paths. However, the installation must issue CESTPAIR to reestablish the duplexed PPRC volume pairs.

CF links

Depending upon the error, OS/390 sysplex (XES) support either redrives the request for up to one second and validates pathing if the operation was unsuccessful, or immediately validates pathing. Path validation tries to identify and activate the message path for up to 2.5 seconds. If the identify and activate processing succeeds, the path is considered good and the request is redriven; otherwise, the CF connectivity is lost. The IBM 9729 will switch to the backup fiber prior to XES performing path validation and redriving the request.

Sysplex Timer

The secondary clock goes offline. This can take from a couple of milliseconds up to a mega-microsecond (1.048576 seconds) depending upon when the fault occurs. The systems in the same data center as the secondary clock do not receive timer signals (from either the local links or the cross-site links) and therefore enter disabled wait states.

3.2.1 Using Two Pairs to Achieve Continuous Availability

To avoid the above problems, you can deploy two *pairs* of IBM 9729s in your Geoplex. Two pairs of 9729s will yield continuous availability for redundant links in the event of single failures provided that certain other steps have been taken such as cable routing and power considerations.

3.2.2 Cable Routing Considerations

Figure 14 shows a deployment of two pairs of 9729s in a two-site geoplex.

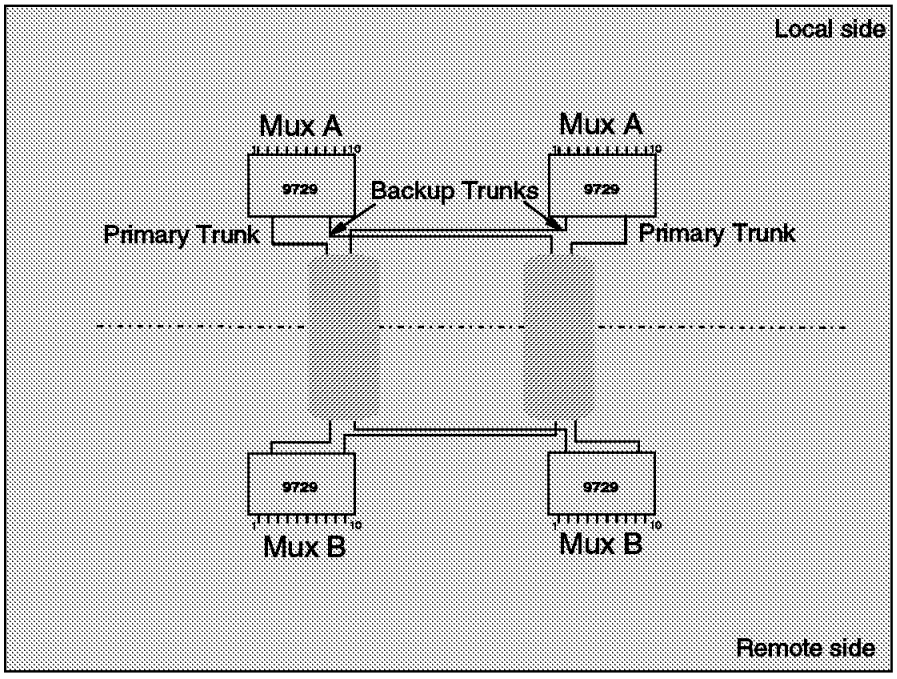


Figure 14. Routing Fiber Paths with Parallel Sysplex Configurations

In this configuration, each 9729 trunk carries one each of the redundant ETR and CLO links between the sites. Also, two different routes are used for the 9729 trunks. One route carries the primary trunk of one 9729 pair and the backup trunk from the other pair. The other route carries the backup trunk for the first 9729 pair and the primary trunk for the second pair.

The following recommendations apply to the CLO and ETR links:

- Each CLO link should be routed across a separate path. The greater the separation, the lower the probability that an environmental failure (accidental or otherwise) will affect both links. In addition, the CLO links should be routed across separate paths from the fiber optic cables that attach the SysPlex Timer to the CPCs. If the CLO links are accidentally severed, the SysPlex Timer in the primary data center will is still able to transmit to the CPCs located in the secondary or remote data center.
- Separately route the redundant fiber optic cables from each SysPlex Timer to the CPC across a separate path.

3.2.3 Power Considerations

Even with two pairs of IBM 9729s deployed, careful power planning is required to maximize the probability that at least one pair is available at all times even in the event of a failure or a power outage. The following recommendations will help to ensure that this is the case:

1. At a minimum, the two 9729 units in each data center (for example, the two A sides) should be connected to separate A/C power circuits. However, an even better solution is to connect them to separate power sources.
2. Protect the power connections by adding a dedicated Uninterruptible Power Supply (UPS) for each 9037-2 and 9729.
3. It is especially important to keep the power source of at least one of the 9729s separate from the source for the SysPlex Timer unit in the primary data center (the primary SysPlex Timer). By doing so, you greatly improve the probability that the OLS signal from the primary timer is received by the secondary timer during a power outage. If the primary SysPlex Timer and the extenders on each CLO link experience a power outage at exactly the same time, the OLS signal may not be received by the secondary timer, resulting in all OS/390 images to be placed in a non-restartable wait state.

3.2.4 Single Failure Scenario

Given this configuration of two pairs of 9729s, using the same scenario described above, the following text describes the effect of losing one of the primary trunks:

ESCON links

The channel subsystem detects the loss of light and redrives the I/O through other functional paths, resulting in the I/O being successful.

If the logical path was lost, the channel will automatically re-establish the paths and present an I/O resource accessibility report. IOS will bring back offline paths and start processing any queued I/O requests.

PPRC links

The primary DASD CU detects the loss of light within 1 second (ESCON architected channel timeout) and attempts to redrive the I/O through other functional paths within 10′s of milliseconds.

The primary DASD CU detects when the links are repaired and automatically reestablishes the paths.

CF links

Depending upon the error, XES support either redrives the request for up to one second and validates pathing if the operation was unsuccessful, or immediately validates pathing. Path validation will try to identify and activate the message path for up to 2.5 seconds. If identify and activate processing succeeds, the path is considered good and the request is redriven; otherwise the link and CF connectivity are lost. The IBM 9729 switches to the backup fiber prior to XES performing path validation and redriving the request; if path validation is unsuccessful, XES loses the link and redrives the request over the other link (assuming it goes to the same CF).

Sysplex Timer

The secondary clock does not go offline since there are two CLO links that go through a different pair of IBM 9729s. Hence, the systems do not enter disabled wait states. (The CPC may switch to the other clock if the active clock time signal is lost).

Both CLO and ETR links become operational automatically once the fiber is restored.

3.3 The 9729 in an ESCON Environment

This section provides information on using the IBM 9729 with ESCON devices.

3.3.1 ESCON Operation

This section provides some background information on the Enterprise Systems Connection (ESCON) architecture. It is not meant to be a complete reference on ESCON. For more information on ESCON, please see the IBM redbook *Enterprise Systems Connection (ESCON) Implementation Guide*, SG24-4662.

The ESCON I/O interface consists of a set of media specifications, physical and logical protocols, and elements of the ESA/390 architecture that allow the transfer of information between an ESA/390 channel subsystem and a control unit or ESCON director (ESCD).

ESCON offers many advantages over the previous S/370 Bus and Tag architecture for connecting I/O channels. One of these advantages is the excellent bandwidth and signal characteristics of fiber optics that allow the previous scheme of parallel transfer to be replaced with high speed serial transmission.

ESCON channels provide 160 Mbps[13] point-to-point links using light emitting diode (LED) transmitters in the 1300nm range over multimode fiber. An ESCON interface uses two fiber stands in order to provide a duplex connection.

ESCON is a circuit-switched protocol. The reason for this is the very fast response times required at the computer channel level. Thus in ESCON protocol you make a connection (just like making a telephone call) between the processor and an I/O device before you can start to transfer data. In practice, the time you need to hold this call is very long in relation to the data transfer time. The result is that when interconnecting computer centers you often need a large number of ESCON channel connections (sometimes up to 50 or 60).

3.3.1.1 Frames and Sequences

ESCON links use a synchronous bit stream to send user data and control information between the channel and a control unit. The bit stream can represent two different types of data constructs: Frames and Sequences.

Frames: The frame is the primary unit of information transfer. It consists of a group of transmission characters organized according to a defined format that includes address and control information, user data, and frame delimiters. The frame also includes a cyclic redundancy check (CRC) field, which assists in detection of transmission errors in the frame.[14]

Sequences: A sequence is a special stream of characters used for certain primitive signaling functions which, because of unusual conditions, cannot be performed reliably using frames. For example, if the error rate on a link is high, frames are likely to fail the CRC test.

---

[13] The maximum data transfer rate of an ESCON channel is 160 Mbps (20 MBps) although the actual baud rate on the fiber is 200 Mbps. The difference is due to the use of line 8/10 coding that is used in the transmission. In most IBM documentation on ESCON the quoted data rate is 200 Mbps.

[14] If a frame is found to be in error, it is discarded and the other side retransmits it.

Each sequence consists of the continuous repetition of a particular ordered set of bits that stops when an event that is defined for that particular sequence occurs.[15] Continuous repetition ensures that the sequence will be correctly recognized by the receiver even in the presence of a high link-error rate.

The following sequences are defined:

Not operational (NOS)

The sender is not receiving a signal or cannot synchronize with the signal it is received.

Offline (OLS)

The sender is offline.

Unconditional disconnect (UD)

The sender does not know whether it is connected to another ESCON interface through the ESCON Director (ESCD) and is attempting to ensure that there is no connection.

Unconditional disconnect response (UDR)

This is the response to UD.

3.3.1.2 Connection Initialization Procedure

One of the first events that has to occur is to activate the link. To perform link initialization, each ESCON interface transmits a prescribed sequence while simultaneously attempting to acquire bit and character synchronization from the other end of the link. When it has acquired character synchronization, the interface indicates that by transmitting a prescribed response sequence.

Using system configuration information from the I/O Configuration Data Set (IOCDS), the channel determines those control units with which it will communicate and then establishes logical paths to these units. Logical-path establishment provides each control unit with the link and logical addresses that have been assigned to the channel image. When a logical path is being established, the channel and control unit are essentially agreeing on the configuration and agreeing that all of the necessary initialization procedures required to support device-level communication have been successfully completed. After this, either the channel or control unit can initiate device-level communication.

Important Note

IBM 9729s installed into an ESCON link do *not* participate in the link initialization because they are transparent to the ESCON devices. The link initialization is done between the nodes defined in the IOCDS. It is not necessary to make any changes in the IOCDS to implement a 9729 link.

The next initialization step is the exchange ID (XID) procedure. In this step, the device at each end of the link sends its unique identifier to the other side. The identifier includes information as such as the type of product and its serial number. It is used for verifying the configuration and for problem determination.

[15] Events that terminate a sequence transmission include the receipt of a response sequence or a timeout as well as several others.

Important Note

The 9729 does not participate in the XID process because it is transparent to the ESCON devices. Therefore it is not defined in the Node descriptor list.

To begin communicating with the control unit, the channel subsystem (CSS) executes a Start Subchannel (SSCH) command.

3.3.1.3 Link Errors

The ESCON architecture also specifies how to handle transmission errors on the link. Transmission errors can be caused by transient noise or hardware malfunctions as well as a failed or failing link. The architecture requires the receiver to detect and respond to the following types of transmission errors:

Link-signal error

The amplitude, or power, of the received signal is below the value required for reliable communication. Or, the receiver has determined that it has lost synchronization with the incoming character stream.

Code-violation error

The receiver has detected an invalid transmission character.

CRC error

A received frame has failed the cyclic redundancy check.

3.3.1.4 Channel I/O Operations

The channel subsystem (CSS) communicates with the control unit via Channel Control Words (CCWs). The CCWs result in a number of frames being sent between the two entities on the link. The frames are of the type command, control, data, or status. The actual frames and number of frames sent are dependent on the CCWs associated with the SSCH and the maximum Data in Block (DIB) size supported by the channel and control unit. The important point that is relevant to a 9729 configuration is that they take the form of a request and a response between the channel and the control unit.

A frame exchange for one CCW read operation is shown in Figure 15 on page 34.

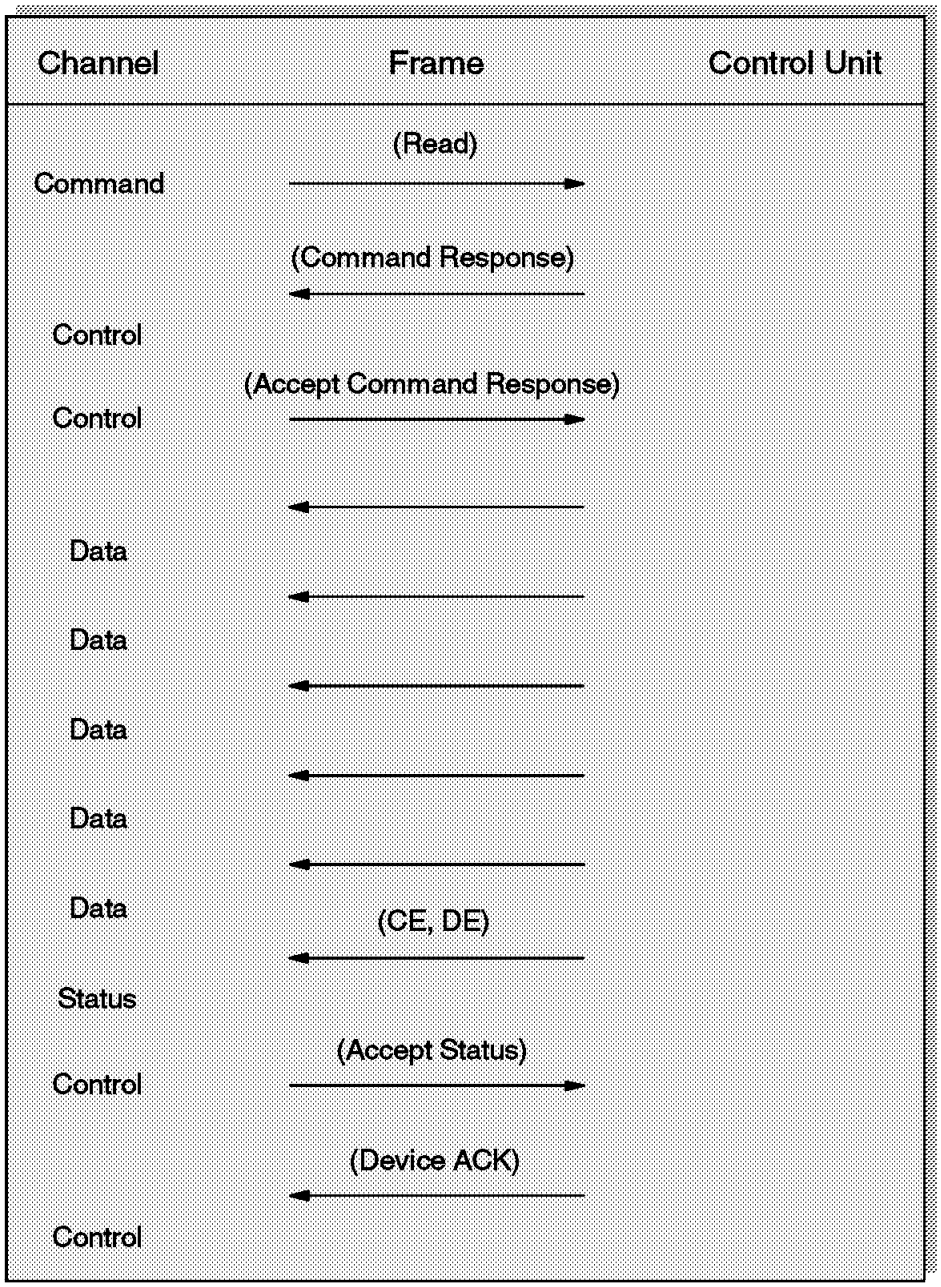


Figure 15. Frame Exchange for a Single Read CCW

As can be seen from Figure 15, CCWs can often take many frames to complete.

3.3.2 Using 9729s with ESCON

The ESCON architecture allows for a maximum transmission distance of a single link of 2 or 3 km depending on the type of fiber that is used. Sometimes, increased distances are needed. A few examples of where increased distances might be needed are:

- Interconnecting multiple computer systems across sites
- Placing printers and terminal controllers near their users
- Placing critical data storage devices to placed in secure locations

Prior to the introduction of the IBM 9729, users did not have many options available which would allow them to extend the distances allowed for an ESCON connection. IBM did announce a feature called the ESCON Extended Distance Feature (XDF) that uses laser transmitters and single-mode fiber to extend the maximum transmission distance to 20 km. However, this feature has been withdrawn from marketing for use with processors since May 1995 even though it is still available for ESCDs and some control units.

Note: It is *not* possible to use the XDF and the 9729 on the same ESCON channel connection.

The introduction of the 9729 has radically alleviated this distance limitation on ESCON connections. With the 9729, we now have the capability to extend an ESCON channel to distances up to 50 km.

3.3.3 ESCON Configurations

ESCON supports two kinds of topologies: point-to-point and switched point-to-point. The link-level and device-level functions and protocols are identical for both topologies.

The point-to-point topology consists of a single link between a channel and a control unit. The IBM 9729 can be used in this topology, for example, to establish a remote tape archive or connect a remote printer.

The switched point-to-point topology consists of a number of channels and control units with their ESCON interfaces each connected by a point-to-point link to a port on an ESCON Director (ESCD). The ESCD permits any channel to communicate with any device attached to any control unit.[16] The IBM 9729 is more powerful in this environment. It can be used to connect or mirror storage devices, ESCDs and control units.

Figure 16 on page 36 shows an example of a remotely connected DASD subsystem. Two pairs of 9729s are used in order to provide full redundancy of the ESCON connections.

[16] For a given installation, the system configuration definition generally restricts the devices with which channels can communicate.

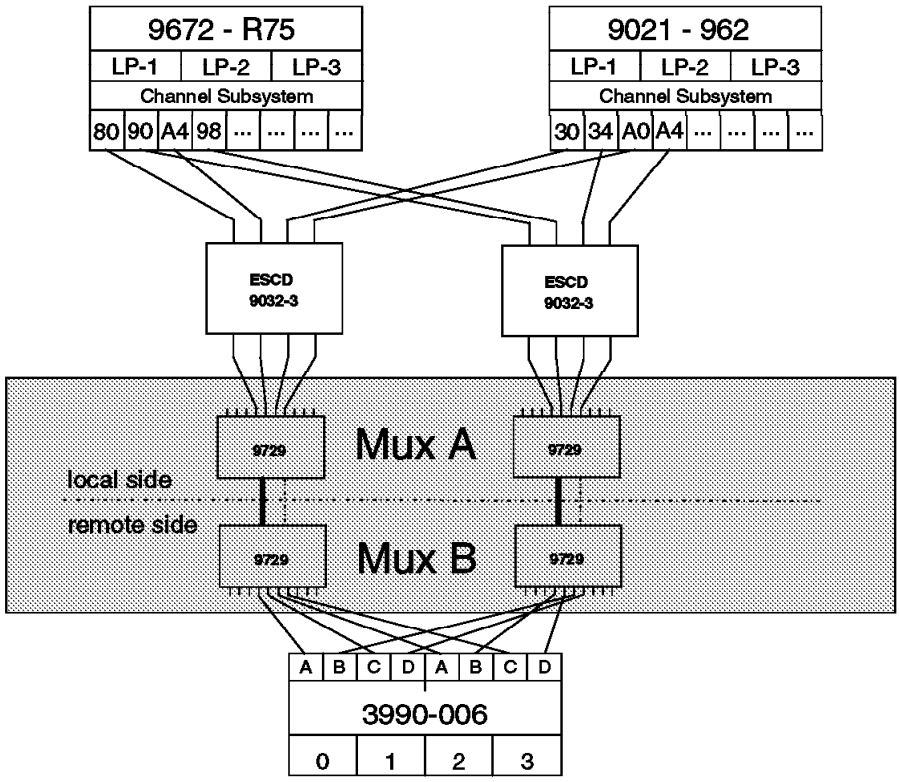


Figure 16. Remotely Connected DASD Subsystem Using IBM 9729s

3.3.3.1 Channel-to-Channel Configuration

The switched point-to-point topology can also be used to connect processors in a Channel-to-Channel (CTC) configuration. An ESCON CTC facility provides a high-bandwidth intersystem network, suitable for the rapid transfer of either control information or large data blocks between cooperating S/390 systems. ESCON CTC transfers can operate at data rates up to 17 MBps.

With the switched point-to-point topology provided by an ESCD, any ESCON I/O channel of one system can connect to an ESCON I/O channel of any other system that is attached to the same ESCON director.[17]

Also, the ESCON Multiple Image Facility (EMIF) allows the processor channel subsystem (CSS) to provide physical path sharing by extending the logical addressing capability of the ESCON architecture to the host images (PR/SM logical partitions).

The IBM 9729 can be used to extend the distances of CTC configurations up to 50 km. Figure 17 on page 37 shows an example of an ES/9000 9021 processor connected to an S/390 9672 CMOS processor in a remote CTC configuration. Two pairs of 9729s are used in order to provide full redundancy of the CTC connection.

[17] For CTC operation, the ESCON I/O channel of one of the communicating systems must have been initialized with the ESCON CTC microcode. This is necessary in order to support current System/370-XA CTC protocols.

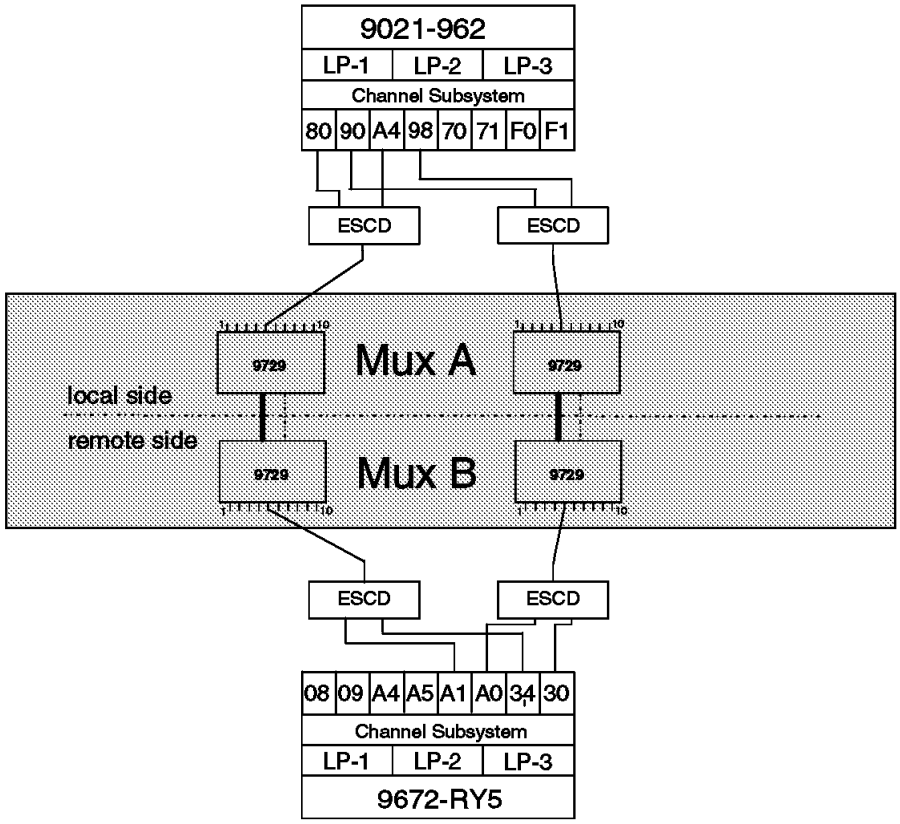


Figure 17. Remote Channel-to-Channel Connection

3.3.3.2 Connecting Conversion Channel (CVC)

To support the migration from System/370 Bus and Tag channels to System/390 ESCON channels, IBM announced two products:

9034 Model 1

This converter is used to connect parallel channel I/O control units to processors with ESCON channels. The control unit attaches to the converter through parallel channel bus and tag cables. The converter then attaches to an ESCON channel operating in ESCON Converter Mode[18] through a fiber optic cable.

9035 Model 2

This converter allows the attachment of IBM 3990 Storage Control Units with ESCON adapters to many current processors with parallel channels.

These converters allow a customer to migrate from existing parallel channels to ESCON channels without modifying the I/O control unit, or the application software and with only a slight modification to the processor hardware configuration.

[18] An ESCON channel can be operated in ESCON Converter Mode by specifying that mode while creating the I/O configuration data set (IOCDS). Changing modes requires reconfiguration of the IOCDS.

The IBM 9729 can be used to extend the links of both of these converters. Figure 18 on page 38 shows 9729s being used to extend the link between a host and a remote printer.

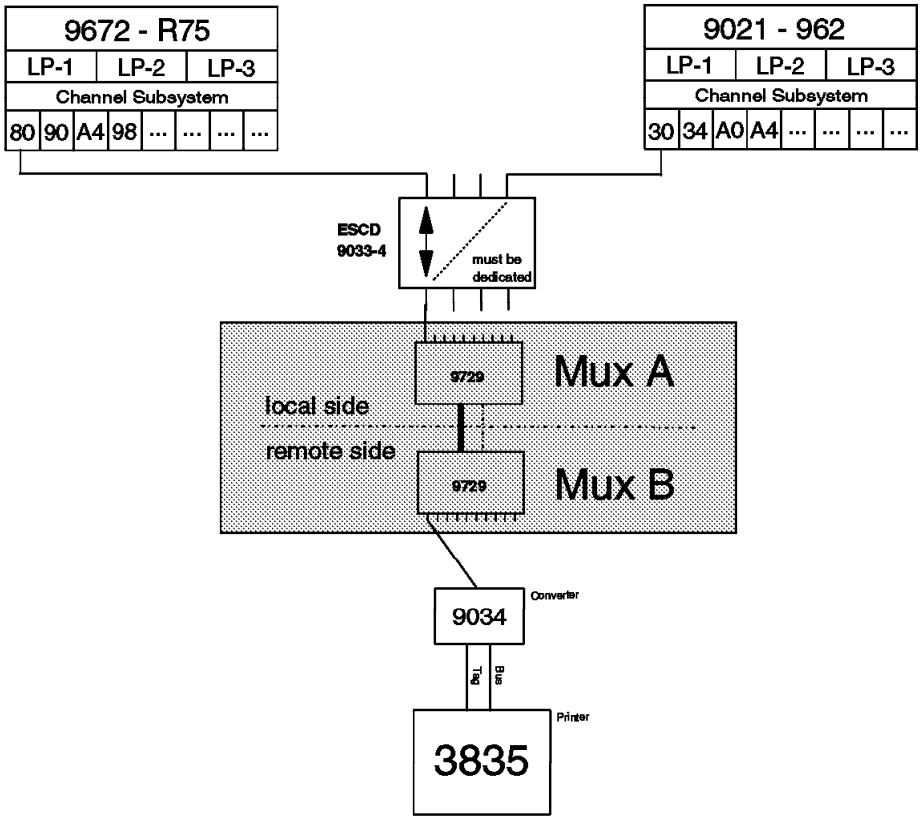


Figure 18. Remote Printer Connected to a Conversion Channel

Similarly, the 9729 can also be used to channel attach 3745 communications controllers. IBM has tested this configuration up to a distance of 20 km.

3.3.4 ESCON Distance Considerations

As discussed in 2.4.5, “Other Distance Limitations” on page 22, some protocols are sensitive to propagation delays. S/370 channel I/O, mostly because of the request/response nature of CCWs, is such a protocol. Because ESCON channels carry S/370 channel I/O, the overall throughput of ESCON connections will be sensitive to propagation delay. Figure 19 on page 39 depicts how the maximum data rate of an ESCON connection decreases as the link distance increases.

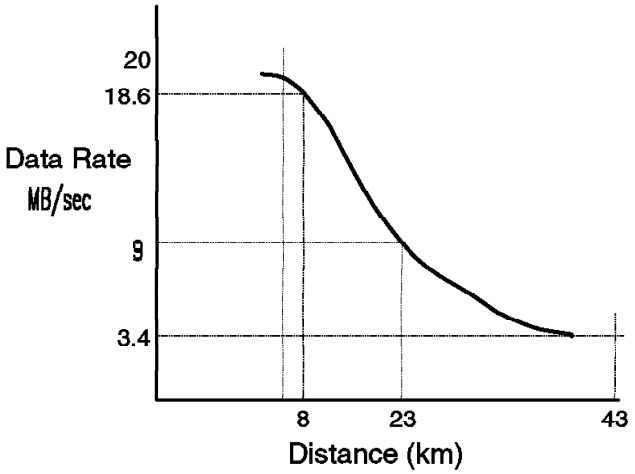


Figure 19. Relative ESCON Throughput As a Function of Distance

As can be seen from Figure 19, the ESCON channel can maintain an 18.6 MBps data rate up to a distance of about 8 kilometers.

Note: This data rate is a maximum rate and it assumes 100% channel utilization which does not represent a typical workload. Actual data rates on ESCON links vary depending on workload characteristics. At a distance of 9 km, we can still drive the channel at a rate of 17.6 MBps. If we increase the channel length up to 23 km, we can still gain a transfer rate of 9 MBps.

Several points should be made about this figure. First, these effects are due solely to propagation delay and the manner in which the ESCON protocols function. A pair of 9729s inserted into a link cause no performance penalty by the multiplexers themselves. It is merely the delays caused by the signal propagation in the fiber that slows the data rate.

Note: This is still very good throughput compared to bus and tag connections, even at long distances. In fact, we can transfer data at about the same rate on an ESCON *serial* channel over a distance of 50 kilometers that we can with a bus and tag *parallel* channel at a distance of only 122 meters.

Second, most ESCON devices cannot handle the maximum data rate anyway. Table 1 shows you the maximum transfer rates of various control units.

Table 1. I/O Data Rate for IBM Control Units

Control Unit and Model	Data Rate (MBps)	Control Unit and Model	Data Rate (MBps)
3172 1,3	1.5	3174 12L/22L	1.5
3490 Cxx	4.5/9	3490 A01	4.5
3490 A02/A10	4.5/9	3490 A20	9
3990 2,3/6	Device speed w/o cache; up to 17 w/cache	RAMAC Array Subsystem	17
RS/6000 /SP2	6 MB on 10 MB channel; 11 MB on 17 MB channel	ESCON CTC	17
9343 Dxx	10/17	3995 133	1.2

You can translate this maximum data rate information into maximum distances allowable between the CPU and the control units. Table 2 on page 40 shows this information.

Table 2. Maximum Distances of Control Units

Control Unit and Model	Maximum Distance (km)	Maximum Distance (mi)	Control unit and Model	Maximum Distance (km)	Maximum Distance (mi)
3172 1,3	43	26.7	3174 12L/22L	43	26.7
3490 Cxx	23	14.3	3490 A01	23	14.3
3490 A02/A10	23	14.3	3490 A20	23	14.3
3990-002/003	15	9.3	3745	43	26.7
3990-006	43	26.7	3746-900/950	43	26.7
RS/6000 /SP2	43	26.7	3900	43	26.7
9343 Dxx	43	26.7	RAMAC Array Subsystem	43	26.7
3995 133	9	5.6	ESCON CTC	50	31

3.3.4.1 RAMAC Performance over a 9729 Link

IBM has completed performance testing for RAMAC 3 arrays[19] connected to host units over a 9729 link. This section presents some of the results from that testing.

Figure 20 on page 41 shows a test of random writes to DASD using a block size of 6.5 KB.

[19] RAMAC is an IBM high-performance, high-capacity, and fault-tolerant storage solution consisting of 9390 storage controllers and 9391 RAMAC Array DASD units.

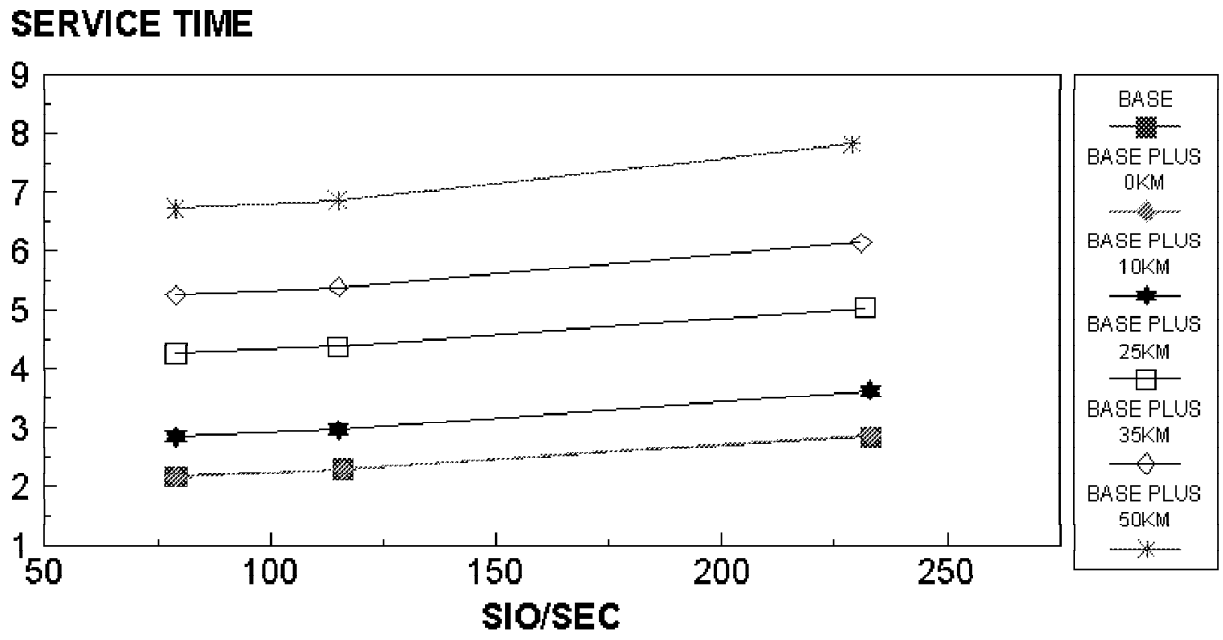


Figure 20. Testing Random Write Operations at Various Distances

Important Note

The line for the *base* in this chart is superimposed on the line for the *base plus 0 km*. The former is a baseline measurement without any 9729s in the link while the latter is a measurement with a pair of 9729s with a very short fiber trunk in between. (All results in this section used the same methodology.)

This is confirmation of the fact that the 9729 units themselves do not add any delay to the link.

Figure 21 on page 42 shows a test of a random workload with a R/W ratio of 2.0 and a read cache hit ratio of 70%. The block size was 4 KB.

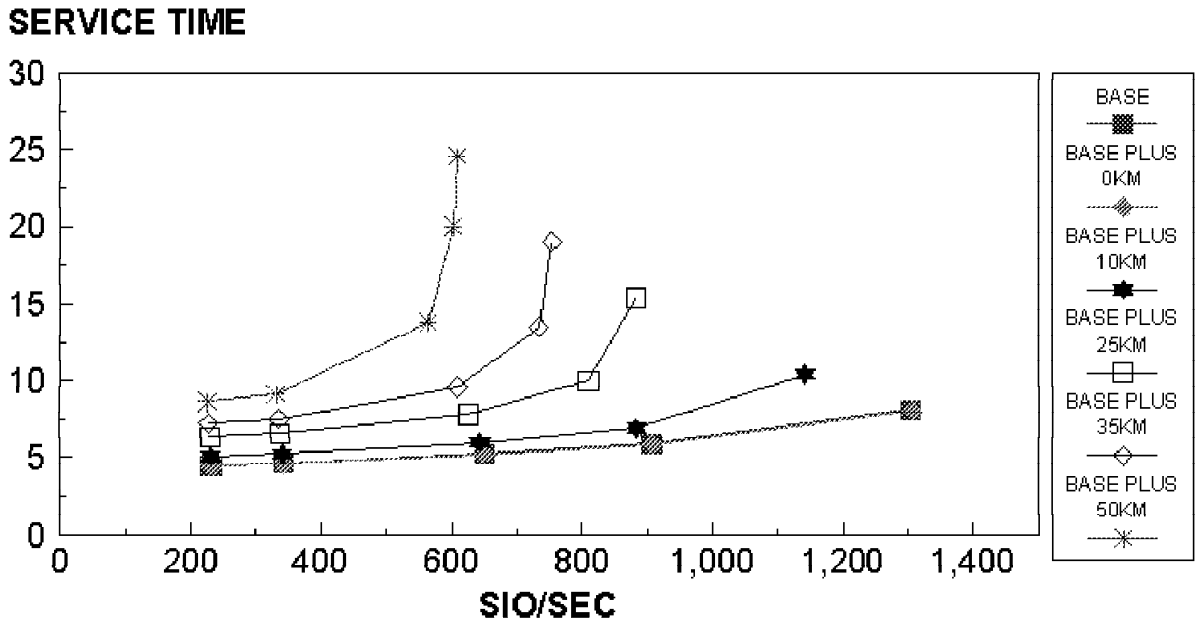


Figure 21. Testing a Random Workload at Various Distances

Figure 22 shows a test of a sequential writes to a QSAM volume using 27 KB blocks. The BUFNO parameter for this test was set to 6.

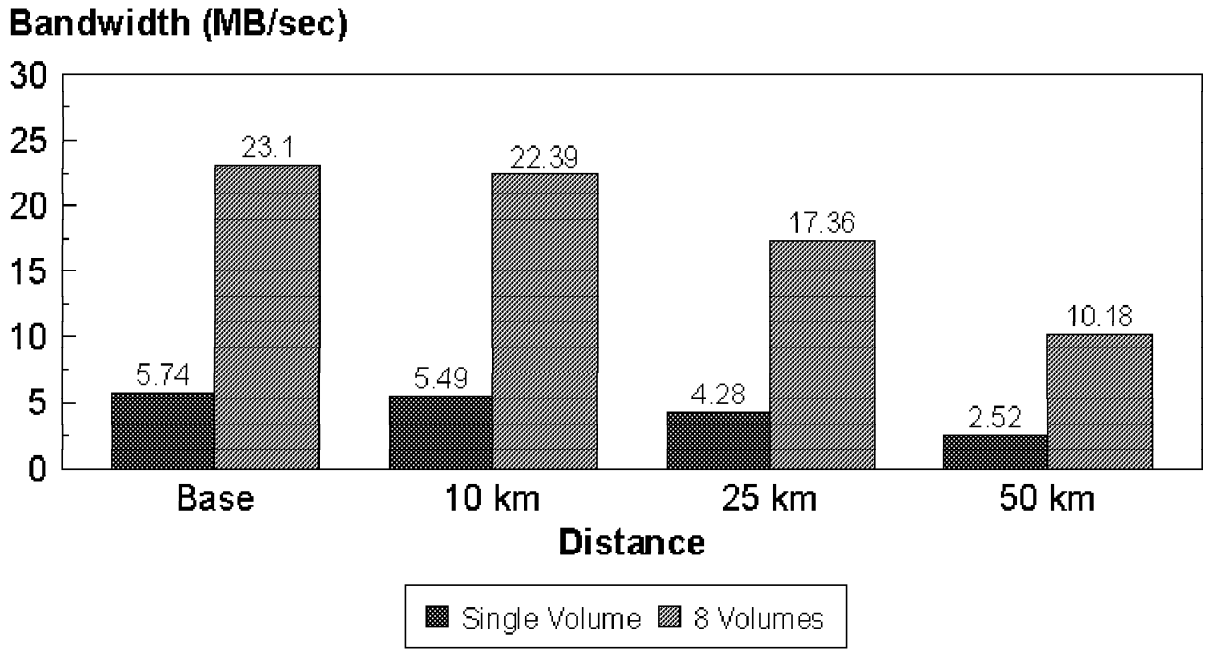


Figure 22. Testing Sequential Writes at Various Distances

3.3.4.2 Distance Limitations on ESCON Tail Circuits

As discussed in 2.4.2, “Jitter” on page 19, the effects of jitter on a communications link can be cumulative. This is true in the case of ESCON tail circuits. Since ESCON interfaces use multimode fiber and since multimode fiber is prone to more dispersion and jitter than single-mode fiber, we have to be careful about the distances between the ESCON devices and the 9729.

Figure 23 shows a point-to-point ESCON connection between a host processor and a control unit. The connection has been extended using a pair of IBM 9729s.

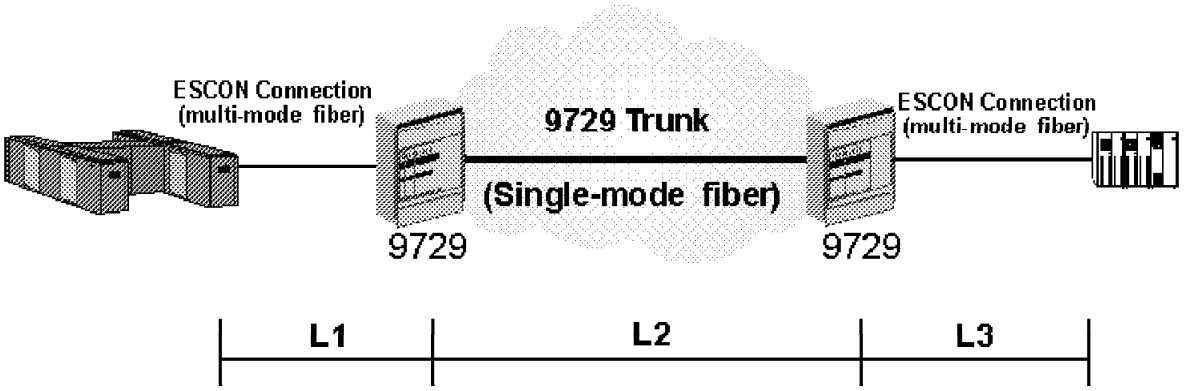


Figure 23. Distance Limitations on ESCON Tail Circuits

The maximum distance of this connection without the 9729s would be 3 km. Since the 9729 re-generates the signal at the IOC interface, you might think that it would be possible to have a 3 km connection between the host and the 9729 (L1) and another 3 km between the 9729 and the control unit (L3). However, since the 9729 does not re-clock the signal, the maximum distance for L1 + L2 is still 3 km.

Even considering this limitation, the 3 km distance is still more than adequate for most customer installations.

3.3.5 ESCON Link Recovery

As discussed in 2.3.1, “Dual Fiber Switching Feature” on page 17, the 9729 has an optional Dual Fiber Switching Feature that electronically switches from one trunk fiber to another if the active trunk fails for any reason (for example, if the cable is cut).

The switchover occurs in less then 2 seconds from the time the active fiber becomes non-operational. Since MVS waits 3 seconds before it drops an ESCON connection and the ISC time out is 10 seconds, this means that the systems can recover from this situation transparently.

The following scenario simulates a failure of the main trunk and takes you step-by-step through the process of the failover from the perspective of the host system.

1. First, display all operations *before* the simulation of the failure of the primary trunk. We use the MVS commands D M=DEV(dev) and D M=CHP(chp) to

be sure everything is online and operational. Figure 24 on page 44 shows these commands.

```
97226 15:01:01.38  D M=DEV(A20)
97226 15:01:02.09  IEE174I 15.00.53 DISPLAY M 490
                   DEVICE 0A20   STATUS=ONLINE
                   CHP                  49 08 A9 9A
                   PATH ONLINE          Y  Y  Y  Y
                   CHP PHYSICALLY ONLINE Y  Y  Y  Y
                   PATH OPERATIONAL     Y  Y  Y  Y

97226 15:01:07.42  D M=CHP(A9)
97226 15:01:08.01  IEE174I 15.01.01 DISPLAY M 497
                   DEVICE STATUS FOR CHANNEL PATH A9
                       0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F
                   00D .  .  .  .  .  .  .  .  .  +  .  .  .  .  .  .
                   0A2 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   0A3 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   0A8 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   0A9 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   0AA +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   0AB +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   ********************** SYMBOL EXPLANATIONS ******
                   + ONLINE    @ PATH NOT VALIDATED   - OFFLINE
                   * PHYSICALLY ONLINE   $ PATH NOT OPERATIONAL
```

Figure 24. All Channels Are Online/Operating

As you can see, CHPIDs 49, 08, A9, 9A are online with operational paths to their devices.

2. Now, we simulate a failure of the primary trunk. (In this case, we unplugged the fiber.) At this point an error occurs on the single-mode fiber link between 9729 A-Side and 9729 B-Side. Figure 25 shows the MVS console when the error occurs.

```
15:01:54.52  IOS581E LINK FAILED REPORTING CHPID=A9 515
                     INCIDENT UNIT  TM=009672/RX5 SER=IBM02-045445 IF=00A9  IC=03
                     ATTACHED UNIT  TM=009032/000 SER=IBM00-010074 IF=00DF
15:01:54.63  IOS581E LINK FAILED REPORTING CHPID=48 516
                     INCIDENT UNIT  TM=009032/000 SER=IBM00-010074 IF=00DF  IC=03
                     ATTACHED UNIT  TM=009672/RX5 SER=IBM02-045445 IF=00A9
15:01:57.50  IOS001E OA3B,INOPERATIVE PATH  A9
15:01:57.53  IOS450E OA3B,A9, NOT OPERATIONAL PATH TAKEN OFFLINE
15:01:59.04  IOS581E LINK FAILED REPORTING CHPID=48 519
                     INCIDENT UNIT  TM=009032/000 SER=IBM00-010074 IF=00DF  IC=04
                     ATTACHED UNIT  TM=009672/RX5 SER=IBM02-045445 IF=00A9
```

Figure 25. MVS Screen at Time of Error

Depending on how many channels are carried over the IBM 9729 trunk, you may see a lot of these messages roll over the MVS console.

If there are CMOS processors installed, the Optical Network icon starts blinking and a message will occur for every channel connected over the link.

If there is an ES9000 processor channel present in the configuration, you could see priority messages, depending on the system configuration.

3. Also at this point, the ESCON director recognizes the dropped link and the status changes on the ESCD console as shown in Figure 26 on page 45.

You will see the Hardware and Connection columns on the Active Matrix go to the OFFLINE status for all relevant ports.

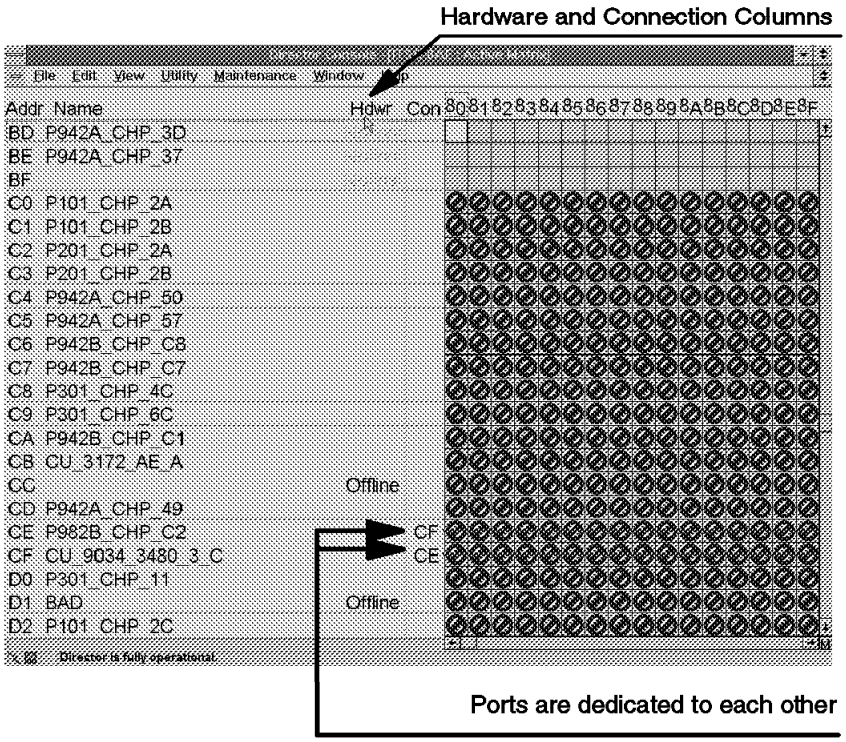


Figure 26. ESCON Director 9032 Model 003 Active Matrix Window

4. Meanwhile, during these events, the 9729s transparently switch to the backup fiber.
5. Now, the ESCON link recovery starts. After both 9729s reinitialize the link, the ESCON interfaces re-establish their links between primary node and secondary nodes. Figure 27 on page 46 shows the messages on the MVS console.

```
15:02:01.49 IOS452I OA91,A9, OPERATIONAL PATH ADDED TO PATH GROUP
15:02:01.49 IOS452I OA92,A9, OPERATIONAL PATH ADDED TO PATH GROUP
            .
            . All Devices
            .

15:02:01.67 IOS582I PATH A9 NOW OPERATIONAL AND BROUGHT ONLINE FOR DEVICE(S): 593
                    OA3B

15:02:01.49 IOS452I OA93,A9, OPERATIONAL PATH ADDED TO PATH GROUP
15:02:01.49 IOS452I OA94,A9, OPERATIONAL PATH ADDED TO PATH GROUP
            .
            . All Devices
            .

15:02:11.32 IOS208I CONTROL UNIT FOR (OOD9,A9) SUCCESSFULLY RECOVERED
15:02:11.87 IOS208I CONTROL UNIT FOR (OA2O,A9) SUCCESSFULLY RECOVERED
15:02:12.33 IOS208I CONTROL UNIT FOR (OA8O,A9) SUCCESSFULLY RECOVERED
```

Figure 27. MVS Screen Showing ESCON Link Recovery

The screen shows all the logical paths were added to path group and the channel returns to an operational state. This also means that all control units were recovered.

6. Finally, we want to verify that all operations have returned to normal, albeit on the backup fiber path. This is done using the same MVS commands as in step 1. Figure 28 shows these commands.

```
97226 15:07:46.44  D M=DEV(A2O)
97226 15:07:46.70  IEE174I 15.OO.53 DISPLAY M 49O
                   DEVICE OA2O   STATUS=ONLINE
                   CHP                   49 O8 A9 9A
                   PATH ONLINE           Y  Y  Y  Y
                   CHP PHYSICALLY ONLINE Y  Y  Y  Y
                   PATH OPERATIONAL      Y  Y  Y  Y

97226 15:07:54.45  D M=CHP(A9)
97226 15:07:55.26  IEE174I 15.01.01 DISPLAY M 497
                   DEVICE STATUS FOR CHANNEL PATH A9
                       0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F
                   OOD .  .  .  .  .  .  .  .  .  +  .  .  .  .  .  .
                   OA2 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   OA3 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   OA8 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   OA9 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   OAA +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   OAB +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
                   ************************ SYMBOL EXPLANATIONS ******
                   + ONLINE    @ PATH NOT VALIDATED   - OFFLINE
                   * PHYSICALLY ONLINE   $ PATH NOT OPERATIONAL
```

Figure 28. MVS Screen Showing All Connections Are Back Online

As you can see, CHPIDs 49, 08, A9, 9A are online with operational paths to their devices, using the alternate fiber.

3.3.6 Problem Determination on ESCON Links

The previous scenario simulates the recovery of a failed link when using 9729s equipped with the Dual Fiber Switch feature. However, there are other failures besides a trunk failure that can occur.

Note: For this scenario, assume that we have one ESCON channel connected to a DASD subsystem channel adapter over a pair of 9729s. For convenience, we refer to the 9729 units as the *A side* and the *B side*. The host side is connected to 9729 A. The DASD control unit is connected to 9729 B. Please refer to Figure 29 to locate the LEDs and test buttons described in the scenario.

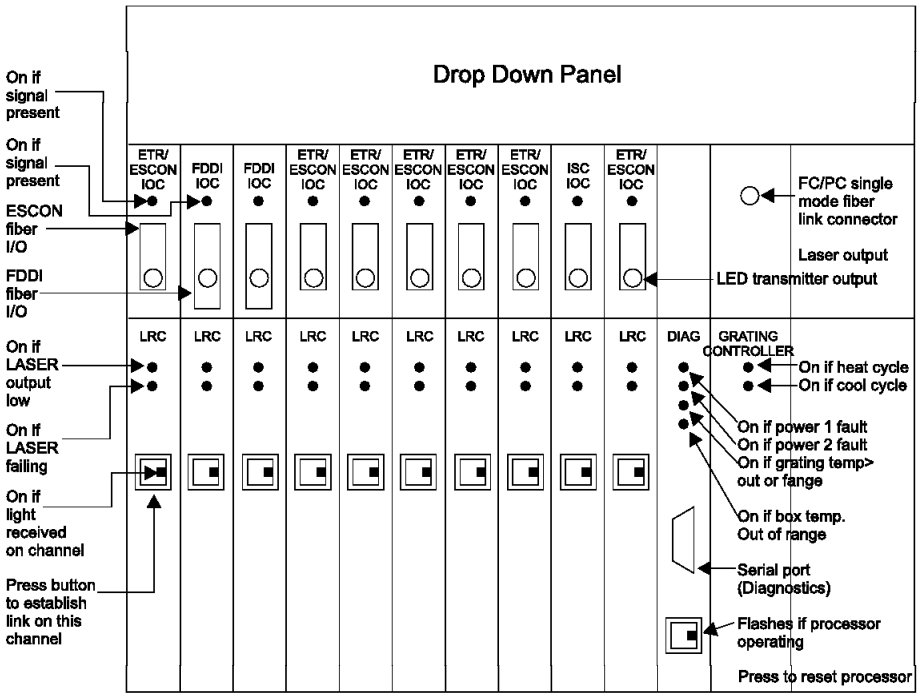


Figure 29. IBM 9729 Front Panel Reference

To determine the source of a failing link, we only have to check a few things:

1. Check 9729 A for the status of the I/O controller (IOC) LED indicators. A green light on an ESCON or ETR/ESCON IOC indicates the presence of light from the ESCON device. If the LED in the relevant slot position is not lit (green), then you must suspect a link problem between the processor and the 9729 A unit.
2. Check the Laser/Receiver Card (LRC) LED indicators on the appropriate channel of the 9729 A unit for the status of *that channel* on the trunk between the two 9729s.
   - An amber Laser Low LED indicates that the laser has drifted outside specification limits.
   - An amber Laser Fault LED indicates that the laser control has been disabled.
   - A green Ready LED indicates that a signal is being received from the corresponding LRC card in the 9729 B unit.
   - A blinking green Ready LED indicates that no signal is being received from the 9729 B unit.

Pressing the Test Laser button causes the LRC card to attempt to establish a connection with the 9729 B unit on this channel. It also resets the amber LED, if lit,[20] so if any LRC error condition is present, you can try to recover using this button. If, after a few moments of pressing this button, the Ready LED does not go to a solid green state, then contact your IBM service representative.

If you found no LED status that indicates an error, you should suspect the fault is on the B side on the ESCON interface to the control unit.

3. Repeat steps 1 and 2 above, as necessary, to isolate the problem on the B side of the connection.

3.4 The 9729 in an External Time Reference (ETR) Network

This section addresses the various factors that have to be considered when attaching an IBM 9037 Sysplex Timer to Central Processing Complexes (CPCs) over IBM 9729 links.

As discussed in Chapter 3, “IBM 9729 in a Large System Environment” on page 25, the 9037 Sysplex Timer is a mandatory hardware requirement for any sysplex environment.

The IBM 9037 Sysplex Timer is known by many names. Most MVS/ESA manuals use the term ETR when discussing the Sysplex Timer. The following names generally refer to the same hardware:

- 9037
- ETR (External Time Reference)
- Reference Timer
- STR (Sysplex Timer Reference)
- Sysplex Timer
- Timer

3.4.1 ETR Network

This section gives an overview of the functions and configurations of an ETR network from the point of view of installing and implementing IBM 9729s in order to extend the ETR and CLO links that are needed in an ETR network. It is not meant to be a complete reference on ETR networks. For more information on the IBM 9037 Sysplex Timer and the functions it provides, please see the redbook *IBM 9037 Sysplex Timer*, SG24-2070.

An ETR network consists of three elements, which are configured in a network:

1. ETR sending unit

   The sending unit is the centralized, external time reference, which transmits ETR signals over dedicated ETR links. It provides a means by which ETR time can be accurately maintained with respect to external standard time services. The IBM 9037 is an example of an ETR sending unit.

[20] The diagnostics card provides the primary method of establishing the connection with the unit on the other side of the trunk. This push button provides a backup method.

2. ETR receiving unit The receiving unit in each CPC receives the ETR signals and includes the means by which the TOD clocks are set and maintained consistent with ETR time.
3. ETR link

   An ETR link is a connection between a sending and a receiving unit.

Figure 30 shows a typical ETR network, which connects the sending unit to all CPCs in an installation. The ETR network may comprise one or more sysplexes and CPCs not belonging to a sysplex.

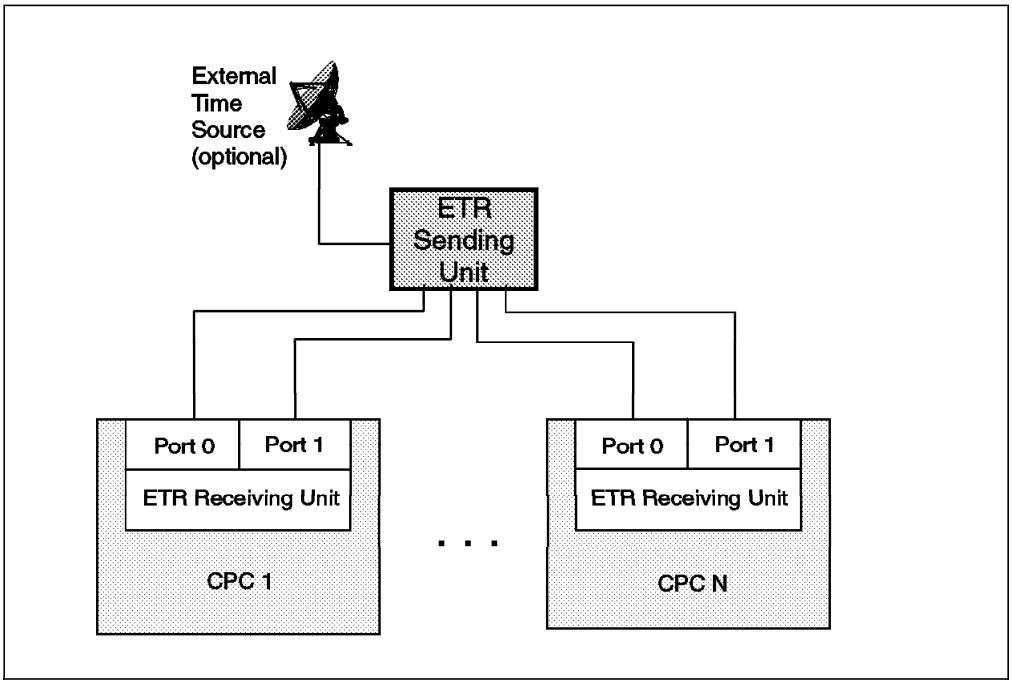


Figure 30. A Typical ETR Network

A fault-tolerant configuration can be provided by coupling and synchronizing two sending units with each other, so that each unit is transmitting consistent ETR timing information. Figure 31 on page 50 shows a typical fault-tolerant ETR network.

The receiving unit at each CPC has two ports; one port is normally connected to a different ETR sending unit of a coupled pair in the same network. This fully duplicated structure minimizes the potential that a single failure can adversely impact the ETR network capability.

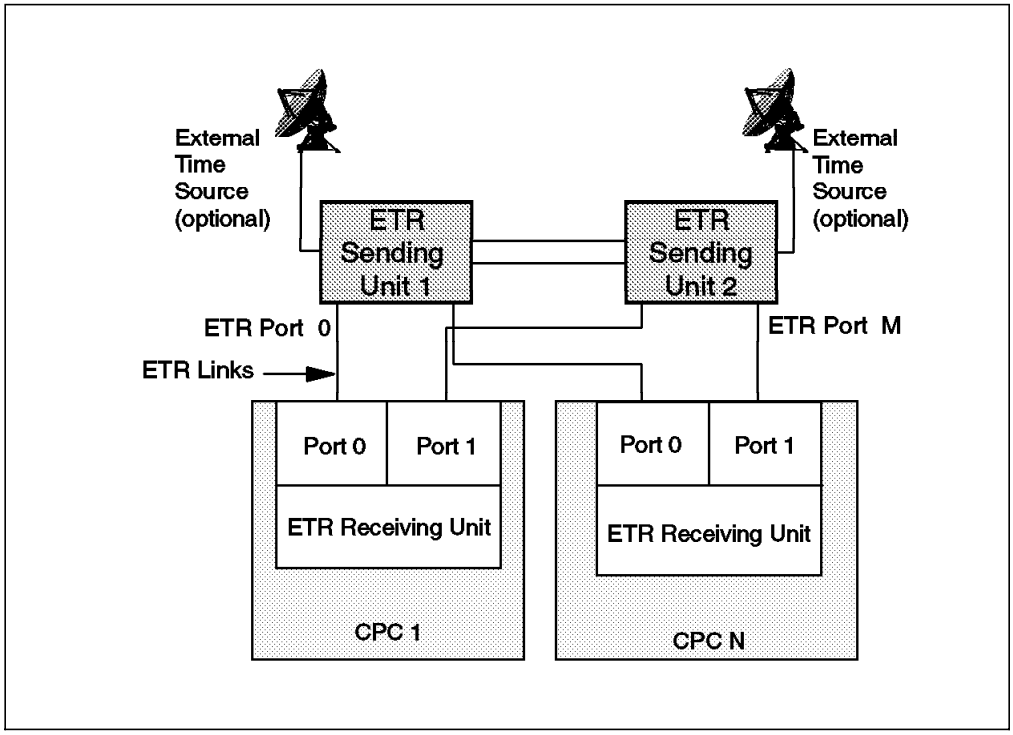


Figure 31. Fault-tolerant ETR Network

3.4.2 9037 Sysplex Timer

The IBM Sysplex Timer is a table-top unit that comes in two different models:

- 9037 Model 001

  The 9037 Model 001 can have up to 16 ETR ports installed. It uses a copper interface for the Control Link Oscillator (CLO) link to another 9037 unit in an expanded availability configuration. The cable length is 2.2 meters and thus the two 9037s must be placed adjacent to one another. Hence, using IBM 9729s to extend these links is not possible.

- 9037 Model 002

  The 9037 Model 002 can have up to 24 ETR ports installed. In addition, the Model 002 has some other enhancements such as more options for connecting the console and improved tracking to an external time source.

  The 9037-002 supports a fiber connection for the CLO links and these links can be extended via IBM 9729s. The interface on the CLO card uses a 62.5/125 micron or 50/125 micron multimode fiber with IBM duplex plugs.

  The maximum supported CLO link distance is 3 km for a 62.5/125 micron cable and 2 km for a 50/125 micron cable. The distance is extendable using IBM 9729s to a maximum distance of 26 km. This is 3 km of multimode fiber between the 9037 and the first 9729 unit, another 3 km between the CPC and the second 9729, with a 20 km single-mode fiber 9729 trunk. In any case, the total cable distance of multimode fiber should not be more than 6 km in total (4 km for 50/125 fibers).

Cables

When implementing an ETR network using IBM 9037s, the customer must supply the following cables:

- The fiber optic cabling used to attach the 9037s to the sysplex attachment feature of each CPC (ETR links). These are bi-directional cables. Each connection from a 9037 port to a CPC attachment port requires two individual fibers, one that carries signals from the 9037 to the CPC and the other that carries signals from the CPC back to the 9037. Each jumper cable is comprised of these two individual fibers.
- Two fiber optic cables used to attach the SysPlex Timer units to each other (CLO links) These are also bi-directional cables.

Keep in mind that when using 9729s to extend ETR and CLO links, you will need two cables for each link, in effect, one for each tail circuit from the device to the 9729 unit.

For example, an ETR link between a single port on a 9037 unit and a single port on a SysPlex Timer attachment feature requires one fiber cable if these two entities are directly attached. If a pair of 9729s is used to extend this link, you will need two of these cables: one from the port on the 9037 to the 9729 A side and another from the 9729 B side to the port on the SysPlex Timer attachment feature.

The same principle holds true for CLO links.

Figure 32 on page 52 shows an example configuration that links two 9037s together over an extended link. Two pairs of 9729s are used to provide multiple fiber paths. Of course, you could also carry ETR links across these trunk connections.

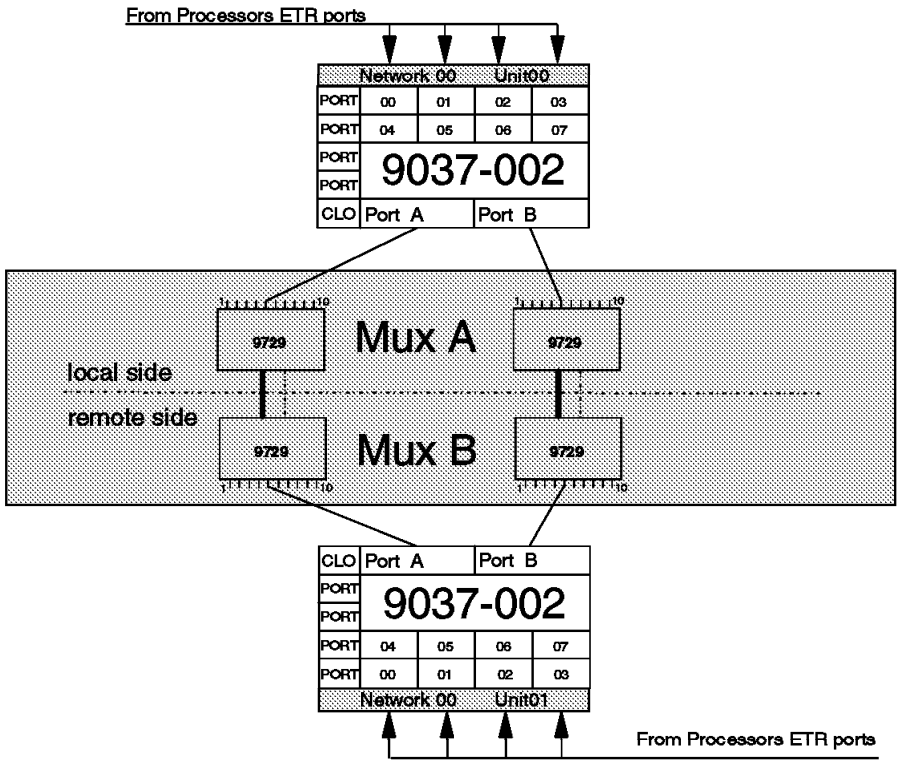


Figure 32. Connection of the CLO Links Using the IBM 9729

3.4.3 SysPlex Timer Unit Configurations

The IBM 9037 is available in three configurations:

- Expanded Availability configuration
- Expanded Basic configuration
- Basic configuration

Parallel Sysplex Availability Recommendation

The recommended configuration in a Parallel Sysplex environment is the 9037 *expanded availability* configuration. This configuration is fault-tolerant to single failures and minimizes the possibility that a failure can cause a loss of time synchronization information to the attached CPCs or CPC sides.

APAR OW19728 documents how Sysplex Timer failures affect OS/390 or MVS images in a sysplex. It also explains the effect on OS/390 or MVS IPL, if ETR signals are not present. In a multisystem sysplex, any OS/390 or MVS image that loses ETR signals is placed into a non-restartable wait state.

3.4.3.1 Expanded Availability Configuration

The 9037 *expanded availability* configuration consists of two 9037s. The 9037-002 expanded availability configuration is shown in Figure 33, and the 9037-001 expanded availability is shown in Figure 34 on page 54. Differences between the 9037-001 and the 9037-002 SysPlex Timers are described in 3.4.2, “9037 Sysplex Timer” on page 50.

In an expanded availability configuration, the TOD clocks in the two 9037s are synchronized using the hardware on the Control Link Oscillator (CLO) card and the CLO links between the 9037s. Both 9037s are simultaneously transmitting the same time synchronization information to all attached CPCs. The connections between 9037 units are duplicated to provide redundancy, and critical information is exchanged between the two 9037s every 1.048576 second (Mμs), so that if one of the 9037 units fails, the other will continue transmitting to the attached CPCs.

Redundant fiber optic cables are used to connect each 9037 to the same SysPlex Timer attachment feature for each CPC or CPC side. Each SysPlex Timer attachment feature consists of two ports, the active or stepping port and the alternate port. If the CPC hardware detects the stepping port to be non-operational, it forces an automatic port switchover and now the TOD steps to signals received from the alternate port. This switchover takes place without disrupting CPC operation. Note that the 9037s do not switch over, and are unaware of the port change at the CPC end.

Note: For an effective fault tolerant expanded availability configuration, Port 0 and Port 1 of the SysPlex Timer attachment feature in each CPC must be connected to different 9037 units.

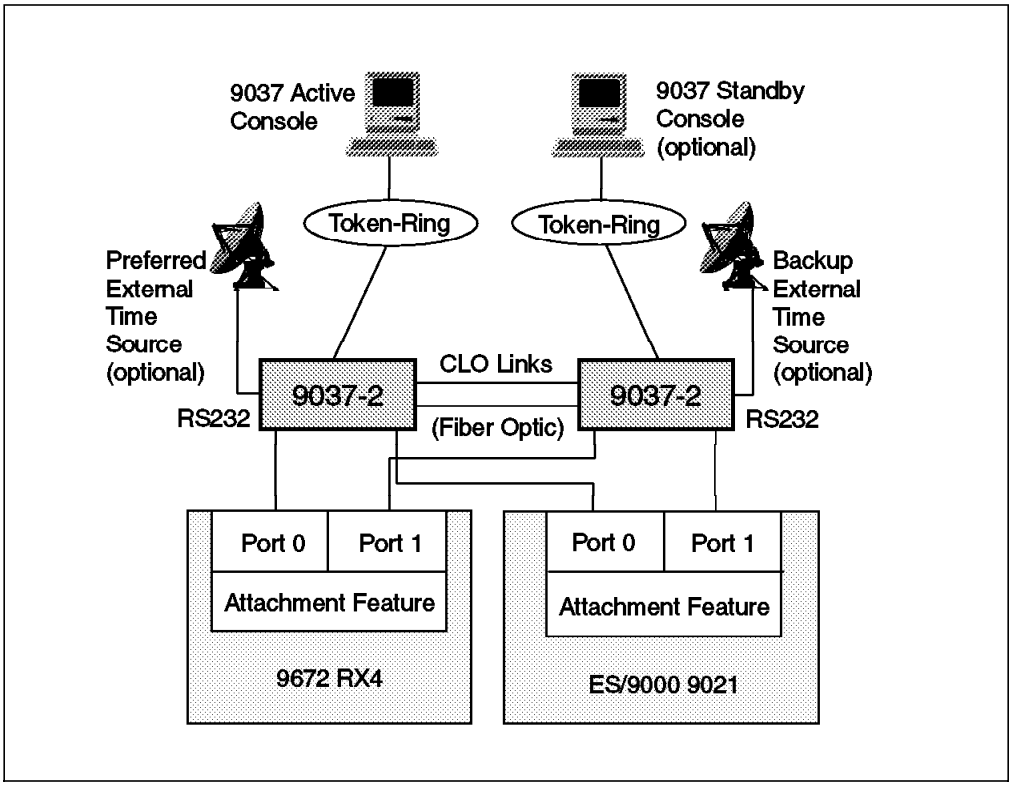


Figure 33. 9037-002 Expanded Availability Configuration

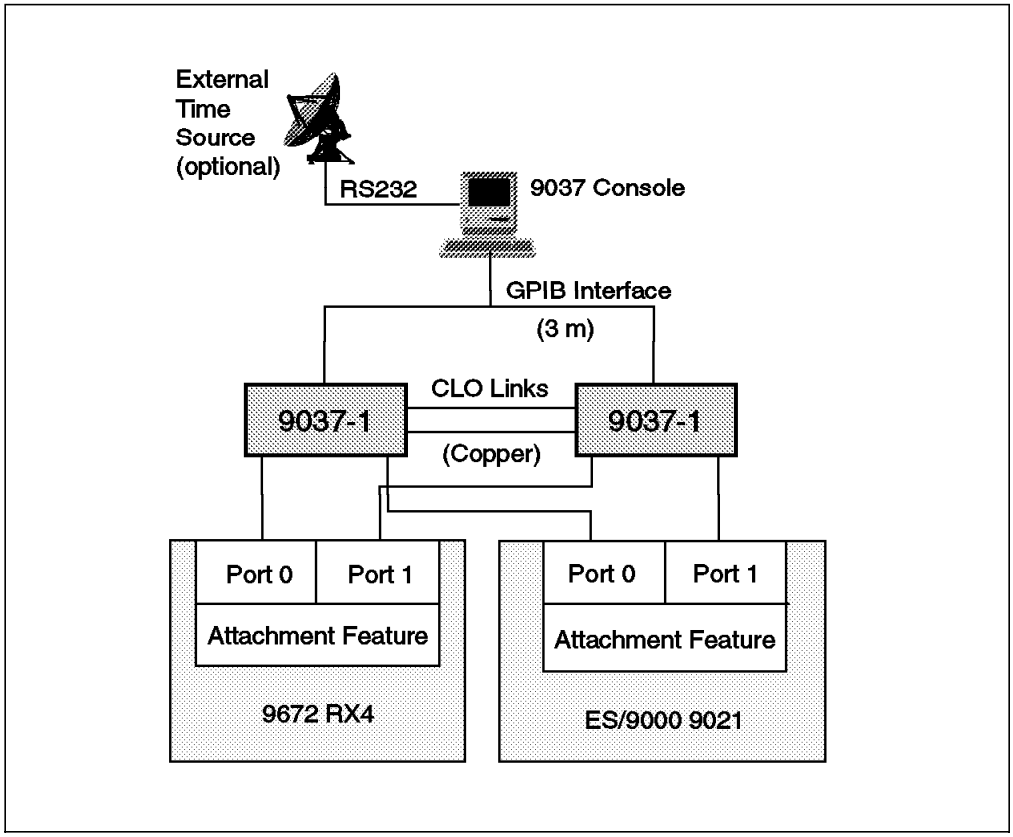


Figure 34. 9037-001 Expanded Availability Configuration

The following notes apply to the SysPlex Timer expanded availability configuration:

Notes:

1. To ensure data integrity among the different MVS images participating in a Parallel Sysplex, the 9037-2 design ensures that if at any time the two 9037-2s lose the capability to synchronize with each other, at least one of the 9037-2 units disables transmission of signals to the CPCs. To implement this rule, the 9037-2s in an expanded availability configuration arbitrate, at IPL time, which unit is primary and which is secondary. If the units cannot synchronize, for example, if both CLO links are accidentally severed, the primary 9037-2 will continue to transmit but the secondary 9037-2 will disable transmission. When the units are synchronized, both transmit the same time synchronization information to all attached CPCs. There is no concept of timer switch over.
2. During a power outage, a 9037-2 unit has sufficient internal capacitance to transmit a special Offline Sequence (OLS) over the CLO link to the other 9037-2. The OLS indicates that its transmissions to the CPCs will soon be disabled. If the 9037-2 unit in the primary data center has a power outage, receipt of the OLS signal by the 9037-2 unit in the secondary data center is critical to allow it to become the primary and continue transmission to the CPCs. If the OLS signal is not received, the 9037-2 unit in the secondary data center will also disable transmission of signals. This will place all MVS images in a non-restartable wait state.
3. The Timer-to-Timer connection was *never* meant to be a one-link connection between the CLO cards. In fact, if there is only one link available and the

9037-002 License Internal Code (LIC) release is at LVL 1.2, the OS/390 will receive an IEA272I message every eight hours until the second link is available.

3.4.4 S/390 Server Sysplex Timer Attachment Features

This section describes the attachment feature considerations required to connect a S/390 Server to a Sysplex Timer.

Two different types of Sysplex Timer attachment features can be used for the 9672 G4, G3, R3, R2, R1, and 2003 (all models). They are:

- Dual port card
- Master card (not available on 9672 R1 models)

Dual Port Card: This card, shown in Figure 35, is recommended for most configurations, since it does not have the same limitations of using the master card. The card has two fiber optic ports, each of which is attached to a SysPlex Timer using a fiber optic cable. In an expanded availability, each fiber optic port should be attached to a different SysPlex Timer in the same ETR network. This ensures that redundant SysPlex Timer paths are made available to all attached CPCs. This card does not have any output ports to redistribute the 9037 signals.

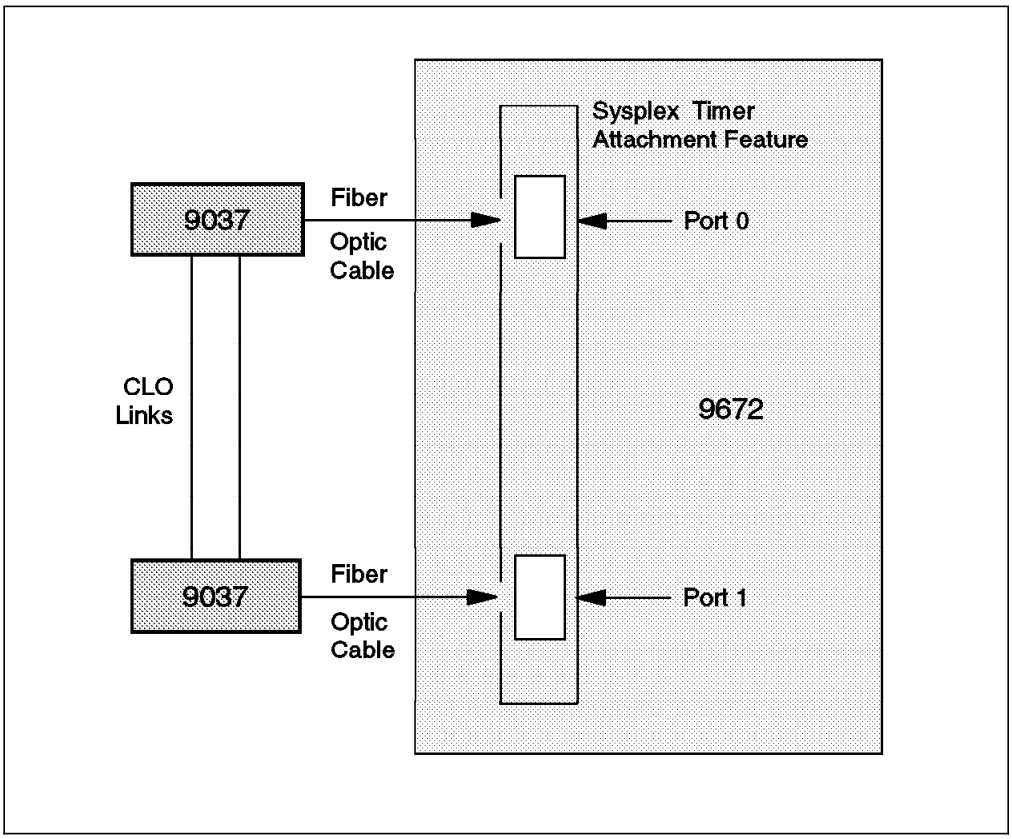


Figure 35. Dual Port Attachment to 9037 Sysplex Timers

Master Card: The master card should be used only in configurations that require a SysPlex Timer port capability in excess of 24 ports. The master card allows you to extend the physical 24-port limitation of the 9037-002 or the 16-port limitation of the 9037-001, by allowing you to attach a 9037-002 to a maximum of 48 S/390 server SysPlex Timer attachment feature ports or a 9037-001 to a maximum of 32 S/390 server SysPlex Timer attachment feature ports.

Each master card has two input ports and one output port. The *master input port* is a fiber optic port that is attached to a SysPlex Timer using a fiber optic cable. The *slave input port* is an electrical port, which receives redriven timer signals from the other master card's output port. The *master output port* distributes timer signals to the *other master card* through a 25-foot external cable. Refer to Figure 36. In an expanded availability, each master card's master input port should be attached to a different SysPlex Timer unit in the same ETR network. This ensures that redundant SysPlex Timer paths are made available to a pair of CPCs.

The following limitations of the master card must be carefully considered before selecting it for use:

- The pair of CPCs must be physically located adjacent to each other, since the external cable is only 25 feet long.
- Whenever one of the CPCs is powered off, the remaining CPC has only one connection to a SysPlex Timer.

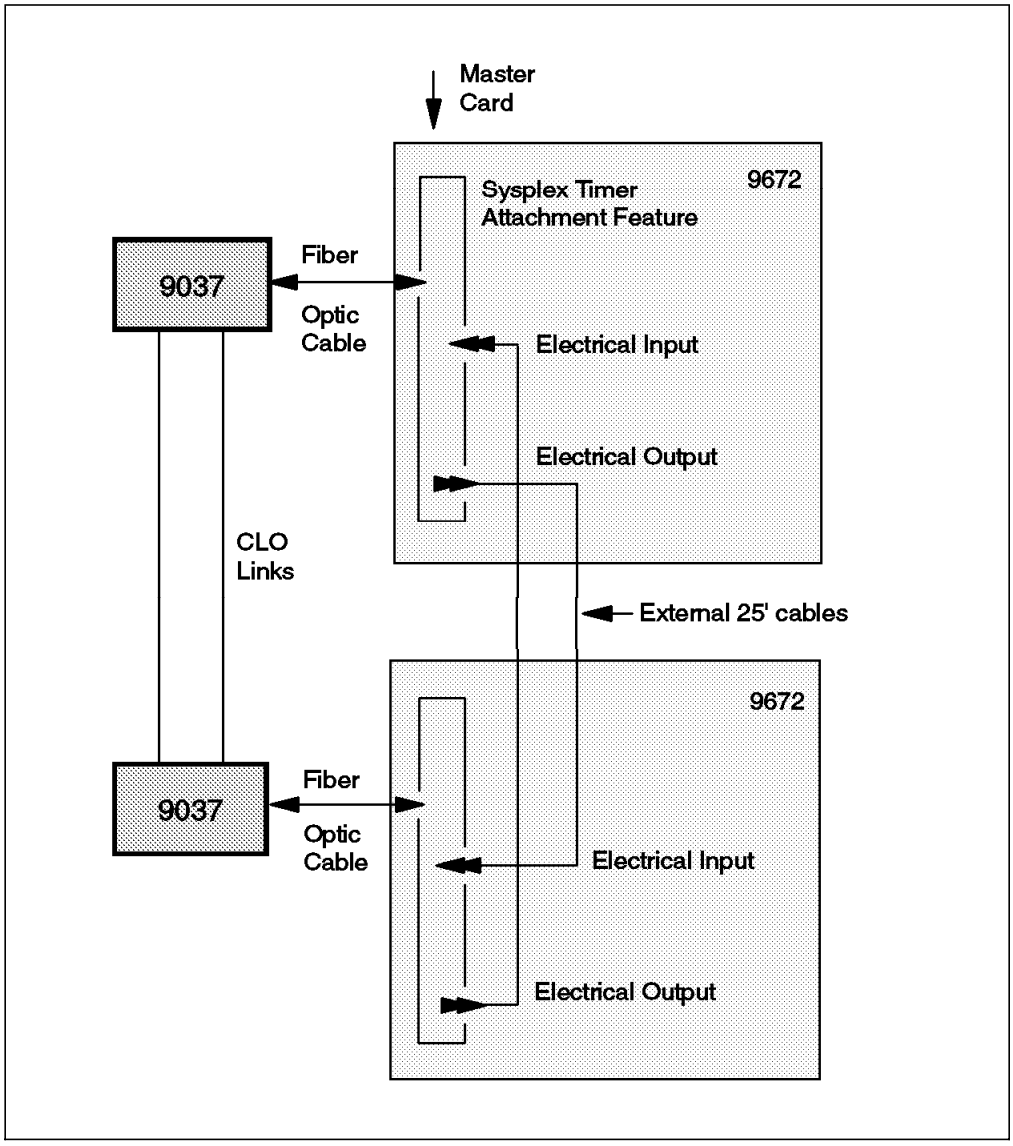


Figure 36. Example of 2 9672 CPCs with Master Cards and External Cables

An example configuration is shown in Figure 37 on page 57. This configuration is fault-tolerant to single failures and minimizes the possibility of a loss of time synchronization to the attached CPCs or CPC sides. Figure 37 on page 57 also shows two pairs of 9729s deployed. This spreads the ETR and CLO links across multiple fiber paths to provide the best availability.

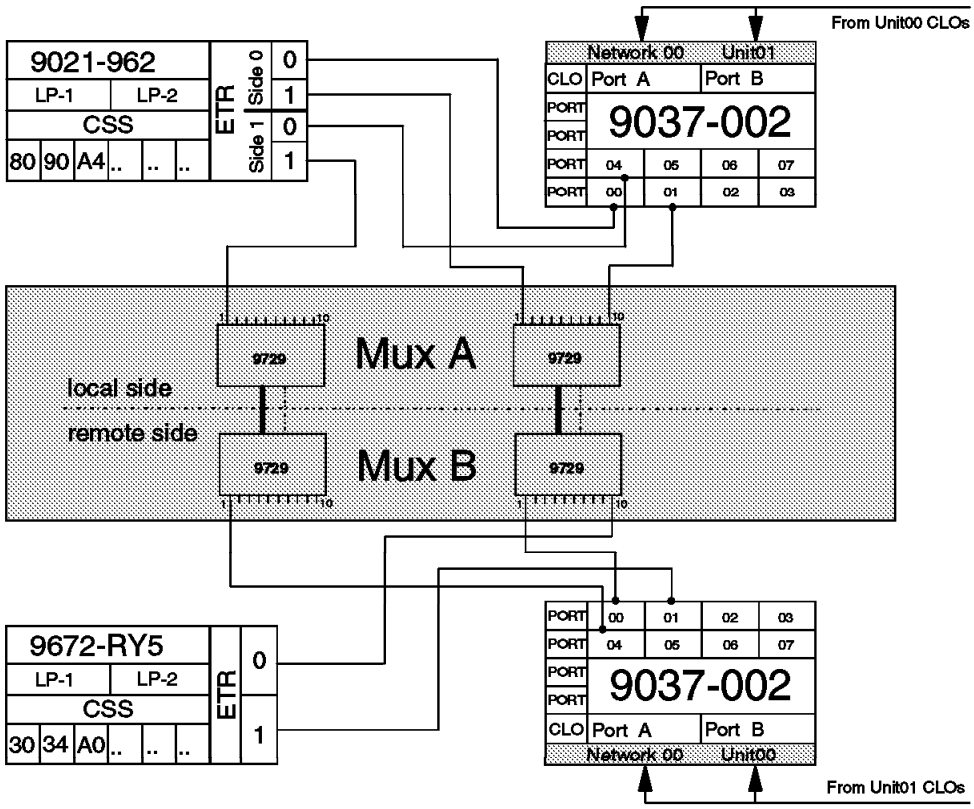


Figure 37. 9037 in Expanded Availability Configuration

3.4.5 Propagation Delays and the 9037

The 9037 automatically compensates for propagation delay through the fiber optic cables, thus allowing each connection from the 9037 to the CPC to be of different length. It is important to note that for the effects of different fiber cable lengths to be nearly transparent to processor TOD clock synchronization, the difference in length between the pair of individual fibers that make up a connection *must be less than 10 m*. Only then can the 9037 perform an effective propagation delay compensation. This specification for the allowable difference in length applies to both the connection between the 9037 and the CPCs and to the connection between SysPlex Timer units in an expanded availability configuration. Maintaining a difference of less than 10 m is not a problem with duplex-to-duplex jumper cables, where the fibers are the same length. However, this is an important consideration when laying trunk cables and using distribution panels.

This mechanism works for both CLO and ETR links. Therefore, the effects of extending CLO and ETR links are almost transparent to the processor's TOD clock synchronization.

Important Note

In order for this mechanism to work, you must ensure that the length of the individual transmit and receive fibers on a given link one are within 10 meters. This is not a problem with duplex-to-duplex jumper cables where the fibers are the same length. However, this is an important consideration when laying trunk cables and using distribution panels.

3.4.6 Extending Existing ETR and CLO Links with 9729s

In this section, we discuss the procedures necessary to relocate a 9037-002 in an expanded availability configuration.[21] For this scenario, we make the assumption that we have an existing Parallel Sysplex that we want to separate geographically.

We make the additional assumptions that we have two 9037s in our Parallel Sysplex and we want to place one 9037 in each site. This will require moving one of them to the new remote site. We have identified all the equipment that will be moved to the remote site. The existing site will be labeled the primary site while the new remote site will be labeled as the secondary.

The steps in our scenario are as follows:

1. First, we need to check that things are operating normally before the secondary site can be powered off and moved to the new location. We need to check:

   a. The CLO links

   b. Every connected ETR port

   You can very easily check the status of the CLO links from the Sysplex Timer console. Figure 38 on page 59 shows the console status screen. An UP arrow on a CLO port indicates that the corresponding CLO link is operational.

[21] Extending a CLO link with a 9037-001s is not possible, since both units have to be located *close* to each other. With the 9037-001, you can only connect an additional system to the existing ETR network.

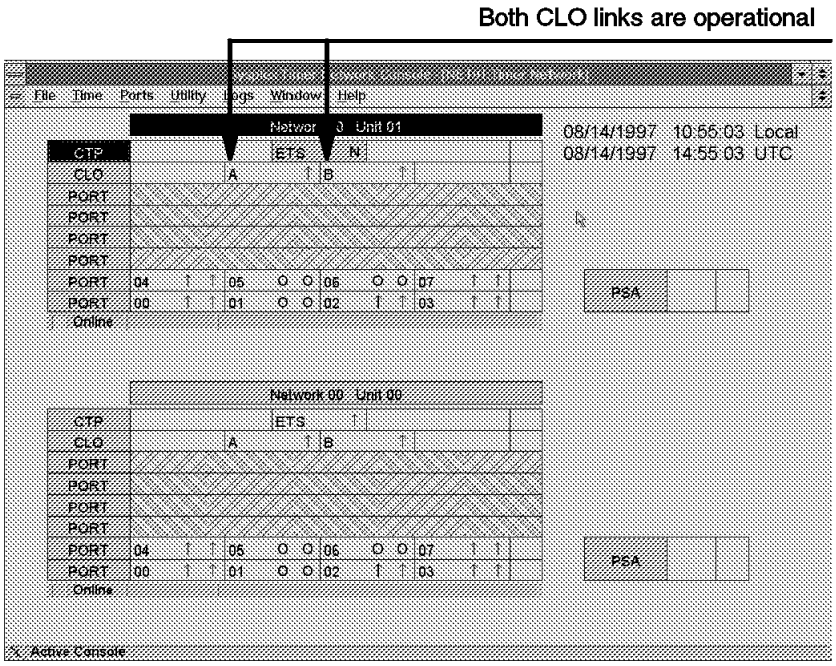


Figure 38. Sysplex Timer Console Status Window

2. Next, we must check the status of all MVS images that run in the ETR mode.[22] These images have to have their ETR ports operational. The verification can be performed at the MVS console by using the following MVS command for each MVS image:

   D ETR,DATA

   Figure 39 shows an example of the output of this command.

```
D ETR,DATA


IEA282I 13.54.57 ETR STATUS 873
SYNCHRONIZATION MODE = ETR    CPC SIDE = 0
  CPC PORT 0   <== ACTIVE       CPC PORT 1
  OPERATIONAL                   OPERATIONAL
  ENABLED                       ENABLED
  ETR NET ID=04                 ETR NET ID=04
  ETR PORT=11                   ETR PORT=04
  ETR ID=01                     ETR ID=00
```

Figure 39. Displaying MVS ETR Link Status

[22] The clock mode is defined in the CLOCKxx member of the SYS1.PARMLIB.

As can be seen in this figure, this CPC is connected to an ETR network with an ID=04. The active port 0 of the CPC is physically cabled to Sysplex Timer Unit 01, port 11. The backup port 1 from the CPC is within the same network, but connected to unit 00, port 04. The active ETR attachment feature shown above is installed at CPC side 0.

Note: It is very important that both CPC ports are connected to different ETR IDs and the backup path *must* show ENABLED and OPERATIONAL.

3. If all MVS ETR links and the CLO links are operational, you are now ready to disable all ports from the unit that must be moved.
4. When all ports have been disabled, this unit will automatically go in an OFFLINE state. It can then be cabled out and moved to the remote site.
5. After the physical installation of the unit is completed at the remote site, the bring up of the 9037 is completed in the reverse order of the steps outlined above.

Cable Routing Consideration

For maximum availability, each duplexed CLO/ETR link should be routed across separate fiber paths. In addition, the CLO links should be routed across separate paths from the ETR links. The greater the separation, the greater the probability that a single failure will not affect both links.

We therefore recommend that two pairs of 9729s be deployed as shown in Figure 37 on page 57. This way, if the CLO/ETR links are accidentally severed, the 9037-2 in the primary data center will still be able to transmit to the CPCs located in the secondary or remote data center.

3.4.7 Error Recovery on ETR Links

As discussed in 2.3.1, “Dual Fiber Switching Feature” on page 17, the 9729 has an optional Dual Fiber Switching Feature that electronically switches from one trunk fiber to another if the active trunk fails for any reason (for example, if the cable is cut). The switchover occurs in less then 2 seconds from the time the active fiber becomes non-operational.

The following scenario simulates a failure of the main trunk and shows the error recovery of the Sysplex Timer units. The recovery occurs without loosing any partition of the Sysplex.

We assume we have a configuration like the one shown in Figure 37 on page 57. We have two pairs of 9729 units and we have two independent fiber trunks between our 9729s.

1. First, display the status of the ETR network to ensure that all the links are operating. To do this, issue the following MVS command:

   ```
   D ETR,DATA
   ```

   Figure 40 on page 61 shows an example of this command. From the figure, you can see that CPC port 0 is the active one.

```
97226 13:54:57.42  D ETR,DATA

97226 13:54:57.60  IEA282I 13.54.57 ETR STATUS 873
                   SYNCHRONIZATION MODE = ETR    CPC SIDE = 0
                     CPC PORT 0   <== ACTIVE       CPC PORT 1
                     OPERATIONAL                   OPERATIONAL
                     ENABLED                       ENABLED
                     ETR NET ID=04                 ETR NET ID=04
                     ETR PORT=11                   ETR PORT=04
                     ETR ID=01                     ETR ID=00
```

Figure 40. Displaying the Status of an ETR Network

2. Next, check the Sysplex Timer Network console to see that the ETR links are enabled and operational. Figure 41 shows an example screen from this console.[23]

   An up arrow on an ETR port is an indication that the ETR port is enabled. Two up arrows indicate that the port is enabled *and* operational.

   Note: It is not possible to determine which port is active from the Sysplex Timer Network Console. This can only be done via the MVS command shown previously.

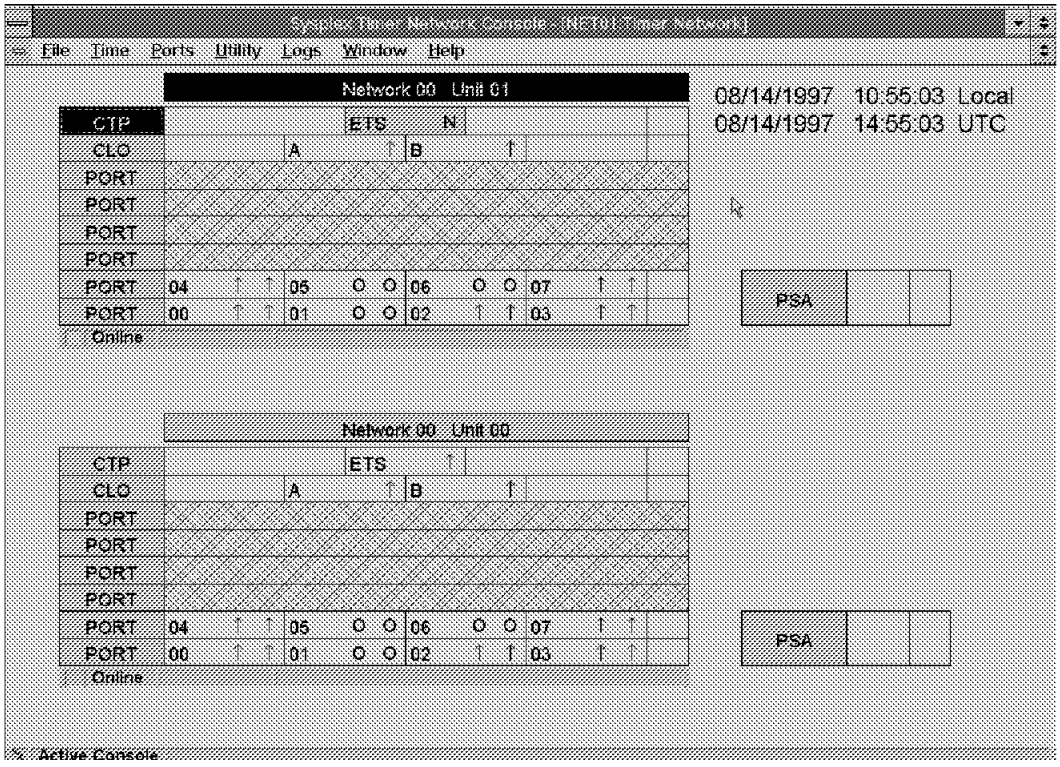


Figure 41. Sysplex Timer Network Console

[23] The appearance of the console is different for 9037 Model 001 and 9037 Model 002. Figure 41 shows the 9037 Model 2 console.

If you have IBM 9672 processors installed, you can check ETR port status from the point of view of the Sysplex Timer Attachment Feature (the CPC′s perspective). However, as with the Sysplex Timer Network Console, you can only check that the link is connected and enabled. It does not show you which link is the active one.

Figure 42 is an example of the status screen for a Dual-Port Attachment feature.[24]

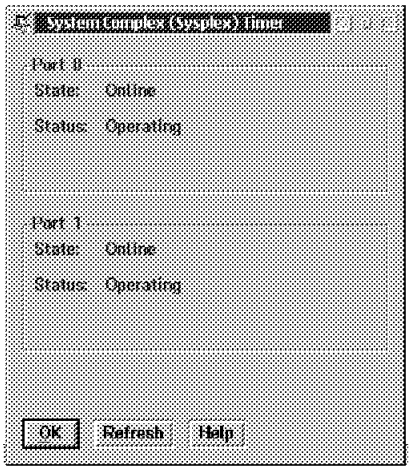


Figure 42. Port Status from 9672 Hardware Management Console (HMC)

3. Now, we simulate a failure of the primary trunk. (In this case, we unplugged the fiber.)

   At this time the active ETR link switch from port 0 to port 1 of the CPC side. The MVS images are not impacted, as the clock is now driven from the secondary 9037 unit in the same network. Also, assuming that the CLO links were split across the two 9729 trunks, one of them will become inactive. However, the 9037s remain in-sync through the other CLO link that is still operational.

   Also at this time, on 9672 processors, the Hardware Message icon starts blinking because a Sysplex Timer problem has been reported. Clicking on the icon will present a window that looks like the one in Figure 43 on page 63.

[24] This function is not available on all processors. It depends on the installed microcode level.

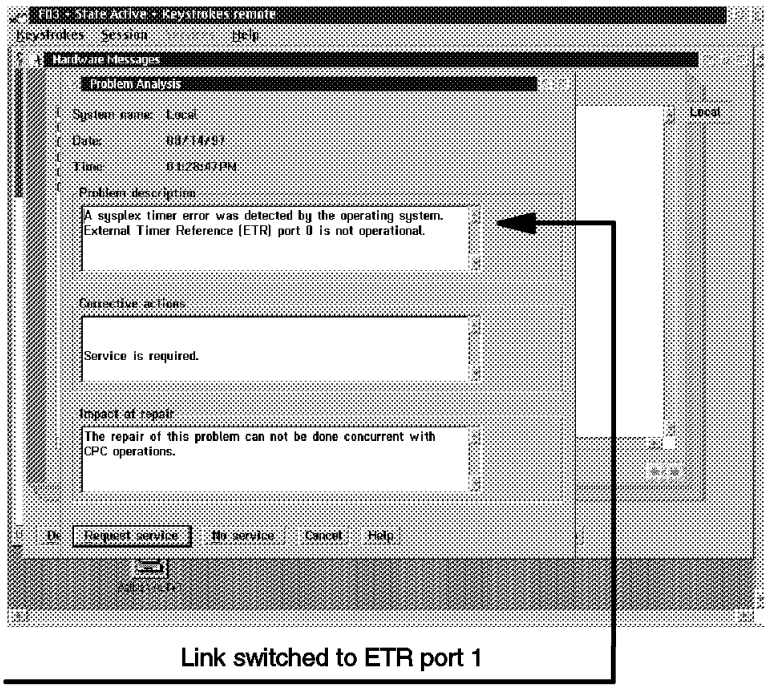


Figure 43. Hardware Error Message

4. During this time, the 9729s have switched to the alternate fiber and have re-established a viable link between them. The 9037s will now automatically re-initialize both the CLO link and the ETR links that were being transported over the 9729 trunk.
5. At this point, all links are operational again. Figure 44 shows the status of the 9672 CPC Sysplex Timer Attachment feature and indicates that both ports are online and operational.

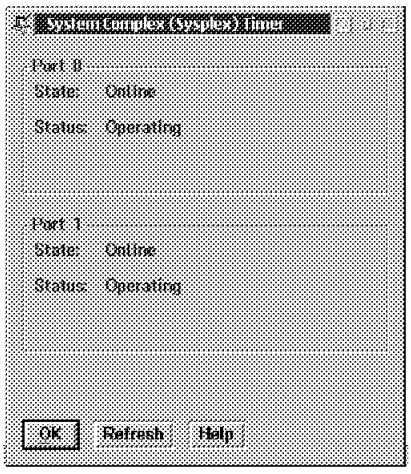


Figure 44. Ports Reflect an Operational State

However, the active CPC port has been changed from the original configuration and the CPC is now using the alternate port as the primary

one. Figure 45 on page 64 shows the active link is now CPC port 1 driven from 9037 unit 00 ETR port 04.

Note: This will not change back to its original state. Further, there is no way to toggle between the active port and the backup port using operator or console functions. The only way to force it back is to initiate a error on the active link.

```
97226 13:57:34.11  D ETR,DATA
97226 13:57:34.35  IEA282I 13.57.34 ETR STATUS 927
                   SYNCHRONIZATION MODE = ETR    CPC SIDE = 0
                     CPC PORT 0       ACTIVE ==>   CPC PORT 1
                     OPERATIONAL                   OPERATIONAL
                     ENABLED                       ENABLED
                     ETR NET ID=04                 ETR NET ID=04
                     ETR PORT=11                   ETR PORT=04
                     ETR ID=01                     ETR ID=00
```

Figure 45. Active Link Switched to Port 1

3.4.8 Problem Determination on ETR Links

The problem determination is the same as for ESCON links. (Please refer to 3.3.6, “Problem Determination on ESCON Links” on page 47.)

However, because the links are duplexed, you must first use the previously described procedures to determine that the links you think should be active are really active and enabled on the Sysplex Timer unit. Remember that an up arrow in the port cell on the Sysplex Timer Network console is an indication for an enabled port.

3.5 Using the 9729 with Coupling Facility Links

The IBM 9729 can be used in a Parallel Sysplex configuration to extend the distance used for coupling facility (CF) links. Coupling facilities are required when creating a Parallel Sysplex.

3.5.1 Coupling Facilities

A coupling (CF) is a specialized Central Processing Complex (CPC) which provides services for all systems in a Parallel Sysplex. Some of these services include common memory and messaging support.

CFs in a multi-CPC production sysplex can be implemented via one or more of the following elements:

- Logical partition on IBM 9674 C0x coupling facility (This is the recommended choice for production systems.)
- Logical partition on 9672-Rxx
- Logical partition on IBM 9021 711-based CPCs with SEC 228270 or higher

CFs run the coupling facility control code (CFCC). The CFCC has the following characteristics:

- CFCC runs only under LPAR. LPAR is mandatory because of the need of some LPAR functions, such as recovery and access to the hardware management console (HMC). CFCC does not have these functions itself. The CPs serving the CF logical partition can be shared or dedicated.
- CFCC is loaded at logical partition activation from the support element (SE) hard disk for IBM 967x CPCs or from the processor controller element (PCE) for IBM 9021 CPCs. LPARs are specified in the IOCDS using either HCD or IOCP directly.
- The major CFCC functions are:
  - Storage management
  - Dispatcher (with MP support)
  - Support for CF links
  - Hardware Management Console (HMC)) (Some console functions are provided by LPAR.)
  - Trace, logout, and recovery functions
  - Model code that provides the list, cache, and lock structure support
- CFCC operates in a continuous wait loop searching for work. This is also called *active wait*.
- CFCC is not interrupt driven. For example:
  - There are no inbound CF interrupts. The CF receiver link does not generate interrupts to CF logical CPs.
  - There are no outbound interrupts. The completion of asynchronous CF operations is detected by MVS using a polling technique.

Areas of the CF storage called structures are allocated for the specific use of CF exploiters. There are three types of structures:

Lock The lock structure is used for serialization of data with high granularity. For example, locks are used by IRLM for IMS/DB and DB2 databases, by CICS for VSAM RLS, and by GRS star for its global RSA information.

Cache Cache is used for storing data and maintaining local buffer pool coherency information. For example, caches are used by RACF database, DB2 databases, VSAM and OSAM databases for IMS, and by CICS/VSAM RLS. Caches contain both directory entries and data entries.

List Lists are used for shared queues and shared status information. For example, lists are used by VTAM/GR, JES2, MVS system logger, JES2 checkpoint data set, tape data sharing, CICS temporary storage, and XCF group members for signalling.

A coupling facility uses two types of channels: sending (CFS) and receiving (CFR). These are defined through the IOCP. The MVS image is always the

sender side and can be shared between the logical partitions while the CFR always belongs to the CF side and must be dedicated.[25]

CF sender links can be shared between PR/SM partitions in the same sysplex connected to a CF. Receiver links cannot be shared between CF LPARs.

Note: There are three possible options for the signalling between systems in a Parallel Sysplex:

- CF structures only
- CTCs only
- A combination of CFs and CTC structures

For low XCF rates of activity, there is little performance differences between CTC and CF signaling paths. At high rates performance is improved by using CTCs. XCF will always use the fastest path for signalling when there is a choice. The recommendation therefore is to configure both CF structures and CTCs and allow XCF to determine the best path.

3.5.1.1 CF Links

A CF link is a high-speed fiber-optic link between a coupling facility and an MVS image in the sysplex.

A CF link on both IBM 9674 Coupling Facilities and IBM 9672 processors is physically implemented the same way: via an Intersystem Channel (ISC) adapter and a coupling link card. The ISC adapter is a motherboard that has two slots on it for coupling link cards. The coupling link cards contain the fiber interface itself. The Intersystem Channel Adapter is not hot pluggable, but the coupling links are.

There are currently two types of coupling link cards available:

- Multimode Coupling Link: Uses a 50 micron multimode fiber optic cable. The link uses a short wave laser light source and can support distances of up to 1 kilometer between the coupling facility and attached systems at a speed of 50 MBps.
- Single-Mode Coupling Link: Uses a 9 micron single-mode fiber optic cable. This link uses a long wave laser light source and can support distances of up to 3 kilometers at a speed of 100 MBps.[26]

The multimode link is not extendable with the IBM 9729. The single-mode link is extendable using 9729s. You can increase the total distance up to 26 km: 3 km from each CPC side to the 9729 units with a 20 km 9729 trunk. Therefore, we recommend that all new and additional links should be ordered as single-mode. There is no limitation on mixing single and multimode links in the same Parallel Sysplex (although both ends of any one CF link have to use the same mode).

---

[25] Both the 9672 and the 9021 711-based processors can build up LPARs which run the CFCC. That means that the channel could be defined in the IOCDS as a sender channel (CFS) or as a receiver channel (CFR), depending on whether the LPAR is defined to be an MVS image or a coupling facility.

[26] The link uses the IBM ISC 1.0625 Gbps interface.

Important Notes

Officially, the 9729 supports a maximum trunk distance of 20 km for ISC channels. However, there is an RPQ available that can extend the distance beyond the 20 km limit. The RPQ number is 8P1786. If you need to extend the trunk beyond the 20 km limit, order the RPQ and specify the details of the installation and the desired trunk distance. IBM will evaluate the request and give you an answer whether the extended distance can be supported.

Also, in testing ISC links over 9729 trunks, it was discovered that there needs to be at least a 4 dB loss on the trunk to prevent detector saturation on the 9729 LRC cards.

IBM will ship a 5 dB attenuator with 9729 pairs that are ordered with ISC features. There will be two attenuators, one per link, for boxes with the Dual Fiber Switch feature. At installation time, the IBM service representative will test the fiber and will install the attenuator if the link loss is less than 4 db.

The attenuator will be shipped in the same box as the product documentation. Since the 9729 transmits and receives on the same fiber, the attenuator is only needed on one end of the link. IBM has arbitrarily decided to ship it with the ″A″ Units of a 9729 pair.

The other main CF link recommendations for achieving the best availability and performance are:

1. Use dedicated links as much as possible.
2. Configure additional links for availability.

Important Note

Extended CF links experience some performance degradation due to propagation delay. Access to the CF takes approximately 10 microseconds longer for each 1 km of link distance. This has a varying effect on performance depending on the frequency and type of access. Please see 3.5.5, “Effect of Distance on CF Links” on page 71 for more information.

3.5.1.2 HiPerLinks

The S/390 High Performance Coupling Links (HiPerLinks) is a performance enhancement for coupling links that provides improved channel efficiency and response times when processing some coupling facility requests. With HiPerLinks, coupling facility link capacity is improved on average by up to twenty percent.

Applications that transfer significant amounts of large data blocks (greater than 4 KB) to/from the coupling facility, such as BatchPipePlex and certain DB2 batch jobs, may see significant improvements in overall elapsed times using HiPerLinks. Also, cross-system coupling facility (XCF) messaging, used by many OS/390 subsystems for inter-system communication, will see improved message delivery time when using XCF structures in the coupling facility.

Using HiPerLinks, message delivery time is comparable to that achieved by XCF using channel-to-channel (CTC) pathing. This allows you to replace XCF-managed CTCs with XCF communication via the coupling facility, reducing

systems management costs through simplified Parallel Sysplex system configuration management, and without compromising performance.

Physically, the HiPerLinks feature is an improved version of the ISC adapter. It implements some functions in hardware that were previously in microcode. As with the ISC adapter, the HiPerLink adapter can accommodate up to two coupling links.

The HiPerLinks feature was introduced in 1997 for the S/390 G3 Servers and S/390 Coupling Facility Model C04. A S/390 G3 Server or Coupling Facility C04 model can operate with either Intersystem Channel Adapter feature or the new HiPerLink feature. The Intersystem Channel Adapter and the HiPer feature are mutually exclusive features on G3 and C04 systems; only the HiPerLinks feature is allowed on G4 and C05 systems.

The IBM 9729 can be used to extend both ISC links and HiPerLinks at distances up to 26 km.

3.5.2 Coupling Facility Configurations

You can use multiple CF links to connect an MVS image to one or more coupling facilities. Multiple CF links (and multiple CFs) allow you to design shared data systems that are fault-tolerant, resulting in a highly available system.

Note: It is strongly recommended that an installation configure at least two CF links between each MVS image and the CF. The system will use all available links when they are available to optimize performance. If one link fails, the load is automatically re-distributed on the remaining links.

Figure 46 on page 69 shows an example Parallel Sysplex configuration that includes two CFs: one in an IBM 9674 coupling facility and another in an LPAR in an IBM 9021 processor. Both partitions run the CFCC microcode. Two pairs of 9729s are used to provide full redundancy for the CF links. Under normal operation, the CF links are balanced over different 9729 paths.

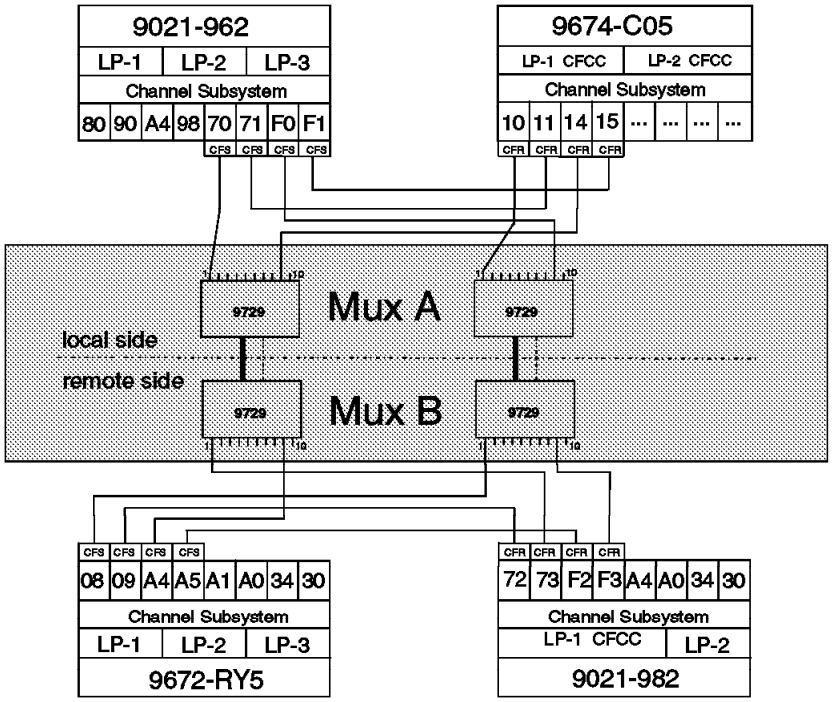


Figure 46. Coupling Facility Configuration Using Two Pairs of 9729s

Important Notes

Having MVS images and a CF in LPARs on the same CPC, while being a valid Parallel Sysplex configuration, is *not* a continuous availability configuration. For continuous availability, the coupling facility component should be physically isolated from the related MVS images; that is, in another LPAR *in another CPC*. Using 9674s is highly recommended for continuous availability configuration.

Also, the usual high-availability configuration principles apply to CF links: Configure one link from each side of an MP to a CF. At the CF, attach the links to different CF adapter cards in different IBB domains (in different cages on the 9674 if possible).

3.5.3 How Many CF Links Do I Need?

The number of CF links you need depends on the types of systems you have as well as the following factors:

- Number of Parallel Sysplexes you want
- Number of physically coupled CPCs
- Number of coupled LPARs
- Number of 9674s required for availability and performance reasons
- Number of extra CF links required for availability and performance reasons

The requests from an MVS image to a coupling facility will balance out over the number of CF links connected to a CF. Keep in mind that each CF link is made up of two subchannels. Each CF link can therefore have two active commands.

It is recommended that no more than 10 percent of the total requests be delayed because of a busy subchannel. This recommendation will usually allow quite high subchannel utilization when two links are configured to the CF from each image as recommended. You can find out the percent of delayed requests from the RMF Subchannel Activity Report. An example of this report is shown in Figure 47.

```
                                      C O U P L I N G   F A C I L I T Y   A C T I V I T Y
                                                                                                                              PAGE   4
        OS/390                SYSPLEX PLEXPERF           DATE 07/08/1997           INTERVAL 030.00.000
        REL. 01.02.00                 RPT VERSION 1.2.0          TIME 15.48.00             CYCLE 01.000 SECONDS

 ------------------------------------------------------------------------------------------------------------------------------
  COUPLING FACILITY NAME = CF7
 ------------------------------------------------------------------------------------------------------------------------------
                                                   SUBCHANNEL  ACTIVITY                          T
 ------------------------------------------------------------------------------------------------------------------------------
           # REQ                                  ----------- REQUESTS -----------   -------------- DELAYED REQUESTS --------------
  SYSTEM   TOTAL                 --BUSY--                  #   -SERVICE TIME(MIC)-          #     % OF  ------ AVG TIME(MIC) ------
  NAME     AVG/SEC -- CONFIG --  -COUNTS-                  REQ     AVG    STD_DEV          REQ    REQ   /DEL    STD_DEV     /ALL

  ECEC1     11229  SCH GEN    4  PTH    0     SYNC         0     0.0       0.0   SYNC     2   0.0%  20883       28788      0.0
              6.2  SCH USE    4  SCH    2     ASYNC     9240  5084.9     17158   ASYNC  177   1.9%   4609       28024     88.3
                   SCH MAX    4               CHANGED      0  INCLUDED IN ASYNC  TOTAL  179   1.9%
                   PTH        2               UNSUCC       0     0.0       0.0
  ECEC4     11812  SCH GEN    4  PTH    0     SYNC         0     0.0       0.0   SYNC     3   0.0%  13976       23367      0.0
              6.6  SCH USE    4  SCH    3     ASYNC     9792  4460.5     16572   ASYNC  422   4.3%   3319        3117    143.0
                   SCH MAX    4               CHANGED      0  INCLUDED IN ASYNC  TOTAL  425   4.3%
                   PTH        2               UNSUCC       0     0.0       0.0
  VCEC8     11864  SCH GEN    4  PTH    0     SYNC         0     0.0       0.0   SYNC     1   0.0%  442.0         0.0      0.0
              6.6  SCH USE    4  SCH    1     ASYNC     9848  4565.7     13711   ASYNC  344   3.5%   2441        3101     85.3
                   SCH MAX    4               CHANGED      0  INCLUDED IN ASYNC  TOTAL  345   3.5%
                   PTH        2               UNSUCC       0     0.0       0.0
  VCECA     12577  SCH GEN    4  PTH    0     SYNC         0     0.0       0.0   SYNC     3   0.0%  853.7       522.0      0.0
              7.0  SCH USE    4  SCH    3     ASYNC    10581  4026.7      5741   ASYNC  403   3.8%   2641        3111    100.6
                   SCH MAX    4               CHANGED      0  INCLUDED IN ASYNC  TOTAL  406   3.8%
                   PTH        2               UNSUCC       0     0.0       0.0
```

Figure 47. RMF CF Subchannel Activity Report

There is an easy rule that you can use to determine the approximate number of CF links that you need to configure on a sending CPC. The rule is based on the measurement of CPC power which is called Millions of Service Units (MSUs). A rough number for CF sizing is one CF link per 15 data sharing MSUs on an IBM CPC.

For example, assume that a 9672-RX4 is rated at 59 MSU. The rule above suggests that four CF links would yield adequate performance on the CPC. Assuming also that the Parallel Sysplex has two 9674s configured, we would configure two links to each 9674 in order to fully distribute the data sharing workload of the CPC.

After you have determined how many CF links you require, you can calculate the corresponding number of 9729 channels that will be required. In the above example, the four coupling facility links would also consume four channels on a pair of 9729s. However, in order to achieve the best system availability, we recommend that two pairs of IBM 9729s be used in order to put spread the links over different fiber paths.

Reminder

One standard length IBM fiber optic jumper cable is provided at no additional charge for the connection between two coupling link features. When you extend the CF link with a pair of 9729s, you will need to order another jumper cable for each CF link to be extended.

3.5.4 9729 CF Link Limitations

At the time of publication of this redbook, you can drive up to five CF links with one pair of 9729s. The 9729 Intersystem Channel (ISC) Input/Output Card (IOC) is used to interface to the coupling link card on both the coupling facility and the MVS side. Because the ISC IOC may only be installed in slots 1, 2, 3, 8, or 9, you can have a maximum of five ISC IOCs in one 9729. This limitation is imposed by the air flow through the 9729 as the ISC IOCs run very hot. The listed slots provide the most cooling since the fan assembly is installed nearest these slots. This restriction may be lifted in the future.

3.5.5 Effect of Distance on CF Links

The CF is accessed through a privileged instruction issued by an MVS component called cross system extended services (XES). The instruction refers to a subchannel. The instruction is executed in synchronous or asynchronous mode. These modes are described as follows:

Synchronous The CP in the MVS image waits for the completion of the request. Under some circumstances MVS will transform a synchronous request to execute asynchronously. For example, it is expected that the operation will take a long time to execute or it was detected that the operation took a certain time and did not complete. Locks are always executed synchronously.

Note: When a synchronous request is executed as asynchronously, it is called a *changed* request.

Asynchronous The CP in the MVS image issues the request, but the CPU does not wait for completion of the request. XES will either return a return code to the requestor, who may continue processing other work in parallel with the execution of the operation at the CF, or XES will suspend the requestor. Currently XES recognizes the completion of the asynchronous request through a dispatcher polling mechanism and notifies the requestor of the completion through the requested mechanism. The completion of the request can be communicated through vector bits in the hardware system area (HSA) of the CPC where the MVS is running. Lists are always executed asynchronously.

Synchronous accesses take longer as the CF link increases in length. This translates to a reduction in the Internal Throughput Rate (ITR) in the coupling facility sender (CFS) CPC. The operating system tries to balance synchronous and asynchronous requests. However, in general, there will be a reduction in the ITR as the CF link distance is increased. Table 3 on page 72 shows the maximum reduction in ITR for each kilometer of CF link distance for different CPCs. Please consider these as maximum values. The actual reduction would depend on application type and workload.

Table 3 on page 72 also shows the maximum reduction in ITR that a CPC would experience over a distance of 26 km (the maximum distance supported for CF links by the 9729).

Table 3. CFS Throughput Reductions Due to Propagation Delay

Sending CPC	% Reduction (%/km)	% Reduction at 26 km
9672-E/P/R1	0.2%	5.2%
9672-R2/R3	0.2 - 0.3%	5.2 - 7.8%
9672-R4	0.2 - 0.4%	5.2 - 10.4%
9021 711	0.7 - 0.9%	18.2 - 23.4%
Note: Link utilization will also increase with distance by 1-2% per kilometer.		

3.5.6 Extending Existing CF Links Using 9729s

A Parallel Sysplex allows Central Processing Complexes (CPCs) and Coupling Facilities (CFs) to be added or removed nondisruptively. Upgrades can be accommodated by taking a CPC/CF out of the Parallel Sysplex and upgrading it, while continuing to run the workload isolated on the remaining CPCs in the Sysplex. The upgraded CPC/CF can then be re-introduced to the Parallel Sysplex when the upgrade and associated testing is complete.

You can also use the same approach to build an entire remote data center and bring it online, moving your CPC(s) and CF(s) without disrupting your operations. Obviously, though, removing or adding a CF requires certain steps to be executed so that the removal or addition happens nondisruptively to the other work within the sysplex. There are both sysplex and multi-system application considerations which need to be taken into account.

This section details procedures for extending coupling facility links using IBM 9729s and the associated operations of shutting down and bringing up the affected coupling facilities.

3.5.6.1 Planning for a CF Shutdown

This section gives you some basic recommendations to follow when planning to move a coupling facility to a remote site. These should be used to prepare for when the CF system will be moved to the remote side of the 9729 link. (Refer to *OS/390 Parallel Sysplex Systems Management*, GC28-1861 for more details of this operation.)

Please consider the following as you plan to shutdown the coupling facility:

- Remove only one coupling facility at a time from the sysplex. Plan to shut down the CF during off-peak hours.
- To prevent any structures from being created in the CF that you are moving, define a new Coupling Facility Resource Management (CFRM) policy that does not include any reference to that coupling facility.
- Resolve any failed-persistent and no-connector structure conditions (if possible) before shutting down the CF.
- In an improved availability configuration, you have to move the structures to an alternate CF. Therefore, you have to make sure that all systems that use the structures to be moved have connectivity to the alternate coupling facility.

- Ensure that the alternate CF has enough capacity to allow the structures to be rebuilt. This includes storage, CF links, CP cycles, and structure IDs.
- If there is not enough capacity, you can possibly avoid a bottleneck situation by rebuilding XCF, JES2, and RACF structures:
  - For XCF you can use redundant paths via a CTC connection before you remove the signal paths through the CF.
  - The JES2 checkpoint could be placed on DASD. But take care to use an different DASD subsystem, if possible, if the other JES2 checkpoint is also located on DASD.
  - For RACF, use RACF in non-data sharing mode and delete the RACF structure on the CF to be removed.

3.5.6.2 Removing a CF from a Sysplex

If the above considerations have been met, you can remove the coupling facility from the Parallel Sysplex. The following step-by-step procedure shows you how to do it:

1. Create a new CFRM policy that does the following:
   - Removes any references to the CF to be shut down
   - Includes in the preference list the name of the CF where all structures are to be rebuilt

   Use different CFRM policy names so that the original policy can be restored after the CF system has been moved to the remote site.
2. Issue the following command on each system connected to the CF to determine the CHPIDs in use by the CF to be removed:

   ```
   D CF,CFNAME=cfname
   ```

   Figure 48 on page 74 shows an example of this command.

```
D CF,CFNAME=CF04

IXL150I  13.54.34  DISPLAY CF 852
COUPLING FACILITY 009674.IBM.02.000000045154
                  PARTITION: 2  CPCID: 00
                  CONTROL UNIT ID: FFFB
NAMED CF04
COUPLING FACILITY SPACE UTILIZATION
 ALLOCATED SPACE                DUMP SPACE UTILIZATION
  STRUCTURES:     133120 K        STRUCTURE DUMP TABLES:        0 K
  DUMP SPACE:      20224 K                 TABLE COUNT:        0
 FREE SPACE:      859648 K       FREE DUMP SPACE:           20224 K
TOTAL SPACE:     1012992 K      TOTAL DUMP SPACE:           20224 K
                               MAX REQUESTED DUMP SPACE:        0 K
   VOLATILE:         YES         STORAGE INCREMENT SIZE:      256 K
    CFLEVEL:           4

COUPLING FACILITY SPACE CONFIGURATION
                          IN USE          FREE         TOTAL
CONTROL SPACE:          153344 K      859648 K     1012992 K
NON-CONTROL SPACE:           0 K           0 K           0 K

SENDER PATH       PHYSICAL            LOGICAL
        25         ONLINE              ONLINE
        65         ONLINE              ONLINE
        A5         ONLINE              ONLINE
        DD         ONLINE              ONLINE

COUPLING FACILITY DEVICE     SUBCHANNEL     STATUS
                   FFE0         1A53        OPERATIONAL/IN USE
                   FFE1         1A54        OPERATIONAL/IN USE
                   FFE8         1A55        OPERATIONAL/IN USE
                   FFE9         1A56        OPERATIONAL/IN USE
                   FFF0         1A57        OPERATIONAL/IN USE
                   FFF1         1A58        OPERATIONAL/IN USE
                   FFF8         1A59        OPERATIONAL/IN USE
                   FFF9         1A5A        OPERATIONAL/IN USE
```

Figure 48. Determining the CHPIDs in Use by the Coupling Facility

Write down the following information regarding the CF to be removed:

- Node descriptor
- Partition (LPAR)
- CPCID
- CHPIDs (listed under the SENDER PATH)

You will need this information later when you take these CHPIDs offline.

3. Start the new CFRM policy by issuing the following command:

   `SETXCF START,POLICY,TYPE=CFRM,POLNAME=newpolicyname`

4. To obtain the names of all allocated structures in the CF that can be removed, issue the following command:

   `D XCF,CF,CFNAME=cfname`

   where `CFNAME` is the name of the CF to be removed.

   If there are no structures allocated in this CF, you will get the following message:

NO COUPLING FACILITIES MATCH THE SPECIFIED CRITERIA

NO STRUCTURES ARE IN USE BY THIS SYSPLEX IN THIS COUPLING FACILITY

The first message implies that the new CFRM policy is active. The second message implies that policy changes are pending and that there are no structures in the coupling facility.

If structures are allocated in the CF (you will get a message stating which ones are allocated), you must remove them from the coupling facility. Use the following procedure to remove them:

a. First, rebuild the structure with the following command:

   ```
   SETXCF START,REBUILD,STRNAME=strname,LOCATION=OTHER
   ```

   This might take several minutes. You will get a message if the rebuild was successful.

   Note: RACF does not support the LOCATION=OTHER parameter. However, RACF will rebuild the cache structure in the CF indicated by the CFRM policy preference list.

   Some structures do not support rebuild. These include:

   - List structure for the JES2 Checkpoint Data Set

     You have use the JES2 reconfiguration dialog to move the checkpoint data set to another coupling facility or to DASD.

   - DB2 Cache Structure for Group Buffer Pools

     You have to stop all DB2 members on the CF to be removed and to restart the members on the alternate CF, where the structures for the group buffer pools are allocated. (For more details, please refer to *DB2 V4 Data Sharing; Planning and Administration*, SC26-3269-01.)

b. Resolve failed-persistent connectors and no-connector conditions. To determine failed-persistent connections, issue the following command:

   ```
   D XCF,STR,STRNAME=ALL,STATUS=(FPCONN)
   ```

   To determine no-connector conditions, issue the following command:

   ```
   D XCF,STR,STRNAME=ALL,STATUS=(NOCONN)
   ```

   Generally, the function or application needs to be reinitialized to resolve these conditions. For information about resolving failed-persistent or no-connector conditions, see the documentation for the application.

c. Force the deletion of the structures.

   Important Note

   Use the FORCE option with caution. The option might cause a loss of structure data, so be sure you understand how the application is using the structure before you delete it.

   The FORCE option of the SETXCF command allows you to delete persistent structures or failed-persistent connectors in a CF.

   - To delete a structure with no connectors, issue the following:

     ```
     SETXCF FORCE,STR,STRNAME=strname
     ```

   - To delete a structure with *only* failed-persistent connectors, issue the following:

```
SETXCF FORCE,CON,STRNAME=strname,CONNAME=ALL
```

You must force the failed-persistent connectors *before* you force the structure. See *OS/390 V2R4.0 MVS System Commands*, GC28-1781-03 for more information.

5. Take the CHPIDs identified in step 2 offline. To do this, issue the following command on each system connected to the CF that you are removing:

   ```
   CF CHP(xx,yy),OFFLINE,UNCOND
   ```

6. To ensure that all CHPIDs were taken offline, issue:

   ```
   D CF
   ```

   on all systems connected to the coupling facility to be moved.

   Note: Identify the CF by the NODE DESCRIPTOR, PARTITION, and CPCID descriptions from the information you recorded in step 2.

   If the CHPIDs do not come offline using the UNCOND keyword, there are probably one or more structures still allocated in the CF. Recheck the previous steps to ensure that all structures and connectors have been removed from the coupling facility that you are shutting down.

   If a structure dump exists on the CF, you can issue the following command to force the deletion of the structure dump:

   ```
   SETXCF FORCE,STRDUMP,STRNAME=strname,STRDUMPID=strdumpid
   ```

   If you need to force the CHPIDs that remain online to the coupling facility, you can use the following command to force the CHPIDs to an offline state:

   ```
   CF CHP(xx,yy),OFF,FORCE
   ```

   Note: When you use the FORCE option on a system to take the last path to the CF offline, all active connectors lose connectivity to the structure.

7. Power off the coupling facility. When the CF is deactivated, any remaining structure data is lost.

8. Move the coupling facility to the remote site and complete the physical installation for that site.

9. Connect the coupling links to the 9729 ISC IOCs that have been reserved for these CF links.

   Note: Remember that only slot positions 1, 2, 3, 8 and 9 can be used to extend CF links.

10. Power on the coupling facility. After the coupling facility has finished the power on reset (POR), you first have to make sure that all CHPIDs are online and operating.

    You will obviously want to check to make sure that all CF links over the IBM 9729s are operating successfully. In case of problems with a link, refer to 3.5.8, “Problem Determination on CF Links” on page 84 to determine where the problem is located.

11. Restore the original CFRM policy.

12. Rebuild any structures that were temporarily rebuilt to an alternate coupling facility. To do this, issue the following command:

    ```
    SETXCF START,REBUILD,STRNAME=strname
    ```

    Note: The LOC=OTHER parameter may be needed depending on the CFRM policy structure preference list.

13. Restart any other subsystems that have been stopped during the coupling facility shutdown procedure.

    Note: When the coupling facility and CFRM policy are restored, some functions might reconnect to the CF automatically depending on the method used to remove the structures.

This completes the procedures for moving a coupling facility.

3.5.7 Error Recovery on CF Links

As discussed in 2.3.1, “Dual Fiber Switching Feature” on page 17, the 9729 has an optional Dual Fiber Switching feature that electronically switches from one trunk fiber to another if the active trunk fails for any reason (for example, if the cable is cut).

The following scenario simulates a failure of the main trunk and takes you step-by-step through the process of the failover from the perspective of the CF network.

Note: This example scenario shows only the rebuild of XCF structures and Global Resource Serialization (GRS) Lock structures. In the case of an actual failure, you could receive a lot more messages, depending on the structures that are defined.

The prerequisite for a good recovery is that at least two structures must be defined on separate CFs.

The scenario is as follows:

1. Display the working CF(s). To do this, use the command:

   ```
   D CF,CFNAME=cfname
   ```

   Figure 49 on page 78 shows this command on an MVS system connected to coupling facility CF03. The channel (sender path) is defined as CFS in the IOCDS.

```
97226 13:54:34.86  D CF,CFNAME=CF03

97226 13:54:34.95  IXL150I  13.54.34  DISPLAY CF 852
                   COUPLING FACILITY 009674.IBM.02.00000
                                     PARTITION: 1  CPCID
                                     CONTROL UNIT ID: FF
                   NAMED CF03
                   COUPLING FACILITY SPACE UTILIZATION
                    ALLOCATED SPACE                 DUMP SPACE UTILIZATION
                     STRUCTURES:     113664 K        STRUCTURE DUMP TABLES:        0 K
                     DUMP SPACE:      20224 K                TABLE COUNT:        0
                    FREE SPACE:      879104 K       FREE DUMP SPACE:           20224 K
                   TOTAL SPACE:     1012992 K      TOTAL DUMP SPACE:           20224 K
                                                  MAX REQUESTED DUMP SPACE:        0 K
                      VOLATILE:         YES         STORAGE INCREMENT SIZE:      256 K
                       CFLEVEL:           4

                   COUPLING FACILITY SPACE CONFIGURATION
                                            IN USE          FREE         TOTAL
                   CONTROL SPACE:          133888 K      879104 K     1012992 K
                   NON-CONTROL SPACE:           0 K           0 K           0 K

                   SENDER PATH       PHYSICAL            LOGICAL
                           05         ONLINE              ONLINE
                           45         ONLINE              ONLINE
                           85         ONLINE              ONLINE
                           C5         ONLINE              ONLINE

                   COUPLING FACILITY DEVICE     SUBCHANNEL     STATUS
                                      FFE4         1A63        OPERATIONAL/IN USE
                                      FFE5         1A64        OPERATIONAL/IN USE
                                      FFEC         1A65        OPERATIONAL/IN USE
                                      FFED         1A66        OPERATIONAL/IN USE
                                      FFF4         1A67        OPERATIONAL/IN USE
                                      FFF5         1A68        OPERATIONAL/IN USE
                                      FFFC         1A69        OPERATIONAL/IN USE
                                      FFFD         1A6A        OPERATIONAL/IN USE
```

Figure 49. Displaying the Status of a Coupling Facility

This command indicates that all four paths are physically and logically online. (In our scenario, we only have one connected MVS image.)

2. Display the GRS lock structures. GRS uses the contention detection and management capability of a lock structure to determine and assign ownership of a particular global resource. Each system only maintains a local copy of its own global resources. The GRS lock structure in the CF has the overall image of all system global resources in use.

   To display the GRS lock structures, issue the following command:

   ```
   D GRS
   ```

   Figure 50 on page 79 shows the result of this command in our scenario.

```
97226 13:54:48.98  D GRS

97226 13:54:49.08  ISG343I 13.54.48 GRS STATUS 858
                   SYSTEM    STATE               SYSTEM    STATE
                   SO0       CONNECTED           SO1       CONNECTED
                   SO2       CONNECTED           SO3       CONNECTED
                   SO4       CONNECTED           SO5       CONNECTED
                   SO6       CONNECTED           SO7       CONNECTED
                   GRS STAR MODE INFORMATION
                     LOCK STRUCTURE (ISGLOCK) CONTAINS 1048576 LOCKS.
```

Figure 50. GRS Star Lock Structure

Note: There is no specific recommendation for the placement of the GRS lock structure. Be aware, however, that if a system loses connectivity to the structure, and the structure is not rebuilt to a CF where the system does have connectivity, the system affected is put into a wait state.

3. Now, we simulate a failure of the primary trunk. (In this case, we unplugged the fiber.) At this point an error occurs on the single-mode fiber link between 9729 A-Side and 9729 B-Side. The failing link messages scroll over the MVS consoles and also several hardware messages appear on the Hardware Management Consoles (HMCs) of the connected 9674 and 9672 (if installed).

   After the primary link has dropped, the MVS console log shows an entry for every link which was driven by the affected 9729 pair. Figure 51 shows an example of the console log messages.

```
97226 13:55:58.11  IOS581E LINK FAILED REPORTING CHPID=45 901
                          INCIDENT UNIT  TM=009672/RX5 SER=IBM02-045445 IF=0045  IC=03
                          ATTACHED UNIT  TM=009674/C05 SER=IBM02-045154 IF=0045
97226 13:55:58.12  IOS581E LINK FAILED REPORTING CHPID=05 903
                          INCIDENT UNIT  TM=009672/RX5 SER=IBM02-045445 IF=0005  IC=03
                          ATTACHED UNIT  TM=009674/C05 SER=IBM02-045154 IF=0020
97226 13:55:41.73  IXL158I PATH 05 IS NOW NOT-OPERATIONAL TO CUID: FFFD 887
                                  COUPLING FACILITY 009674.IBM.02.000000045154
                                                    PARTITION: 1  CPCID: 00
97226 13:55:41.73  IXL158I PATH 45 IS NOW NOT-OPERATIONAL TO CUID: FFFD 888
                                  COUPLING FACILITY 009674.IBM.02.000000045154
                                                    PARTITION: 1  CPCID: 00
```

Figure 51. Error Messages in the MVS Console Log after a Trunk Failure

Both, the CF-Sender CPC and the CF-Receiver CPC will recognize the dropped link. Hardware messages get posted on each of them. Figure 52 on page 80 shows the CF Channel Information from the CF Sender (the MVS image) on the 967x HMC. It contains information about connection status, configuration details and a node descriptor from the attached side.

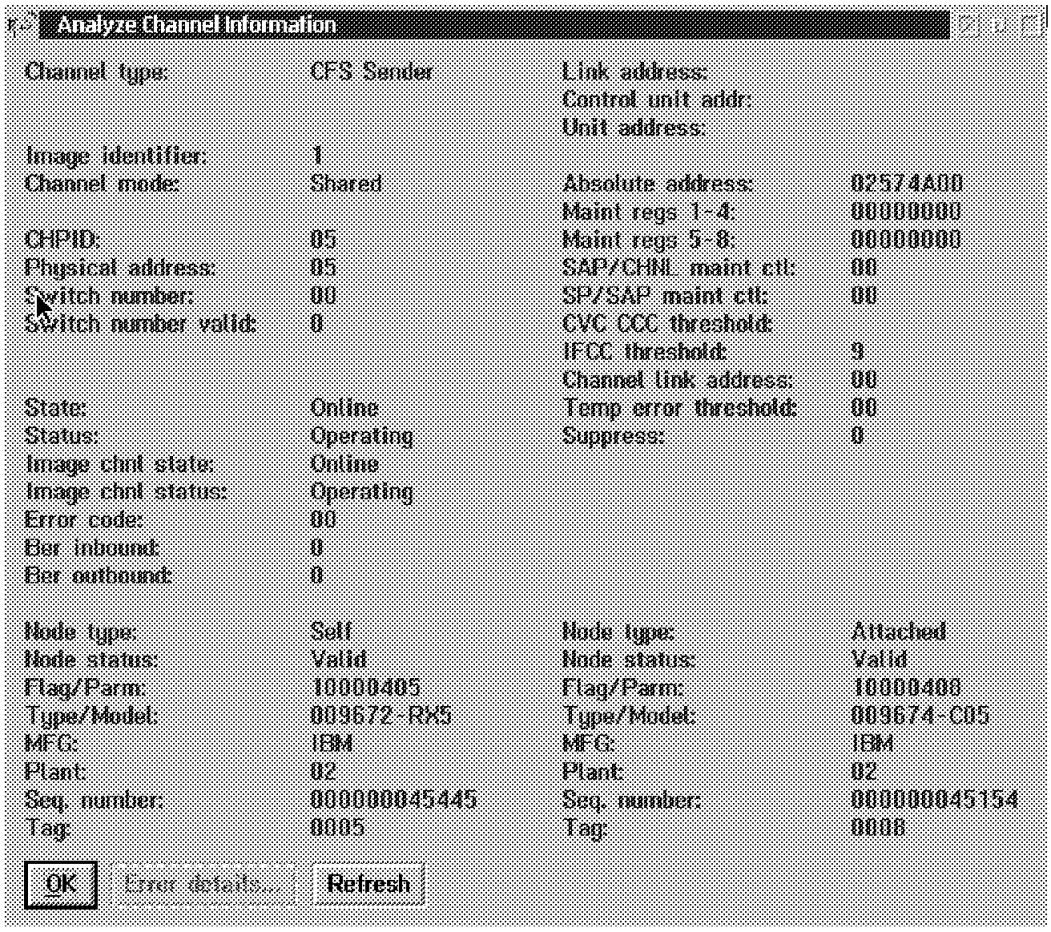


Figure 52. CF Channel Information from CF Sender

Figure 53 on page 81 shows the Problem Analysis window from a CF sender or receiver on 967x HMCs. This window appears when you click on the blinking **Optical Network** icon on the 967x HMC. It contains information about the node and the CHPIDs on the link that was broken.

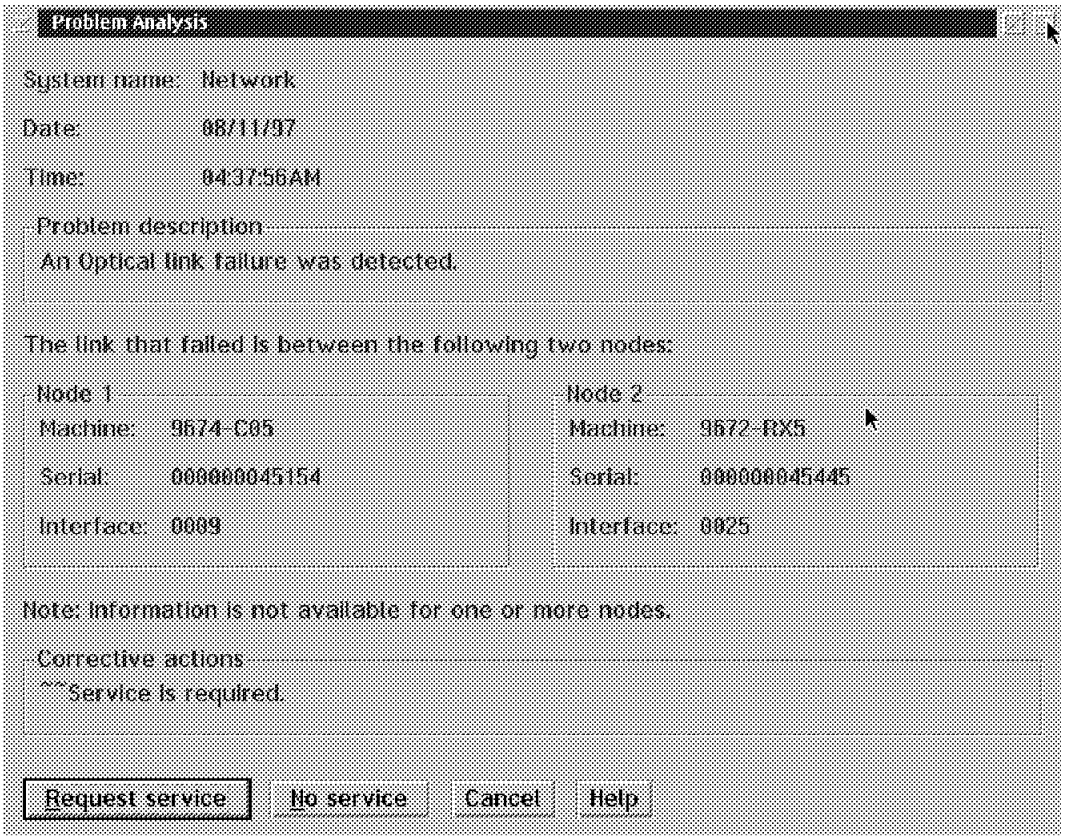


Figure 53. CF Channel Link Failure Information from CF Sender or Receiver

4. Display the XCF structure rebuild.

   Figure 54 shows the rebuild of the XCF structures.

   Note: XCF will always rebuild the structure for any connection failure irrespective of whether an active policy is in place or not.

```
97226 13:55:41.79  IXC467I RESTARTING PATHOUT STRUCTURE IXCSIG7 LIST 20 891
                           USED TO COMMUNICATE WITH SYSTEM SO4
                           RSN: I/O APPARENTLY STALLED
                           DIAG073: 08960208 0000DBED 0000DBE8 0000DBE9 0000DBE1
97226 13:55:41.88  IXC467I RESTARTING PATHOUT STRUCTURE IXCSIG4 LIST 8 892
                           USED TO COMMUNICATE WITH SYSTEM SO1
                           RSN: I/O APPARENTLY STALLED
                           DIAG073: 08960208 0000DAF1 0000DAF0 0000DAF1 0000DAE9


97226 13:55:42.94  IXC466I OUTBOUND SIGNAL CONNECTIVITY ESTABLISHED WITH SYSTEM SO1 896
                           VIA STRUCTURE IXCSIG4 LIST 8
97226 13:55:43.76  IXC466I OUTBOUND SIGNAL CONNECTIVITY ESTABLISHED WITH SYSTEM SO4 897
                           VIA STRUCTURE IXCSIG1 LIST 20
```

Figure 54. XCF Structure Rebuild

5. After approximately 1.5 seconds, messages are posted to indicate that the affected links are operational again. Figure 55 on page 82 shows these messages.

```
97226 13:55:58.26  IXL157I PATH 45 IS NOW OPERATIONAL TO CUID: FFFD 904
                                 COUPLING FACILITY 009674.IBM.02.000000045154
                                                   PARTITION: 1  CPCID: 00
97226 13:56:03.89  IXL157I PATH 05 IS NOW OPERATIONAL TO CUID: FFFD 908
                                 COUPLING FACILITY 009674.IBM.02.000000045154
                                                   PARTITION: 1  CPCID: 00
```

Figure 55. All CHPIDs Back in an Operational State

Figure 56 is a capture of an coupling facility operating system messages window. From here you have a direct interface to the coupling facility control code (CFCC), you can check the link status from the CF point of view.

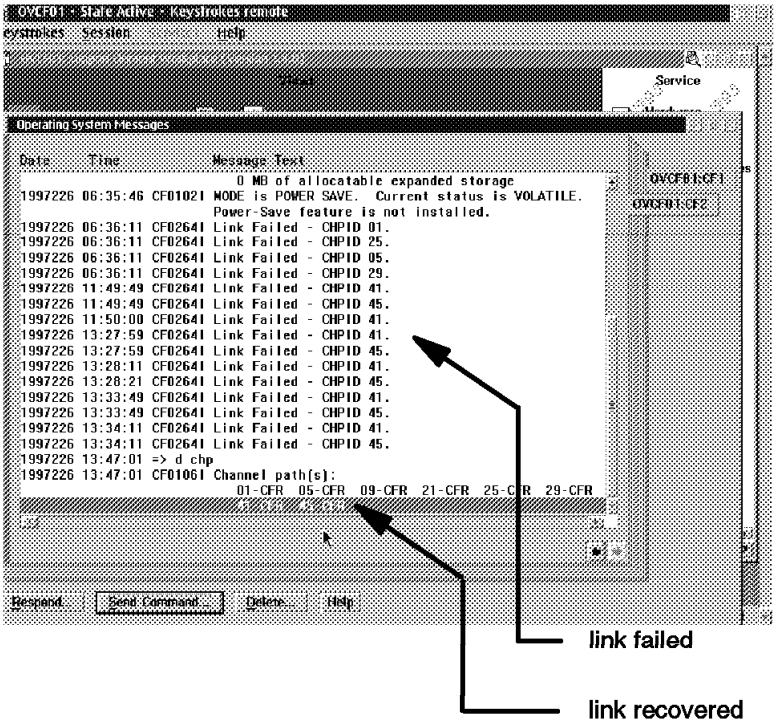


Figure 56. CF Channel Link Failure Information from CF Receiver

6. Display the status of the coupling facility to make sure that everything has returned to an operational state. Figure 57 on page 83 shows this command.

```
97226 13:56:27.51  D CF,CFNAME=CF03
97226 13:56:27.53  IXL150I  13.56.27  DISPLAY CF 913

                   COUPLING FACILITY 009674.IBM.02.000000045154
                                     PARTITION: 1  CPCID: 00
                                     CONTROL UNIT ID: FFFD
                   NAMED CF03
                   COUPLING FACILITY SPACE UTILIZATION
                    ALLOCATED SPACE                 DUMP SPACE UTILIZATION
                     STRUCTURES:     113664 K        STRUCTURE DUMP TABLES:        0 K
                     DUMP SPACE:      20224 K                TABLE COUNT:        0
                    FREE SPACE:      879104 K       FREE DUMP SPACE:           20224 K
                   TOTAL SPACE:     1012992 K      TOTAL DUMP SPACE:           20224 K
                                                  MAX REQUESTED DUMP SPACE:        0 K
                      VOLATILE:         YES         STORAGE INCREMENT SIZE:      256 K
                       CFLEVEL:           4

                   COUPLING FACILITY SPACE CONFIGURATION
                                           IN USE          FREE          TOTAL
                   CONTROL SPACE:          133888 K      879104 K     1012992 K
                   NON-CONTROL SPACE:           0 K           0 K           0 K

                   SENDER PATH       PHYSICAL            LOGICAL
                           05         ONLINE              ONLINE
                           45         ONLINE              ONLINE
                           85         ONLINE              ONLINE
                           C5         ONLINE              ONLINE

                   COUPLING FACILITY DEVICE     SUBCHANNEL     STATUS
                                      FFE4         1A63        OPERATIONAL/IN USE
                                      FFE5         1A64        OPERATIONAL/IN USE
                                      FFEC         1A65        OPERATIONAL/IN USE
                                      FFED         1A66        OPERATIONAL/IN USE
                                      FFF4         1A67        OPERATIONAL/IN USE
                                      FFF5         1A68        OPERATIONAL/IN USE
                                      FFFC         1A69        OPERATIONAL/IN USE
                                      FFFD         1A6A        OPERATIONAL/IN USE
```

Figure 57. All CHPIDs Back in an Operational State

7. Display the GRS lock structure again to make sure that all systems have been re-connected to the GRS Star. Figure 58 on page 84 shows this command.

```
97226 13:56:43.90  D GRS
97226 13:56:44.01  ISG343I 13.56.39 GRS STATUS 919
                   SYSTEM    STATE               SYSTEM    STATE
                   SO0       CONNECTED           SO1       CONNECTED
                   SO2       CONNECTED           SO3       CONNECTED
                   SO4       CONNECTED           SO5       CONNECTED
                   SO6       CONNECTED           SO7       CONNECTED
                   GRS STAR MODE INFORMATION
                     LOCK STRUCTURE (ISGLOCK) CONTAINS 1048576 LOCKS.
```

Figure 58. GRS Structures Rebuild

3.5.8 Problem Determination on CF Links

For the most part, problem determination on coupling facility links is the same as for ESCON links (refer to 3.3.6, “Problem Determination on ESCON Links” on page 47).

The ISC IOC has two LED indicators:

- A green light indicates the presence of light on the fiber.
- An amber light indicates a laser fault.
- If both the amber and the fault LEDs are on, it indicates the loss of signal at the far end of the fiber.

Also, in the Parallel Sysplex environment, you must be aware of XES periodic path monitoring.[27] When path monitoring receives certain kinds of link errors, XES retries those operations an excessive number of times resulting in an unnecessary CPU overhead.

To avoid unnecessary CPU overhead, we recommend that you take the failing CHPID offline for problem determination procedures or while you wait for service.

3.5.8.1 CF Link Incident Examples

If you are running a Parallel Sysplex with IBM 9021, 9121, 9672 processors or 9674 Coupling Facilities, link incidents may be generated for Coupling Facility/ISC channels when varying an MVS image or CPC out of the sysplex.

The following command removes a system out of the sysplex:

```
V XCF,sysname,OFFLINE
```

After issuing this command, you should see an MVS wait code with PSW=000A 0000 8000 40A2. Along with that you may see a link incident with an incident code of 04 (IC=04) generated for the CF/ISC channel that belongs to the image/CPC that has been varied out. Figure 59 on page 85 shows an example of this incident.

[27] Path monitoring and validation are performed on all physically configured CF links in order to ensure that the links are still viable and to detect when non-operational links return to an operational state.

```
D/T=IBM 9672,MOD=E03,SER=00xxxxx,IF=0010
IQ=09, IC=04, FLAGS=10, NODE PARAMETERS=000410
SAC=34, LINK FAILURE BETWEEN TWO NODES

DEV DEPENDENT 0-7=  0000000000000000  BYTES  8-15= 0000000000
      BYTES 16-23=  F4F8F9F900000000  BYTES 24-31= 0000000000
      BYTES 32-35=  00000000
ESCON ASSOCIATED DATA,2ND     FAILING GROUP:
D/T=IBM 9674,MOD=C01,SER=00yyyyy,IF=0013
IQ=00, IC=00, FLAGS=10, NODE PARAMETERS=000413
DEV DEPENDENT 0-7=  0000000000000000  BYTES  8-15= 0000000000
      BYTES 16-23=  0000000000000000  BYTES 24-31= 0000000000
      BYTES 32-35=  00000000
```

Figure 59. Link Failure after Varying System Out of Sysplex

If this link failure occurs after removing a system from the Parallel Sysplex and the incident code equals 04, then no service action is required.

If the incident code is anything other than 04, or there is a permanent loss of connectivity, then link problem determination should be performed.

If the system experiences BIT-ERROR-THRESHOLD (BER), it is an indication that the error rate between two nodes exceeded the defined threshold. Figure 60 shows an example of a BER with an IC=02.

```
ESCON ASSOCIATED DATA,PRIMARY FAILING GROUP:
D/T=IBM 9674,MOD=C01,SER=YYYYYYY,IF=0010
IQ=00, IC=00
SAC=31, HIGH ERROR RATE BETWEEN TWO NODES
DEV DEPENDENT 0-7= 0000000000000000
      BYTES  8-15= 0000000000000000
      BYTES 16-23= 0000000000000000
      BYTES 24-31= 0000000000000000
      BYTES 32-35= 00000000

ESCON ASSOCIATED DATA,2ND FAILING GROUP:
D/T=IBM 9672,MOD=R63,SER=XXXXXXX,IF=00A4
IQ=05, IC=02
DEV DEPENDENT 0-7= 0000000000000000
      BYTES  8-15= 0000000001F6F3C6
      BYTES 16-23= F4F8F9F900000000
      BYTES 24-31= 0000000000000000
      BYTES 32-35= 00000000
```

Figure 60. Link BER with an IC=02

The BIT-ERROR-THRESHOLD IC=02 indicates a problem with a transmitter. The problem could be on either side of the link. If 9729s are used, there are four possible transmitters that could be implicated since the 9729s add two more lasers to the equation. Also, keep in mind that the 9729s are transparent to the node descriptor.

We can use the BER incident report to isolate the problem. We should suspect the node *opposite* the node where the BER was reported. In Figure 60, this would be the 9674-C01 (or the 9729 unit on that side of the coupling facility link).

3.6 The 9729 in a Remote Copy Environment

IBM remote copy is a solution for fast and effective recovery in the event of an application site outage. Using IBM 3990 Storage Controllers, the remote copy function shadows data in real time to a remote site, guarantees the sequence of write updates at the recovery site, and provides easy-to-use recovery and monitoring support. Remote copy offers application performance protection, data currency options, and data independence.

The IBM 9729 can be used to extend the distances between the control units used in the remote copy configuration. This gives you much more flexibility in designing DASD subsystems and remote data centers.

3.6.1 Remote Copy Functional Overview

Basically, remote copy creates two mirrored DASD volumes. Only one volume is online to the operating system. The other one is synchronized via subsystem microcode. Remote copy performs this shadowing of the data from the application site to the recovery site using one of two methods: asynchronous (XRC) or synchronous (PPRC) copying.

1. Extended Remote Copy (XRC)

   XRC is an asynchronous implementation of remote copy. It allows application systems to update data at the application site in the usual way and then manages the process of passing the updates to the recovery site after they have completed at the application site. Asynchronous operation results in minimal performance impact to application systems at the application site. However, the currency of data at the recovery site usually lags slightly behind the currency of data at the application site because of updates in transit. XRC does ensure data integrity at the recovery site by applying secondary updates in the same sequence as they were applied at the application site across many storage controls and their attached devices.

2. Peer-to-Peer Remote Copy (PPRC)

   PPRC provides a synchronous data copying capability by sending updates directly from the application site storage control to the secondary storage control, thus maintaining data currency for each write operation In the event of an outage at the application site, there will be no data loss due to the data at the backup site being out of sync.

Table 4 shows you a comparison of XRC and PPRC. As you can see from the table, PPRC and XRC differ in their effect on DASD I/O performance, the degree of data currency at the time of a disaster, use of system resources, and operational control.

Table 4 (Page 1 of 2). Comparison of XRC and PPRC

Topic	Extended Remote Copy	Peer-to-Peer Remote Copy
Application independent	Yes	Yes
Type of solution	Hardware and software	Hardware and non-operating system S/W
Design priority	Minimize impact of performance	Data current at recovery site
Recovery	Entire session	Volume by volume
Recovery site data currency	All except bytes in transit	All data kept current
Channel type	ESCON or parallel	ESCON only for PPRC links

Table 4 (Page 2 of 2). Comparison of XRC and PPRC		
Topic	**Extended Remote Copy**	**Peer-to-Peer Remote Copy**
Supported Copy distance	ESCON or network distance	ESCON distance between 3990 Model 006s
Copy operation	Asynchronous	Synchronous
Application site storage control	3990 Model 006	3990 Model 006
Recovery site storage control	3990 Model 003 or 006, RAMAC Array Subsystem	3990 Model 006

The two approaches also differ in the amount of distance that can be supported between the application site and the backup site. The maximum distances supported for a PPRC configuration are the same as they are for other applications using ESCON channels. However, XRC sites can be separated by distances greater than those supported by ESCON.

Using IBM 9729s can be advantageous in both configurations. In this redbook, we focus on the PPRC configuration since the deployment of IBM 9729s in the XRC configuration is very similar to a normal ESCON configuration from the link point of view.

3.6.2 Remote Copy Components

The PPRC function is implemented in 3990 Model 006 DASD subsystem and its Licensed Internal Code (LIC). Additional components include the ESCON connections between the 3990 storage controllers at the application site and the recovery site.

3.6.2.1 Function of the 3990 Model 006

3990 Model 006s with the remote copy LIC are required at both the application and recovery sites.[28]

PPRC is implemented almost entirely in the 3990 Model 006 Licensed Internal Code (LIC). Software commands are available to initiate, monitor, and recover PPRC-managed data. Enhancements to the DASD error recovery procedure (ERP) offer a choice of PPRC configurations for disaster recovery.

Because the write to the recovery site controller is synchronous with the application system I/O, data is written to cache and Non-Volatile Storage (NVS) to allow *device end* status to be signaled as soon as possible. Therefore, it is important that adequate capacity planning be performed to evaluate performance based on NVS size and other recovery site 3990-6 storage control resources.

3.6.2.2 3990 Model 006 ESCON Links

The 3990 Model 006s use standard ESCON ports that connect via multimode fiber optic cables at distances up to 3 km. When required distances exceed 3 km, you have several options to extend the links:

- ESCON Directors (with standard multimode fiber ports)

[28] A Model 006 can be the application storage control for some volumes and the recovery storage control for other volumes at the same time.

This configuration can provide a maximum distance between 3990s of 9 km assuming 2 ESCDs with three 3 km multi-mode fiber links between the devices.

- ESCON Directors with XDF ports

  The maximum supported distance using the ESCD extended distance feature (XDF) ports is 20 km.

- IBM 9036 Channel Extenders

  The maximum supported distance using 9036s is also 20 km.

- IBM 9729s

  Using IBM 9729s, the current supported distance is 43 km. This distance can even be extended further via an RPQ up to 50 km.

Clearly, the 9729 provides the best method for extending the distance for PPRC configurations.

3.6.2.3 PPRC Volumes

There are two types of volumes that are used in a PPRC configuration:

- Primary

  Primary volumes are the ones located at the application site that are being shadowed to the recovery site under PPRC control. A primary volume can be copied to only one secondary volume. Primary volumes under PPRC control must be attached to a 3990-6.

- Secondary

  Secondary volumes are PPRC volumes located at the recovery site that contain shadowed primary volume data. The secondary volumes are physically protected from non-PPRC updates and should be offline to all connected hosts. Secondary volumes under PPRC control must be attached to a 3990-6.

A PPRC volume can be only a primary or secondary, not both at the same time.

3.6.2.4 PPRC Prerequisites

To establish an Peer-to-Peer Remote copy connection via the 9729 Optical Wavelength Division Multiplexer, the following prerequisites are necessary:

- **3990 Model 006 Prerequisites:**
  - PPRC requires that primary and secondary volumes located at the application site and recovery site, respectively, be attached to 3990-Model 006s.[29]
  - PPRC requires subsystem cache and NVS to be active on both the application site and recovery site 3990-006s.[30] In addition, caching and DASD Fast Write (DFW) should be enabled for the primary volume. The recovery site 3990 Model 006 must have DFW status enabled on the

[29] It is possible to set up PPRC pairs on volumes attached to the same 3990 Model 006 storage control. This setup is useful for both testing and device migration because PPRC supports copying to devices with the same or higher capacity.

[30] The PPRC CESTPAIR command to establish a PPRC pair will be rejected if the recovery site 3990′s NVS or cache is not active. The same is true for the application site. In this case, the command reject for the primary will be given reason code 0D for primary NVS or 0E for primary cache.

secondary volumes. Any automated procedures for establishing PPRC pairs should ensure that these subsystem functions are active.

- Each 3990 Model 006 must have the remote-copy-capable LIC.
- The PPRC storage subsystems must be connected over ESCON links. (This will usually be through ESCON Directors but direct connections are also supported.)

- **9729 Prerequisites:**

  For every PPRC link to be extended over a 9729 trunk, one pair of ETR/ESCON cards has to be installed (one in each 9729 unit). The cards must be installed in the same IOC slots of the 9729 A and B sides.

- **Software Prerequisites:**

  The TSO command functions for MVS/DFP 3.1.1 and above (supplied as PTFs) must be installed. These can be added dynamically to an MVS system and become active after a refresh using the command LINKLIB (F LLA,REFRESH). Optional error recovery procedures (DASD ERP) are supplied as PTFs to MVS/DFP 3.2.0 and above. These ERPs are strongly recommended for disaster recovery scenarios where operators or 3990 Model 006 storage controls may be unaware of volume status after the loss of the application site, or links between the sites.

3.6.3 PPRC Data Flow

A PPRC write operation between the storage control in the application site and the recovery site is synchronous with the primary volume's I/O operation. Figure 61 depicts a write operation to the primary volume and the sequence of events that occur in the remote copy.

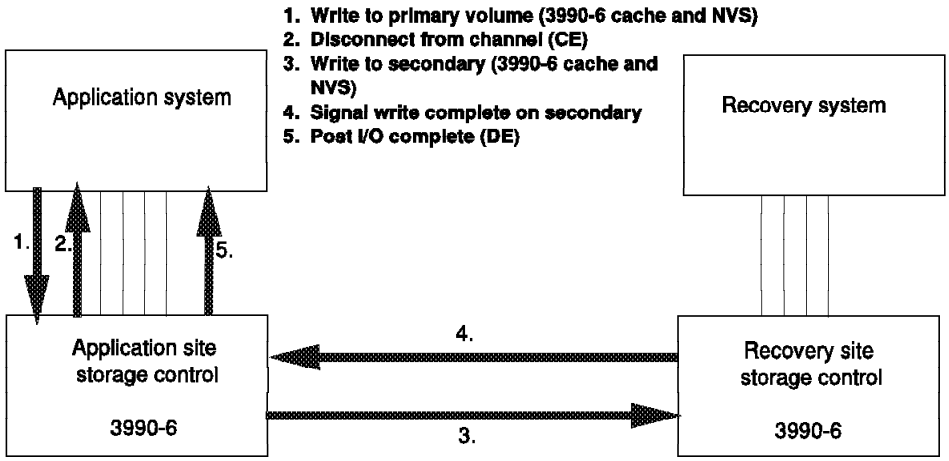


Figure 61. PPRC Data Flow

The sequence of events are outlined as follows:

1. Write to primary volume (3990 Model 006 cache and NVS)

   The application system writes data to a primary volume on a 3990 Model 006 subsystem.

2. Disconnect from channel with Channel End (CE)

Once data has been transferred to the 3990′s cache and NVS, the 3990 sends back channel end status to the application system.

3. Write to the recovery site storage control (3990 cache and NVS)

   The application site initiates an I/O channel program to copy the data to the recovery site 3990. The PPRC copy function does not consider the application system DASD write operation complete until the data has been written to the recovery system. Hence, the performance of all writes will be degraded, since the application must wait until the data is stored at the recovery site. To minimize this impact, a PPRC write operation only transfers the data to the recovery site 3990′s cache and NVS.

   The additional time necessary to perform this synchronous write is estimated to be on the order of 4 to 6 ms which is about the equivalent of a DFW operation.[31]

   Immediately after the application writes to a volume that is part of a PPRC pair, the application site 3990 tries to locate a path to the recovery site storage control.[32] Thus, channel utilization and 3990 storage path resources also have to be considered.

   Note:

   For maximum availability, the paths between storage controllers should be routed across two pairs of IBM 9729s.

   The application site 3990 cache size can be increased to improve workload read performance. This additional cache may not benefit the write content of a workload, but it can offer faster response times for reads, which in turn can result in lower overall averages. The Cache Analysis Aid program can assist in calculating the new hit rates that are likely to result from larger cache sizes.

4. Signal write complete on recovery site storage control

   The recovery site storage control unit signals that the write is complete to the application site 3990.

   Performance of PPRC write operations to the recovery site storage control may be impacted by other operations that use the recovery site 3990 for access to non-PPRC volumes. These operations use the 3990 storage paths and add to the subsystem load. At some point, utilization of the recovery site 3990 from non-PPRC secondary volumes accesses may cause an application site 3990 to experience delay when writing to the which in recovery site, which in turn will delay completion of the primary application I/O.

   Ensure that the configurations do not exceed the capacity of the recovery site storage control. The recovery site 3990 must be configured with adequate NVS such that the PPRC writes can be satisfied as quickly as possible. The recovery site 3990 will guarantee a DFW hit for the secondary volume as long as there is enough NVS.

---

[31] This estimate assumes a 4 KB record update on an ESCON channel within normal machine room distances of up to 100 m.

[32] This connection works in a way similar to dynamic path reconnect (DPR) for normal 3990 Model 006 operations. The application site storage control will connect to the first available path, on any available 3990 system adapter.

Note

The NVS size at the recovery site 3990 should be equal to the sum of all of the application sites' 3990 NVS size.

5. Post I/O complete

   On completion of the transfer, the application site 3990 returns a device end to the operating system.

3.6.4 PPRC Availability Configuration

The choice of hardware configuration to support PPRC depends on a number of factors that will be unique to your installation. At a minimum, the 3990 Model 006s must have ESCON capability and the 9729s must have ETR/ESCON IOC cards to take part in a PPRC configuration. The following recommendations should also be adhered to:

- Balance the PPRC links over the cluster.

  The 3990 Model 006 has two clusters: Cluster 0 and Cluster 1. Each cluster should contain the same amount of System Adapter (SA) cards.

- Balance the PPRC links over the SA cards.

  One cluster can have either 2, 4, or 8 ESCON ports. The number of ESCON ports will affect how many paths the 3990s will be able to use for host and PPRC communication. A 3990 Model 006 with only four ESCON ports (two 2-port SA cards) will have to be upgraded to provide ports for PPRC use or be connected to an ESCON Director so that existing ports can be used for both host connection and PPRC paths.

- Balance the PPRC links over the IBM 9729 trunks.

  One pair of 9729s can have up to 10 full-duplex ESCON channels installed. Therefore we can drive up to 10 links with one pair. However, it is strongly recommended to install two pairs of 9729s to avoid a single point of failure for the PPRC links.

- Balance the links through the ESCON Director.

  The ESCON Director Models 002 and 003 have a base of 28 ports. On a single card, the four ports have consecutive addresses. For maximum resilience to a single four-port card failure, spread the connections for host, PPRC, and controller across multiple cards.[33] For highest availability, use the ESCON Director Model 003. If suitably configured, this model has no single points of failure and facilitates concurrent upgrades and maintenance. If an installation already has ESCON Director Model 002s, the PPRC paths should be evenly distributed across the available directors, if possible.

3.6.4.1 PPRC Configuration without ESCON Directors

This non-switched PPRC configuration has two 3990 Model 006s, each with eight ESCON ports (two 4-port SA cards). The ESCON channels are used to connect the application and recovery host processors to the application and recovery site 3990 Model 006 subsystem, respectively. The PPRC paths are defined using the dedicated fiber links between the application and recovery site 3990 Model 006s.

[33] As a consequence, the ports for PPRC links will not have consecutive addresses.

Because PPRC operations are a synchronous process, it is recommended that all four 3990 interfaces be defined to have a path to the recovery site storage control unit even though you might share these paths across only two physical ESCON connections. Defining the maximum number of PPRC paths in this way avoids potential contention for 3990 paths.

3.6.4.2 PPRC Configuration with ESCON Directors

When compared to non-switched configurations, PPRC ESCON Director configurations offer several advantages in terms of flexibility and resource sharing. The ESCON Director permits economical sharing of physical resources by translating the destination addresses in ESCON frames and switching director connections to route the data to the appropriate address. Figure 62 shows a PPRC configuration using two ESCDs.

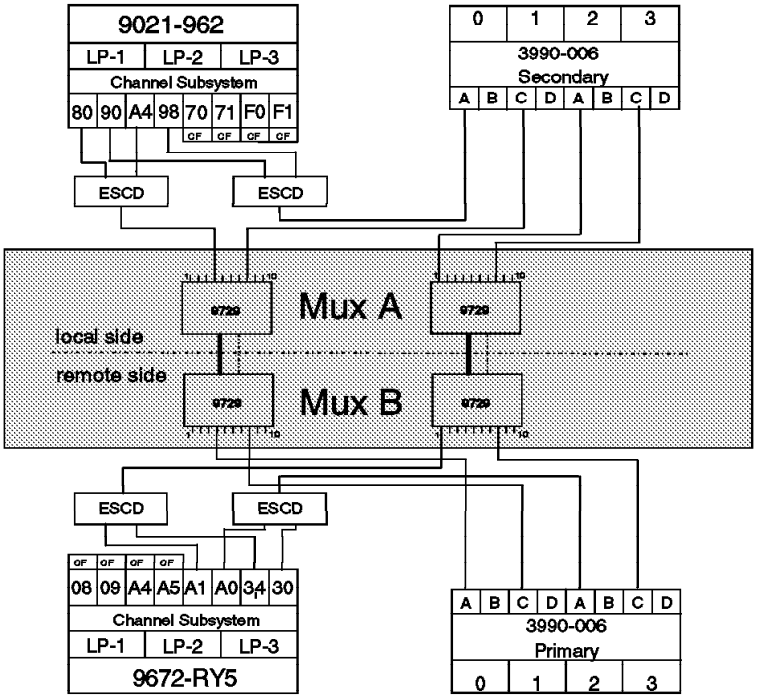


Figure 62. PPRC Configuration Using Two Pairs of 9729s

While Figure 62 shows only two PPRC links connected to each other for better illustration, we recommend using four, which will provide better performance and availability.

The 9729 Optical Wavelength Division Multiplexer is transparent to the links, so there is no need for any configuration changes when 9729s are installed between ESCDs.

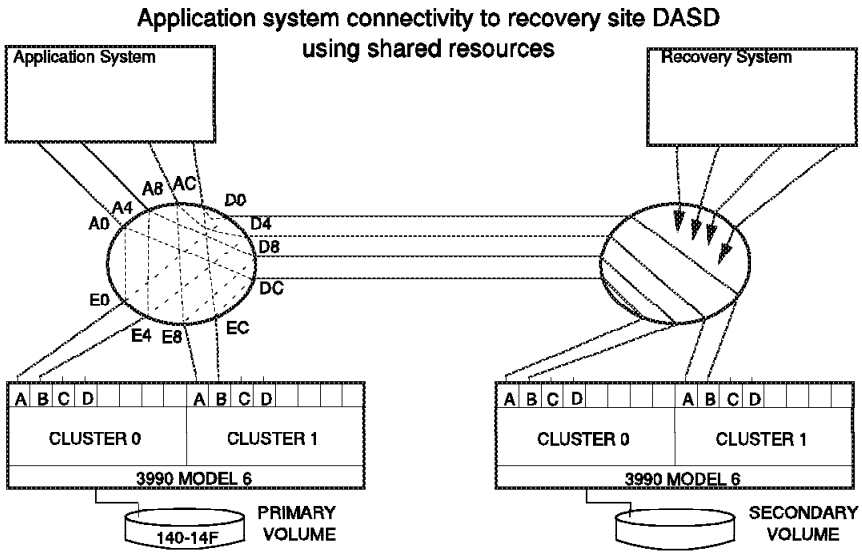


Figure 63. PPRC Configuration with ESCON Directors

Figure 63 shows a PPRC configuration that includes two chained ESCON directors (one at the application site and one at the recovery site).

The application has four-path access to the application site 3990 Model 006 through the ESCON Director. As an example of one path, the ESCON Director port A0 switches to ESCON Director port E0 to permit access to the primary volumes. The dotted lines indicate a total of four paths to the application site DASD.

The sharing of ESCON Director, 3990 Model 006, and fiber link resources is also possible with configurations that require application system connectivity to the recovery site DASD. For example, in the figure above, the following paths can be established:

- Application system paths to primary volumes

  For example: ESCD port AC switching to ESCD port EC
- Application system paths to recovery site DASD

  For example: ESCD port AC switching to ESCD port D0
- PPRC paths from primary volumes to the secondary volumes

  For example: ESCD port EC switching to ESCD port DC

All of this switching can be achieved by using the same physical resources as before.

Note: The recovery system can also connect to the recovery site 3900 Model 006 as shown by the arrowheads in the diagram. However, because switching can only take place in one of the two chained ESCON directors, the connections for PPRC data transfer at the recovery site ESCON Director are dedicated. Thus, the recovery system′s access to the recovery site 3990 Model 006 must involve the use of additional ESCON Director ports and 3990-6 system adapter interfaces.

3.6.5 Establishing and Monitoring PPRC Links

There are seven TSO commands that control PPRC operations. These are provided as maintenance to MVS/DFP 3.1.1 and above. The commands are available for PPRC operations as soon as TSO is active. You can issue the PPRC commands to automatically initiate PPRC subsystem activity.

In this section, we discuss several of the basic PPRC commands. The complete command set as well as their full syntax can be found in *DFSMS/MVS Version 1 Remote Copy Administrator's Guide and Reference*, SC35-0169-03.

3.6.5.1 Establishing a PPRC Path

The CESTPATH command establishes a path between the application site 3990 Model 006 and the recovery site 3990 Model 006. It uses the following parameters:

DEVN Defines the device number of any volume behind the application site 3990 Model 006.

PRIM Describes the primary volume's 3990 storage control. It defines the SSID and the serial number of the application site 3990.

SEC Describes the secondary volume's 3990 storage control. It defines the SSID and serial number of the recovery site 3990.

LINK Defines which 3990 interfaces on the application site 3990 will be used to connect to the recovery storage control. In addition, it provides the ESCON Director port addresses if they are used for the PPRC path. There can be multiple links specified in each CESTPATH command.

The link parameter contains an eight-digit value that defines the physical components in the path. It does not define any part of the 9729 The syntax of the link parameter is as follows: `LINK (X'AAABCCDDD)` where:

- AAA identifies the 3990 Model 006 cluster number:
  - 000 is cluster 0.
  - 001 is cluster 1.
- B identifies the 3990 Model 006 interface:
  - 0 is interface A.
  - 1 is interface B.
  - 2 is interface C.
  - 3 is interface D.
  - 4 is interface E.
  - 5 is interface F.
  - 6 is interface G.
  - 7 is interface H.
- CC identifies the outgoing ESCON Director port address:
  - Any valid ESCON port address
  - 00 for non-switched configurations
- DD is always 00 for ESCON.

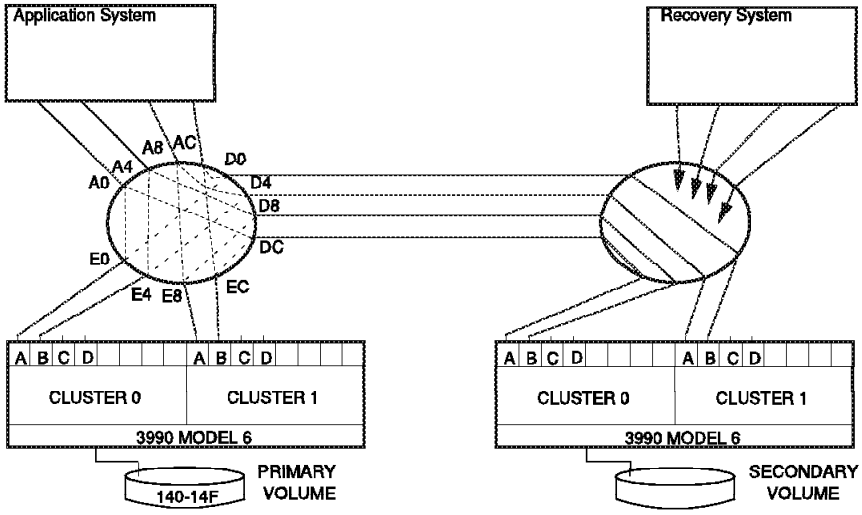


Figure 64. A Sample PPRC Configuration

Figure 64 shows a sample PPRC configuration.

Using the above figure as a reference, the following example illustrates the command to establish four paths between the 3990 attaching device number 14A (subsystem ID of 2010 with a serial number of 3990-73-62512) and the 3990 with an SSID of 2011 and a serial number of 3990-73-43156.

```
CESTPATH DEVN(X'014A')
         PRIM(X'2010'62512)
         SEC(X'2011'43156)
         LINK(X'0000D000')
             (X'0001D400')
             (X'0010D800')
             (X'0011DC00')
```

Figure 65. Establishing a Four-Path PPRC Configuration

3.6.5.2 Monitoring a PPRC Path

You have three methods to check that a PPRC path is functioning normally:

- ICKDSF

  ICKDSF can be used to monitor logical paths used by PPRC operations. The ANALYZE NODRIVE NOSCAN command performs this function.

- ESCON Director

  PPRC paths and operations are not known to the ESCON manager, as they are not defined in either the IOCP or MVSCP definitions within the hardware configuration dialog (HCD). As a result, when making any change to an ESCON Director switch matrix you must be aware of all existing PPRC paths.

  The recommended way of minimizing the risk of inadvertently interrupting PPRC pairs at the ESCON Director is to adopt a naming convention for ESCON Director ports so that it is obvious to an operator at the ESCON

Director hardware console or an ESCON manager screen that the ports in question are used by a PPRC pair.

Figure 66 shows you a sample window from the 9032 Model 003 Active Matrix window. The Hdwr column is host controlled and gives you the status of the port. A blank in this column indicates that the port is online.

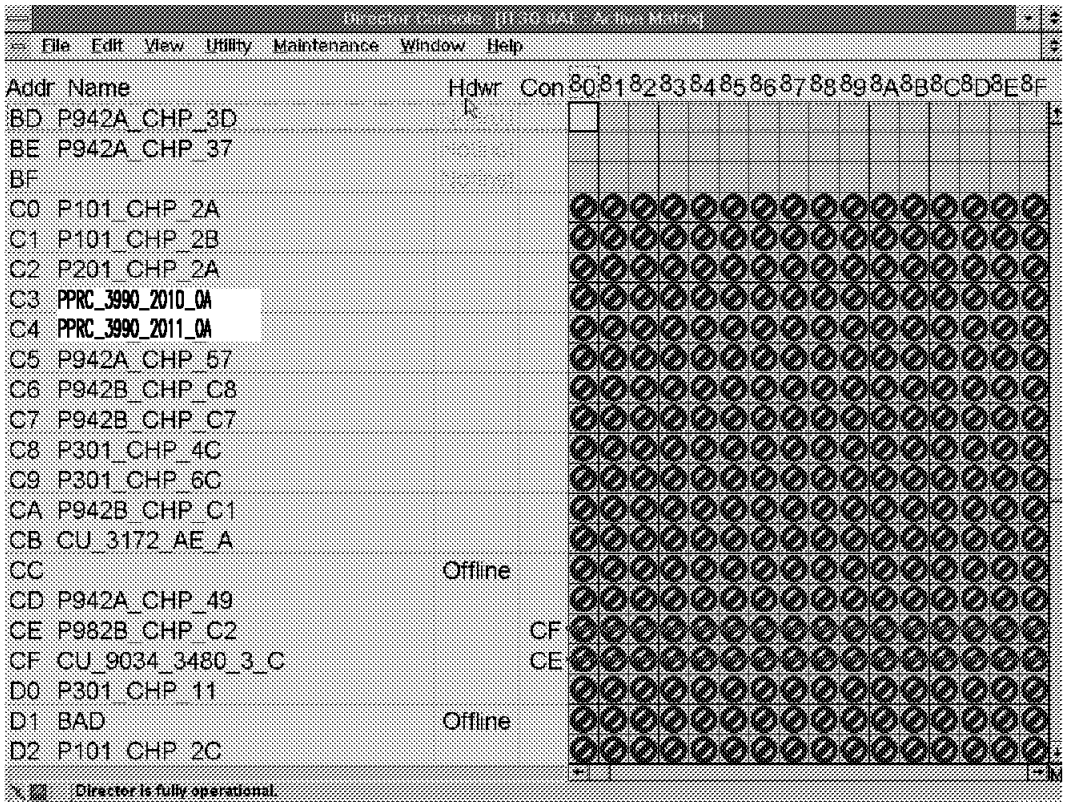


Figure 66. ESCD Active Matrix Window

An operator can modify the active switch matrix for the director on the ESCON Director hardware console. This matrix holds a description that must be manually entered for all ports on the ESCON Director. Each port address is allowed a 24 alphanumeric character field that should be maintained to accurately reflect the connectivity of each port. The ESCON manager product can also remotely view and update the description field. A meaningful address name can help avoid confusion.

- The TSO CQUERY command

  The CQUERY command is used to query the status of one volume of a PPRC volume pair, or all paths associated with the storage control for the named device number. Figure 67 on page 97 shows you the result of a sample CQUERY command. These results are returned to the TSO user ID that issued the command and copied to the SYSLOG.

```
****************** PPRC CQUERY REMOTE COPY - PATHS *********************

PRIMARY UNIT: SERIAL#=000007362512    SSID=2010

              FIRST          SECOND          THIRD          FOURTH
            SECONDARY      SECONDARY       SECONDARY      SECONDARY
 SERIAL NO: 000007343156   ............   ............   ............
      SSID:     2011           0000            0000           0000
     PATHS:      4              0               0              0

            SAID DEST S*   SAID DEST S*   SAID DEST S*   SAID DEST S*
            ---- ---- --   ---- ---- --   ---- ---- --   ---- ---- --
          1:0000 D000 01
          1:0001 D400 01
          1:0010 D800 01
          1:0011 DC00 01


S*=PATH STATUS
00=NO PATH           01=ESTABLISHED               02=INIT FAILED
03=TIME OUT          04=NO RESOURCES AT PRIMARY   05=NO RESOURCES AT SECONDARY
06=SERIAL#MISMATCH   07=RESERVED                  08=RESERVED
09=RESERVED          10=CONFIGURATION ERROR
```

Figure 67. The Result of the CQUERY Command

The information on the panel describes the paths from the primary storage control, as specified on the CQUERY command, to the secondary storage control. Up to four paths are displayed.

In the example shown, the application site 3990 has four paths to a recovery site storage control configured via an ESCON Director.

Note: The 9729 Optical Wavelength Division Multiplexer will not appear in the list but, a path shown as established by the CQUERY command is an indication that the 9729 is operating properly on that channel.

You can see that all four paths are established because the status column (S*) contains 01 which indicates that the path is established.

3.6.6 Link Error Reporting and Recovery

The 3990 ESCON system adapter cards transmit light continuously, so the links will be established automatically if the fiber trunk from the application site to the recovery site is functioning. The IBM 9729 will indicate via the LED status on the ETR/ESCON IOC and LRC cards whether the link is up across the 9729 trunk. (See 3.3.6, “Problem Determination on ESCON Links” on page 47)

The link failure detection and reporting is done by the 3990 System Adapter (SA) card. When it detects a link failure (for example if the fiber between the 9729s is cut) the SA card generates a link incident record (LIR) and sends it to the host. The host uses the LIR to recognize and register the error, and to initiate the proper corrective action.

Note

The LIR includes only the two nodes defined for the link. The node description for PPRC links includes only 3990s and ESCDs (if installed). The 9729s are not part of the node description; therefore they will not appear in the incident report.

Figure 68 shows an example of an LIR report from a 3990 with switched PPRC links.

```
FAILURE DATE 082997,TIME 1013 , VIA LOGROUTE
ESCON ASSOCIATED DATA,PRIMARY FAILING GROUP:
D/T=IBM 3990,MOD=006,SER=7351104,IF=0011,CLUSTER=1,INTERFACE=B
IQ=28, IC=03, FLAGS=20, NODE PARAMETERS=000100
SAC=35, LINK FAILURE BETWEEN TWO NODES,MULTIPLE INTERFACES FAILING NODE
DEV DEPENDENT 0-7=  0000000000000000  BYTES  8-15= 0000000000000000
      BYTES 16-23=  0000000000000000  BYTES 24-31= 0000000000000000
      BYTES 32-35=  00000000
 P/N0000079G3428 QTY01 CARD -PRI 9032-003 INTERFACE CARD
ESCON ASSOCIATED DATA,2ND     FAILING GROUP:
D/T=IBM 9032,MOD=003,SER=0021578,IF=00CA
IQ=28, IC=03, FLAGS=00, NODE PARAMETERS=000A00
DEV DEPENDENT 0-7=  2000980000000302  BYTES  8-15= 21F8014ACA050005
      BYTES 16-23=  0005000000000000  BYTES 24-31= 0000000000000000
      BYTES 32-35=  00000000
```

Figure 68. Link Incident Report (LIR)

The SA card can generate an LIR for any link connected to it but *must* transmit the LIR over another link on the same card.

Figure 69 shows two clusters of a 3990 Model 006 that illustrate both the correct and incorrect configuration design for PPRC installations.

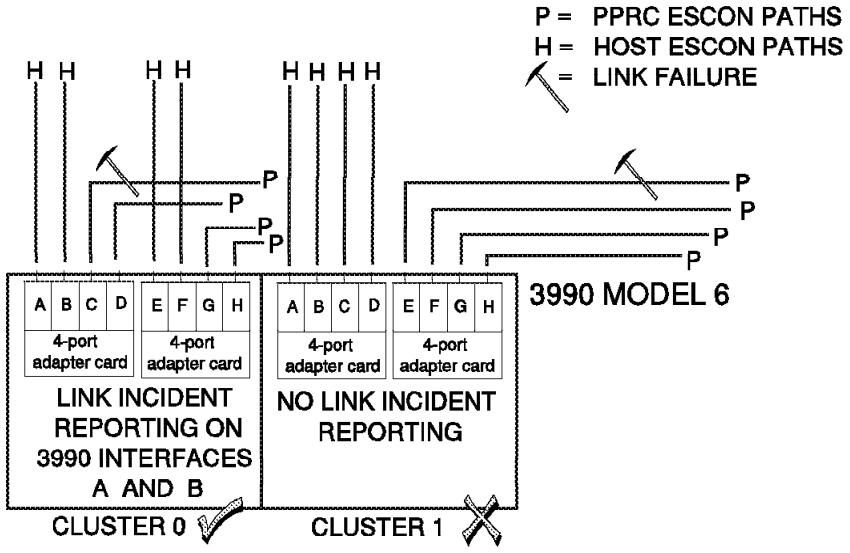


Figure 69. PPRC System Adapter (SA) Configuration

The correct way, cluster 0 (shown on the left), contains two 4-port SA cards. The host ESCON paths and the PPRC ESCON paths are split across SA cards. Interfaces A and B on one card are connected to the host channels while interfaces C and D on the same card are connected to the recovery site 3990 Model 006. Similarly, interfaces E and F on the second 4-port card are connected to the host channels, and interfaces G and H on the same card are connected to the recovery site 3990 Model 006. This technique improves overall availability and also allows concurrent maintenance to take place.

If a link failure occurs on a PPRC path that uses the link connecting interface C on the 3990 Model 006, the SA card can generate the LIR and transmit it to the host using the links connecting 3990 Model 006 interfaces A or B. Cluster 0 has been configured to support the LIR and ensure that there is a method for notifying the host of link incidents.

In contrast, cluster 1 demonstrates the incorrect way of configuring PPRC links. In cluster 1, the two 4-port cards are configured in such a way that each card uses all of its interfaces to connect either the host channels or the recovery site 3990. Interfaces A, B, C, and D all connect to the host, and interfaces E, F, G, and H all connect to the recovery site 3990 Model 006. Thus, if there is a link failure on the PPRC path that uses interface E on cluster 1, there is no way of notifying the host that a failure has occurred. The LIR can only be transmitted through a link on the same card that detected the failure in the first place. The cluster 1 configuration results in link failure going undetected and the appropriate recovery action not taking place.

Appendix A. Fiber Optic Technology

This appendix provides the reader with some basic principles of fiber optic technology and fiber optic communication. Also, the major advantages of optical fibers over copper cables are given. This material is from the IBM redbook, *ESCON Implementation Guide*, SG24-4662 and is being reprinted here for your convenience.

For a more thorough description of fiber optic technology you may want to look at an excellent publication entitled *Introduction to Fiber Optics for IBM Data Communications*, ZZ81-0215.

A.1 Optical Fiber Advantages

Optical fibers have several important advantages over traditional copper cables. Optical fibers are therefore increasingly being used to replace copper cables not only in communications systems, but in various other applications. The advantages include the following:

- Larger bandwidth

  A fiber optic cable can sustain a much higher data rate than a copper cable. Typical bandwidths are 500 MHz to 100 GHz over 1 km for different kinds of optical fibers.

- Physical characteristics

  A fiber optic cable is very flexible. It is small and light, and consequently it is much easier to install than a copper cable. Even though a fiber optic cable is very flexible, care should be taken not to bend the cable too sharply. The IBM 3044 Jumper Cable has a minimum bend radius of 48 mm (approximately 2 inches). It is also highly resistant to environmental conditions such as temperature, water, light, and radiation. There are many benefits when using fiber optic cables.

- No electromagnetic interference

  A signal transmitted in a fiber optic cable does not generate electromagnetic interference (EMI), does not have radiated electromagnetic susceptibility (RES), and is not affected by radio frequency interference (RFI). These problems have to be considered when a transmission uses copper cables.

- Low loss

  The light signal travelling in a fiber optic cable will lose its energy very slowly, usually at less than 1 dB/km.

- Small crosstalk and high security

  Since there is no electrical radiation from the fiber optic cables, they have a high degree of security when compared to copper cables. Although fiber optic cables are not easy to tap, it is possible. Security must be taken into consideration when “dark fibers” from telephone companies or other vendors providing network services are used. As the transmitted light is converted to electrical signals in repeaters, data may be exposed.

- No common ground

  The fiber optic cables use light, not electricity to transmit signals, so the signals are not affected by the electrical disturbances to ground. Also, when devices are not electrically connected there will be no ground loops, or currents, caused by differing electrical ground potential.

A.2 Optical Fibers

An optical fiber functions as a kind of waveguide for light. It is usually made of silica glass. The fiber itself has a central core and a surrounding cladding of slightly different glass material. These are protected by a plastic or nylon coating, sometimes called a buffer. The physical size of an optical fiber is determined by the diameter of the core and cladding, expressed in microns (um). One micron is a millionth (1/1,000,000) of a meter. A fiber optic cable having a core diameter of 62.5 um and a cladding of 125 um is designated as 62.5/125 um optical fiber. You can see an example of an optical fiber cable in Figure 70.

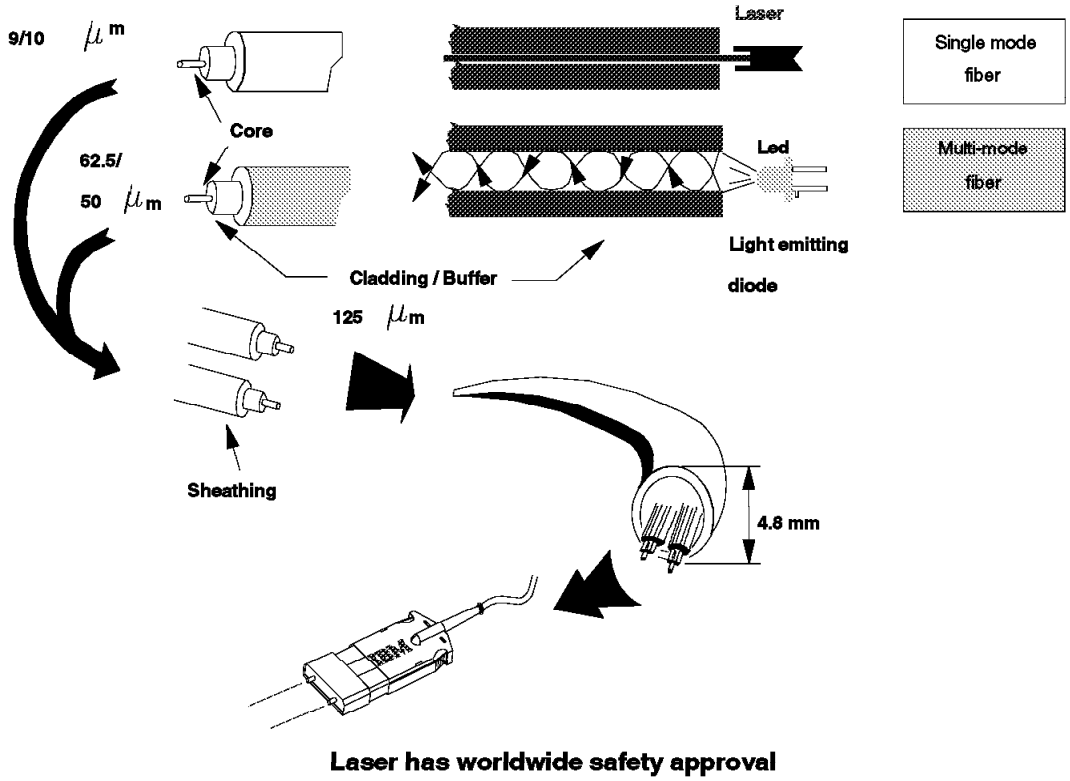


Figure 70. Fiber Optic Cable

The optical fibers are coated with several kinds of protective layers. The material and the thickness of these layers depend on the usage of the cable. An optical fiber cable normally contains a pair of fibers, or several pairs of fibers. The former is called a jumper cable and the latter a trunk cable. The outer diameter of a jumper cable is typically 4.8 mm, or 0.19 inch, while a trunk cable with 72 pairs of fibers measures about 12.5 mm, or half an inch in diameter. Optical fiber cables are manufactured for different environments and they should always be used within their design criteria.

A.3 Propagation of Light in an Optical Fiber

The light used for transmitting signals in fiber optic communication systems could be visible light or invisible light such as infrared or ultraviolet. However, only certain frequencies are suitable for use in optical fibers, as the attenuation of the signal varies with the wavelength. The wavelengths of 850 nanometers (nm) and 1300 nm have the lowest attenuation, so these are the most commonly used wavelengths. The power of light used to transmit a signal in an optical fiber is relatively small; 10-50 milliwatts. Still, you should follow all the safety precautions when you work with devices transmitting light. As the light used is invisible, you may not know if the transmission is on. The light used may cause harm to human tissues and even blindness if aimed at an eye.

A.4 Bandwidth

Light travels in a vacuum at a speed of 300,000,000 meters per second. In glass, the speed is slower, 200,000,000 meters per second. However, the data transmission capacity is determined by the bandwidth used. The higher the bandwidth, the greater the amount of data carried through an optical fiber. Normally the bandwidths for optical fibers are given in MHz-km. A bandwidth of 500 MHz-km denotes that a 500 MHz signal can be transmitted over a 1 km distance. This is the typical bandwidth for multimode fibers. Bandwidths for single-mode fibers are in the GHz range, typically 100 GHz over a 1 km distance. Using lower frequencies we can send light signals, or pulses, over longer distances. The distance is limited due to dispersion, or scattering, of the signal so that the pulses cannot be separated from each others at the receiving end. There are several reasons for dispersion and we will discuss them later.

A.5 Transmission Modes

There are two transmission modes we can use to send light signals through an optical fiber: single-mode or multimode. The optical fibers used are called single-mode or multimode fibers. These optical fibers have different physical dimensions and light transmission characteristics.

The term *mode* is related to the number and variety of wavelengths that may be propagated through the core of an optical fiber. In other words, it describes the propagation path of a light ray in the core of an optical fiber.

When the core has a comparatively large diameter, light rays enter the optical fiber at different angles compared to the central axis of the core. These light rays will reflect from the interface of the core and cladding and their paths in the core follow a zig-zag pattern. Light rays also enter the core parallel to the central axis of the core. These light rays will follow a straight path through the core. We have now light rays in different modes travelling in an optical fiber. This is called multimode transmission.

Small cores allow only the light rays almost parallel to the central axis of the core to enter an optical fiber. So we have fewer modes travelling in the fiber. This is called single-mode transmission, where only one frequency of light, or one mode, is propagated.

A.5.1 Single-Mode Fiber

A single-mode fiber usually has a core diameter of 8 to 10 microns and a cladding diameter of 125 microns. The light source used for single-mode optical fibers is a *laser* (Light Amplification by Stimulated Emission of Radiation). A laser commonly used in communication systems is generated by a semiconductor laser diode (LD). Light generated by a laser diode is very coherent. The maximum distance for a single-mode optical fiber link is 20 km.

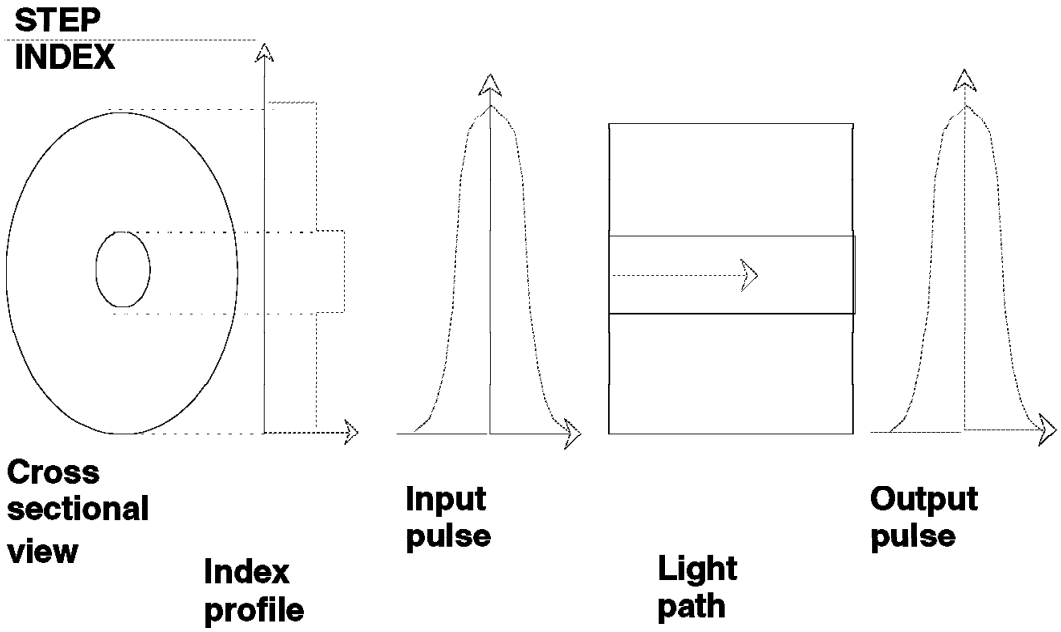


Figure 71. Single-Mode Fiber Profile

A.5.2 Multimode Fiber

As described, multimode fibers carry several modes concurrently. The 62.5/125 or 50/125 um optical fiber cables are supported by IBM. A light emitting diode (LED) is usually the light source for multimode fibers. The maximum distance for a multimode optical fiber link is 3 km using a 62.5/125 um optical fiber.

Even though the distance achieved by multimode fibers is relatively short compared to that for single-mode fibers, multimode fibers have several advantages and therefore were selected for campus-wide communications in the ESCON architecture. For several years to come, the multimode fiber is the right choice, although further in the future single-mode fibers will be used more frequently. If you are planning an installation, you should consider installing both multi- and single-mode fiber cables even though the single-mode cables would have no use yet.

A.6 Refractive Index Profile

Beside transmission mode, another important term is refractive index profile. The refractive index profile describes the relationship between the index of the core and the cladding. There are two commonly used types of refractive index profiles: step and graded.

A.6.1 Step Index

In a step-index fiber, the core has a uniform index, but there is an abrupt change in the refractive index between the core and the cladding.

A.6.2 Graded Index

In a graded-index fiber the core has a gradually decreasing refractive index from the center of the core outward to the cladding boundary. This causes the light rays propagated in a graded-index optical fiber to travel at almost the same speed even though they have different geometrical paths. The graded-index fibers are made by doping the glass material used to manufacture the fibers. Figure 71 and Figure 72 describe the step and graded indices of fiber optic cables.

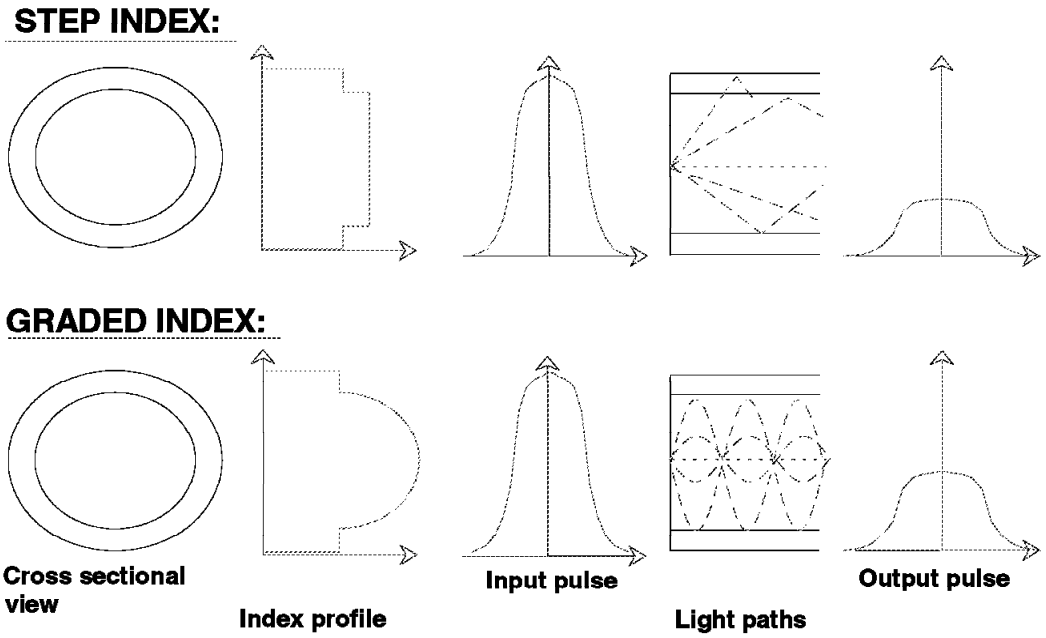


Figure 72. Multimode Fiber Profiles

A.7 Dispersion

Light rays travelling parallel to the central axis of the core, called lower modes, will be faster than the rays reflecting from the interface of the core and the cladding. These reflected rays are called higher modes. If we have both lower and higher mode rays in a multimode optical fiber, the rays will arrive at slightly different times at the receiver. This is called mode delay, or modal dispersion. Modal dispersion may prevent differentiation of the incoming pulses in a receiver. Modal dispersion could significantly limit the useful distance of a multimode optical fiber. Using the graded index optical fibers for multimode transmission decreases the modal dispersion, as both the higher and lower modes proceed at the same speed.

Dispersion is also caused by different wavelengths in the light signal transmitted into an optical fiber. The different wavelengths travel with different speed and so they can cause merging of the pulses. This is a common problem in single-mode fibers. It is best avoided using a light source of coherent light. The impurities in the material used for optical fibers can also cause dispersion, as lower modes may convert to higher modes.

A.8 Light Sources and Detectors

To make use of optical fibers, we have to convert the electrical pulses that electrical devices use to light pulses, and vice versa. This conversion is done using transmitters and receivers. Since we usually have a pair of fibers for a fiber optic link, each end of the link needs a transmitter and a receiver. Let us have a look what kind of devices are used.

A.8.1 Transmitters

A transmitter emits, or sends, light into an optical fiber. The light rays should all be of the same wavelength. Remember what we learned earlier about dispersion. The emitted light works as a carrier. For data transmission, we have to add the data in the carrier. We can do this by modulating the emitted light. As described later in this manual, data is transmitted through an optical fiber serially. So if data appears as parallel signals to our transmitter, it must first be converted to serial signals.

The devices used in transmitters can be LEDs or laser diodes. LEDs generate light that has a narrow range of wavelengths around the central wavelength. It is used for multimode fibers. An LED is a small and quite inexpensive component. Laser diodes are rather complicated devices and therefore usually more expensive than LEDs. Their advantages are the intensity and accuracy of the light. The light is more coherent than the light from a LED. Also, a light of high power can be generated. Laser devices are used with single-mode fibers.

A.8.2 Receivers

The light travelling through an optical fiber must be detected to convert it back to an electrical signal. Detection is done by light-sensitive devices called photodiodes. After being detected, the signal is amplified and usually digitized.

A.9 Fiber Optic Standards

In addition to the ESCON architecture, there are other standards based on fiber optics. This section provides information on some of the standards that are either available or being designed.

You may often hear the terms proprietary and open systems in conjunction with fiber optic standards. The proprietary systems cover architectures such as S/390 and ESCON. They are owned as the property of somebody, usually a company. Open systems are architectures that different vendors may use to produce products. These products can be used in an installation connected to products of any vendor, assuming that all the products meet the same standard.

A.9.1 Fiber Channel Standard

Fiber Channel Standard (FCS) is ANSI standard X3T9.3. It became available for the first public review late in 1992. The base for FCS is laser technology with a bandwidth of 1 GHz. It will be first utilized by NIC and scientific applications providing a high performance parallel interface (HIPPI). In the Fiber Channel Standard, switching is not architected; only the ends are defined. Switching in the network is based on the destination and source addresses. There may be several switches in a network. It may become possible to use FCS as a high-speed backbone network. It may also be used as a general transfer vehicle for upper-level protocols, such as TCP/IP, HIPPI, SBCC, and so on. These upper-level protocols may be intermixed on a link.

A.9.2 Synchronous Optical Network (SONET)

SONET is a U.S. standard for the internal operations of optical networks. It is closely related to a system called Synchronous Digital Hierarchy (SDH), which is a CCITT recommendation for the internal operations of optical networks.

SONET can mix packages, or frames, from different sources and carry them over a network. It uses a family of rates of 51.84 Mbps. Using multiples of these rates, SONET is capable of reaching gigabit rates per second. Within SONET, several bandwidths can be transported simultaneously.

A.9.3 Integrated Services Digital Network (ISDN)

ISDN describes and specifies a digital user interface to a public communication network. It does not specify the internal operations of a network, but the interfaces to it and services it provides.

There are three generic types of ISDN:

- Narrowband ISDN can utilize 64 Kbps copper links primarily on a switched services basis.
- Wideband ISDN is a form of ISDN where a user is able to access a wider synchronous data channel utilizing several 64 Kbps copper links.
- Broadband ISDN is a cell-based packet switching system. It does not offer synchronous links as narrow- and wideband ISDN, but Asynchronous Transfer Mode (ATM) cell switching.

ISDN is already in use for carrying speech and data simultaneously over a channel. It is aimed at many applications, not only for computing, but consumer services at offices, homes, and other locations.

A.9.4 Asynchronous Transfer Mode (ATM)

ATM is a protocol that has been accepted as the basis for the broadband ISDN service. It is aimed at high-speed cell switching systems. ATM is suitable for all types of traffic, including voice, data, image, and video. Information is transferred in cells that contain a header defining the route through a network. Each cell may contain data, or payload, of 48 bytes.

A.9.5 Fiber Distributed Date Interface (FDDI)

FDDI was developed by ANSI. It was originally proposed as a standard for fiber optical computer channels, but has become a generalized standard for operation of a LAN at 100 Mbps. The FDDI standard was approved in 1991, and there are many FDDI devices available on the market.

FDDI is primarily intended for operation over optical fiber, but recently has been proposed for operation over standard copper wire (shielded twisted pair). Using multimode optical fiber, a FDDI ring may be up to 200 km in length with a maximum of 500 stations. Data transfer takes place in frames or packets. The maximum frame size is 4500 bytes and contains a header with the physical destination address of a FDDI station. The ring protocol of FDDI is conceptually similar to the token-ring LAN (IEEE 802.5), but differs significantly in detail.

FDDI was approved quite early, and so it has been implemented and accepted by many customers.

Appendix B. Special Notices

This publication is intended to help IBM and customer I/T professionals understand and use the capabilities of the IBM 9729 Optical Wavelength Division Multiplexer. The information in this publication is not intended as the specification of any programming interfaces that are provided by IBM. See the PUBLICATIONS section of the IBM Announcement for the IBM 9729 Optical Wavelength Division Multiplexer for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM′s product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM′s intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, 500 Columbus Avenue, Thornwood, NY 10594 USA.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM (″vendor″) products in this manual has been supplied by the vendor and IBM assumes no responsibility for its accuracy or completeness. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer′s ability to evaluate and integrate them into the customer′s operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other

operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

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

AIX	AIXwindows
CICS	DB2
DFSMS/MVS	Enterprise System/9000
Enterprise Systems Connection Architecture	ES/9000
ESA/390	ESCON
Extended Services	IBM
IMS	Multiprise
MVS	MVS/DFP
MVS/ESA	NetView
OS/2	OS/390
Parallel Sysplex	PR/SM
RACF	RAMAC
RMF	RS/6000
S/370	S/390
SP2	Sysplex Timer
System/370	System/390
VTAM	

The following terms are trademarks of other companies:

C-bus is a trademark of Corollary, Inc.

Java and HotJava are trademarks of Sun Microsystems, Incorporated.

Microsoft, Windows, Windows NT, and the Windows 95 logo are trademarks or registered trademarks of Microsoft Corporation.

PC Direct is a trademark of Ziff Communications Company and is used by IBM Corporation under license.

Pentium, MMX, ProShare, LANDesk, and ActionMedia are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

Tru-Wave is a Trademark of AT&T.

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

Appendix C. Related Publications

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

C.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see “How to Get ITSO Redbooks” on page 113.

- *High-Speed Networking Technology: An Introductory Survey*, GG24-3816-02
- *Enterprise Systems Connection (ESCON) Implementation Guide*, SG24-4662
- *OS/390 MVS Parallel Sysplex Configuration Cookbook*, SG24-4706-00
- *System/390 MVS Parallel Sysplex Continuous Availability*, SG24-4502-00
- *System/390 MVS Parallel Sysplex Migration Paths*, SG24-2502
- *System/390 Parallel Sysplex Performance*, SG24-4356-02
- *OS/390 Parallel Sysplex Application Considerations*, SG24-4743-00

C.2 Redbooks on CD-ROMs

Redbooks are also available on CD-ROMs. **Order a subscription** and receive updates 2-4 times a year at significant savings.

CD-ROM Title	Subscription Number	Collection Kit Number
System/390 Redbooks Collection	SBOF-7201	SK2T-2177
Networking and Systems Management Redbooks Collection	SBOF-7370	SK2T-6022
Transaction Processing and Data Management Redbook	SBOF-7240	SK2T-8038
Lotus Redbooks Collection	SBOF-6899	SK2T-8039
Tivoli Redbooks Collection	SBOF-6898	SK2T-8044
AS/400 Redbooks Collection	SBOF-7270	SK2T-2849
RS/6000 Redbooks Collection (HTML, BkMgr)	SBOF-7230	SK2T-8040
RS/6000 Redbooks Collection (PostScript)	SBOF-7205	SK2T-8041
RS/6000 Redbooks Collection (PDF Format)	SBOF-8700	SK2T-8043
Application Development Redbooks Collection	SBOF-7290	SK2T-8037

C.3 Other Publications

- *9729 Operator's Manual*, GA27-4172-02
- *9729 Maintenance Information*, GY27-0357-02
- *OS/390 V2R4.0 MVS System Commands*, GC28-1781-03
- *OS/390 Parallel Sysplex Systems Management*, GC28-1861
- *IBM 9037 Model 2 Planning Guide*, SA22-7233-01
- *Using the 9037 Model 2 Sysplex Timer*, SA22-7230-01
- *Planning for the 9037 Model 2*, SA22-7233-00
- *9672/9674 Managing Your Processors*, GC38-0452-05
- *9672/9674 Hardware Management Console Guide*, GC38-0453-05
- *ESA/390 ESCON I/O Interface*, SA22-7202-02

- *Using the 9032 Model 5 ESCON Director*, SA22-7296-00
- *ESCON Physical Layer*, SA23-0394-02
- *DB2 V4 Data Sharing; Planning and Administration*, SC26-3269-01
- *DFSMS/MVS Version 1 Remote Copy Administrator's Guide and Reference*, SC35-0169-03

How to Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, CD-ROMs, workshops, and residencies. A form for ordering books and CD-ROMs is also provided.

This information was current at the time of publication, but is continually subject to change. The latest information may be found at http://www.redbooks.ibm.com/.

How IBM Employees Can Get ITSO Redbooks

Employees may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Redbooks Web Site on the World Wide Web**

  http://w3.itso.ibm.com/

- **PUBORDER** — to order hardcopies in the United States
- **Tools Disks**

  To get LIST3820s of redbooks, type one of the following commands:

  ```
  TOOLCAT REDPRINT
  TOOLS SENDTO EHONE4 TOOLS2 REDPRINT GET SG24xxxx PACKAGE
  TOOLS SENDTO CANVM2 TOOLS REDPRINT GET SG24xxxx PACKAGE (Canadian users only)
  ```

  To get BookManager BOOKs of redbooks, type the following command:

  ```
  TOOLCAT REDBOOKS
  ```

  To get lists of redbooks, type the following command:

  ```
  TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET ITSOCAT TXT
  ```

  To register for information on workshops, residencies, and redbooks, type the following command:

  ```
  TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ITSOREGI 1998
  ```

- **REDBOOKS Category on INEWS**
- **Online** — send orders to: USIB6FPL at IBMMAIL or DKIBMBSH at IBMMAIL

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (http://www.redbooks.ibm.com/redpieces.html). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

How Customers Can Get ITSO Redbooks

Customers may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Online Orders** — send orders to:

	IBMMAIL	**Internet**
In United States:	usib6fpl at ibmmail	usib6fpl@ibmmail.com
In Canada:	caibmbkz at ibmmail	lmannix@vnet.ibm.com
Outside North America:	dkibmbsh at ibmmail	bookshop@dk.ibm.com

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	(long distance charges apply)
(+45) 4810-1320 - Danish	(+45) 4810-1020 - German
(+45) 4810-1420 - Dutch	(+45) 4810-1620 - Italian
(+45) 4810-1540 - English	(+45) 4810-1270 - Norwegian
(+45) 4810-1670 - Finnish	(+45) 4810-1120 - Spanish
(+45) 4810-1220 - French	(+45) 4810-1170 - Swedish

- **Mail Orders** — send orders to:

IBM Publications Publications Customer Support P.O. Box 29570 Raleigh, NC 27626-0570 USA	IBM Publications 144-4th Avenue, S.W. Calgary, Alberta T2P 3N5 Canada	IBM Direct Services Sortemosevej 21 DK-3450 Allerød Denmark

- **Fax** — send orders to:

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	(+45) 48 14 2207 (long distance charge)

- **1-800-IBM-4FAX (United States)** or **(+1)001-408-256-5422 (Outside USA)** — ask for:

  Index # 4421 Abstracts of new redbooks
  Index # 4422 IBM redbooks
  Index # 4420 Redbooks for last six months

- **On the World Wide Web**

Redbooks Web Site	http://www.redbooks.ibm.com/
IBM Direct Publications Catalog	http://www.elink.ibmlink.ibm.com/pbl/pbl

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (`http://www.redbooks.ibm.com/redpieces.html`). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

IBM Redbook Order Form

Please send me the following:

Title	Order Number	Quantity

First name Last name

Company

Address

City Postal code Country

Telephone number Telefax number VAT number

- Invoice to customer number
- Credit card number

Credit card expiration date Card issued to Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

Index

Numerics

19-inch rack-mountable card cage 15
8P1786 67
9021 64
9032 45
9034 Model 1 37
9035 Model 2 37
9037 Model 001 50
9037 Model 002 50
9037 Sysplex Timer (Model 2) 7
9037 Sysplex-Timer Model 002 External Time Reference (ETR) 16
9390 40
9391 40
9672 64
9674 64
9729 Intersystem Channel (ISC) 71
9729-001 6
9729-041 6

A

A side 14
Active Matrix 45
active trunk 17
active wait 65
aging 20
AIX 23
alternate fiber 63
amplifier 12
APD 19
Asynchronous Transfer Mode 108
ATM 3, 7, 19
See also Asynchronous Transfer Mode
attenuation 12, 18, 20, 67, 103
availability 17

B

B side 14
backup path 5, 17
bandwidth 4, 101, 103
BatchPipePlex 67
bibliography 111
BIT-ERROR-THRESHOLD (BER) 85
Broadband ISDN 107

C

Cache 65
card cage 16
carrier concentrations 13
CF shutdown 72
CFCC microcode 68
CFR 65
CFS 65
Channel Control Words (CCWs) 33
channel extenders 6
channel spacings 21
channel subsystem (CSS) 33
channel-to-channel (CTC) pathing 67
channel-to-channel configuration 36
chirp 13
chirped in-fiber Bragg gratings 19
CHPIDs 44
CICS 65
CICS/VSAM 65
circuit-switched protocol 31
cladding 102, 105
class 1 laser product 14
client/server applications 3
clock recovery 16
clock synchronization 5
Code Division Multiple Access (CDMA) 4
coherency 65
concave mirror 11
continuous availability 69
Control Link Oscillator (CLO) 16, 50
cooling 71
core 103, 105
cost 5
coupling facility 64
coupling facility control code (CFCC) 65, 82
Coupling Facility Resource Management (CFRM) 72
cross system extended services (XES) 71
cross-system coupling facility (XCF) 67
crosstalk 12, 18, 20, 21, 101
cyclic redundancy check 33

D

dark fibers 101
data center consolidations 3
DB2 65, 67
detectors 13, 106
diagnostic card 15, 17, 22
dispersion 17, 103, 106
dispersion compensating fiber 18
dispersion-shifted fiber 21
distortion 18
Distributed Feedback (DFB) laser 13
doping 105
drift 13
dual fiber I/O card 16
Dual Fiber Switching Feature 17
dual port card 55

E

electromagnetic interference 101
electromagnetic spectrum 5
electronic tuning 13
Enterprise Systems Connection (ESCON) 6, 10
ES/9000 processors 2, 7
ESA/390 architecture 31
ESCON 104
ESCON Card 16
ESCON Directors 2, 7
ESCON Extended Distance Feature (XDF) 35
ESCON Multiple Image Facility (EMIF) 36
ESCON-capable Control Units 7
Ethernet 23
Ethernet collision domain 22
Ethernet-F 7
Exchange ID (XID) 32
Expanded Availability configuration 52
External Time Reference (ETR) 48

F

fan 71
fans 15
Fast Ethernet 7, 16
FCS
  See Fiber Channel Standard
FDDI 3, 19
  See also Fiber Distributed Date Interface
FDDI card 16
feedback control 13
Fiber Channel Standard 107
Fiber Distributed Data Interface (FDDI) 7
Fiber Distributed Date Interface 108
Fiber Optic Standards 107
  Asynchronous Transfer Mode 108
  Fiber Channel Standard 107
  Fiber Distributed Date Interface 108
  Integrated Services Digital Network
  Synchronous Optical Network 107
fiber optic technology 101
Fiber Optics 101
  Advantages 101
  bandwidth 103
  Dispersion 106
  Graded index 105
  Laser 104
  LED 104
  Multimode fiber 104
  Propagation of light 103
  Receiver 106
  Refractive Index Profile 105
  Single-mode fiber 104
  Step Index 105
  Transmission Modes 103
  Transmitter 106
Four Wave Mixing (FWM) 20, 21
frames 31
Frequency Division Multiplexing (FDM) 4
front panel 15
front panel reference 47
fused fiber coupler 12

G

Global Resource Serialization (GRS) 77
Graded Index 105
grating assembly 11, 15
grating controller 17
grating controller card 12, 15

H

Hardware Management Console (HMC 65
Hardware Management Consoles (HMCs) 79
hardware system area (HSA) 71
harmonics 21
high performance parallel interface (HIPPI) 107
HiPerLinks 67
hot-pluggable 17

I

I/O Configuration Data Set (IOCDS) 32
IMS 65
incident code 85
Indium Phosphide Avalanche PhotoDiode (APD) 13
input/output cards (IOCs) 10
insertion loss 12
Integrated Services Digital Network 107
Integrated Services Digital Network (ISDN) 107
Inter-System Coupling (ISC) Card 16
Inter-System Coupling (ISC) Channel 7
interference 20
interference (crosstalk) 18
Internal Throughput Rate (ITR) 71
Intersystem Channel (ISC) adapter 66
intranets 3
IOC 17
IOCP 65
IOCs 15
ISC 84
ISDN
  See Integrated Services Digital Network

J

JES2 65
jitter 16, 17, 19

L

LAN-to-LAN connectivity 3
Laser 104
laser (Light Amplification by Stimulated Emission of Radiation) 104

laser cavity 13
laser diodes 106
laser safety 14
laser/receiver cards (LRC) 13
laser/receiver cards (LRCs) 10
lasers 17
LED
See Light emitting diode
LED indicators 47, 84
LED status indicators 16
LEDs 106
Light emitting diode 104
light sources and detectors 106
lights-out mode 2
line costs 1
linewidth 13
link budget 20
link errors 33
link recovery 43
Lists 65
Littrow grating 21
Littrow reflective grating 11
lock structure 65
long-haul communication networks 4
loss 12, 20, 101
lost synchronization 33
LPAR 65
LRC 17
LRCs 15

M

Management Information Base (MIB) 22
master card 55
maximum distance 22
mean time between failures (MTBFs) 17
Millions of Service Units (MSUs) 70
modal dispersion 106
mode 103
Multimode fiber 104
MVS 65
MVS console 45
MVS console log 79

N

narrow linewidth 19
Narrowband ISDN 107
NetView/6000 23
network management 22
Network Management Station 23
noise 12, 18, 20
non-linear effects 21

O

OC-3 3, 7, 16
Offline Sequence (OLS) 54
Open Fiber Control 15
Open Fiber Control (OFC) 16
operational costs 2
Optical Carrier Level 3 (OC-3) 7
Optical fibers 102
See also Fiber Optics
Optical Network icon 80
optical networks 4
optical time-division multiplexing (OTDM) 4
OSAM 65
OW19728 52

P

pair gain 5
Parallel Sysplex 64
PBX equipment 3
photodiodes 106
physical configuration 15
pluggable modules 15
point-to-point topology 35
power 33, 103
power level 20
power planning 29
power supplies 15
PR/SM logical partitions 36
prism 11
Problem Analysis window 80
problem determination 47, 64, 84
propagation delay 22, 39, 57, 67
protocol independence 5

R

RAMAC 3 40
re-clocking 20
re-clocks 16
re-timing 16
receiver 13, 20, 106
receiver sensitivity 20
receivers 106
redundancy 17
refraction 11
refractive index 105
Refractive Index Profile 105
reliability 17
remote site backup 2
remote tape archiving 3
RMF Subchannel Activity Report 70
RPQ number 67

S

S/370 Bus and Tag architecture 31
S/390 Multiprise 2000 Server 7
S/390 Multiprise 2000 Servers 2
S/390 Parallel Enterprise Servers 2, 7
safety 103

scattering 103
sequences 31
shifted fiber 18
signal re-shaping 19
Simple Network Management Protocol (SNMP) 22
Single-mode fiber 104
SNMP agent 23
SNMP-DPI subagent 23
SONET
See Synchronous Optical Network
SONET/SDH 16, 19
speed of light 22
standard fiber 21
star coupler 12
Start Subchannel (SSCH) command 33
Step Index 105
Stimulated Brillouin Scattering (SBS) 21
Stimulated Raman Scattering (SRS) 21
structures 65
subagent 23
switched point-to-point topology 35
Synchronous Digital Hierarchy (SDH) 107
Synchronous Optical Network
Synchronous Optical Network (SONET) 107
Synchronous Optical Networks (SONET) Optical Carrier Level 3 (OC-3) 7
Sysplex Timer Attachment Feature 62
Sysplex Timer console 58
Sysplex Timer Network console 61
Sysplex Timers 2
System/370 Bus and Tag 37

T

tail circuits 43
TCP/IP 23
temperature changes 13
temperature control system 12
Test Laser button 48
throughput 38
Time Division Multiplexing (TDM) 5
TOD clock synchronization 57
token-ring network 23
transmission errors 33
Transmission Modes 103
transmitter 13, 18, 106
transmitter power 20
transmitters 106
tune 13

U

Uninterruptible Power Supply (UPS) 30
unshifted fiber 21

V

VSAM 65
VTAM/GR 65

W

waveguide 102
Wavelength 103
wavelength allocation 14
Wavelength Division Multiplexing (WDM) 4, 9
wavelength range 9
Wideband ISDN 107
Windows 95 23
Windows NT 23

ITSO Redbook Evaluation

IBM 9729 Optical Wavelength Division Multiplexer
SG24-2138-00

Your feedback is very important to help us maintain the quality of ITSO redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at http://www.redbooks.ibm.com
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Which of the following best describes you?
__Customer __Business Partner __Solution Developer __IBM employee
__None of the above

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction _____

Please answer the following questions:

Was this redbook published in time for your needs? Yes____ No____

If no, please explain:

_____

_____

_____

_____

What other redbooks would you like to see published?

_____

_____

_____

Comments/Suggestions: **(THANK YOU FOR YOUR FEEDBACK!)**

_____

_____

_____

_____

_____

SG24-2138-00
Printed in the U.S.A.