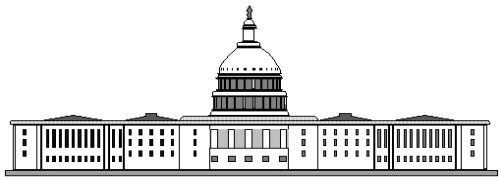
zSeries Peer Mode Links

John Hughes
IBM Washington Systems Center
Advanced Technical Support
jjhughes@us.ibm.com



1.0 Introduction

The intent of this paper is to provide a discussion on the use and definition of peer mode channels on an IBM z900 and z800 processors. It is hoped that the examples will clear up some of the confusion and misconception regarding this new capability.

For the examples, IOCP format was used rather than using HCD panels since experience has shown that the IOCP statements seem to be a better way to visualize what is going on. HCD or HCM is obviously the preferred method to be used for setting up any processor definitions.

2.0 Some History and Level Setting

IBM has provided coupling links for a long time now so it is presumed there is some knowledge on the part of the reader regarding this area. Peer mode was introduced with the z900 machines and has resulted in an increase in the terminology for link technology which should be discussed up front so that there is no confusion.

You will more than likely see in the documentation references to *legacy links*, *peer mode links*, *compatibility mode*, etc. Basically, links used on the earlier generation 9672 and 9674 machines are referred to as *legacy links*. These could be either the fiber type of links (ISC link), the copper cable Integrated Cluster Bus Links (ICB), or the Internal Coupling Channel Link (IC) introduced on the G5/G6 processors. The following will list and describe the different link types, where used, and the associated HCD/IOCP definitions:

1. **ISC-2:** This refers to the fiber type of legacy link which uses single mode fiber. It is better known by its feature code of #0008. There was a multimode type of link (feature code #0007) but this type of link cannot have connectivity to a z900 or z800 machine. An ISC-2 link can only be installed on a 9672 or 9674 type of machine.
   HCD/IOCP: CFS, CFR

2. **ISC-3:** This refers to the fiber type of link which uses single mode fiber and is the fiber link used on a z900 or z800 processor. It can connect to an ISC-2 on a G3 or later 9672/9674 or to another ISC-3 on a z900 or z800 machine. If it is connected to a 9672/9674 machine, it needs to run in what is called *compatibility mode*. If it is connected to another ISC-3, then it can be run in either compatibility mode or in peer mode. The actual mode is determined by the way the CHPID is defined in the HCD/IOCP.
   HCD/IOCP: CFS,CFR - Compatibility Mode, CFP - Peer Mode
3. **IC:** This refers to the Internal Coupling Channels for both the z900/z800(IC-3) and the 9672(IC-2) machines. However, it must be noted that when an IC channel is used on the z900 or z800, it **MUST** be defined as a peer mode link. The IC channels do not use real hardware but rather are defined with CHPID numbers that are not used by any of the installed hardware. **Note:** If any card is installed in a machine or port on a card activated that has a CHPID associated, it uses a CHPID even if it is not used or not defined in HCD.
   HCD/IOCP: ICS, ICR - 9672 Machine, ICP - z900
4. **ICB-2:** This refers to the ICB feature that is used on the 9672 processors (as feature code 0992) and can also be ordered on the z900 (also as feature code 0992). An ICB-2 feature on a z900 allows connectivity to an ICB-2 feature on either another z900 or on a 9672. An ICB-2 must always run in compatibility mode. An ICB-2 is not an option on the z800.
   HCD/IOCP: CBS, CBR
5. **ICB-3:** This is a new ICB feature that is only available on the z900 or z800 processor (as feature code 0993). An ICB-3 can only connect to another ICB-3 feature and must run in peer mode.
   HCD/IOCP: CBP

Please note, an ICB-2 feature can **ONLY** connect to an ICB-2 feature. An ICB-3 can **ONLY** connect to an ICB-3 feature. Furthermore, an ICB-3 connection must be run in peer mode and can only be between z900 or z800 processors. The reason for all this discussion is that there has been a misconception that an ICB-3 could be run in either compatibility mode or peer mode depending on its definition in HCD. **THIS IS NOT TRUE!** An ICB-3 can only run in peer mode. To illustrate this, look at the following diagrams representing the previous generation hardware (G5/G6) and the z900:

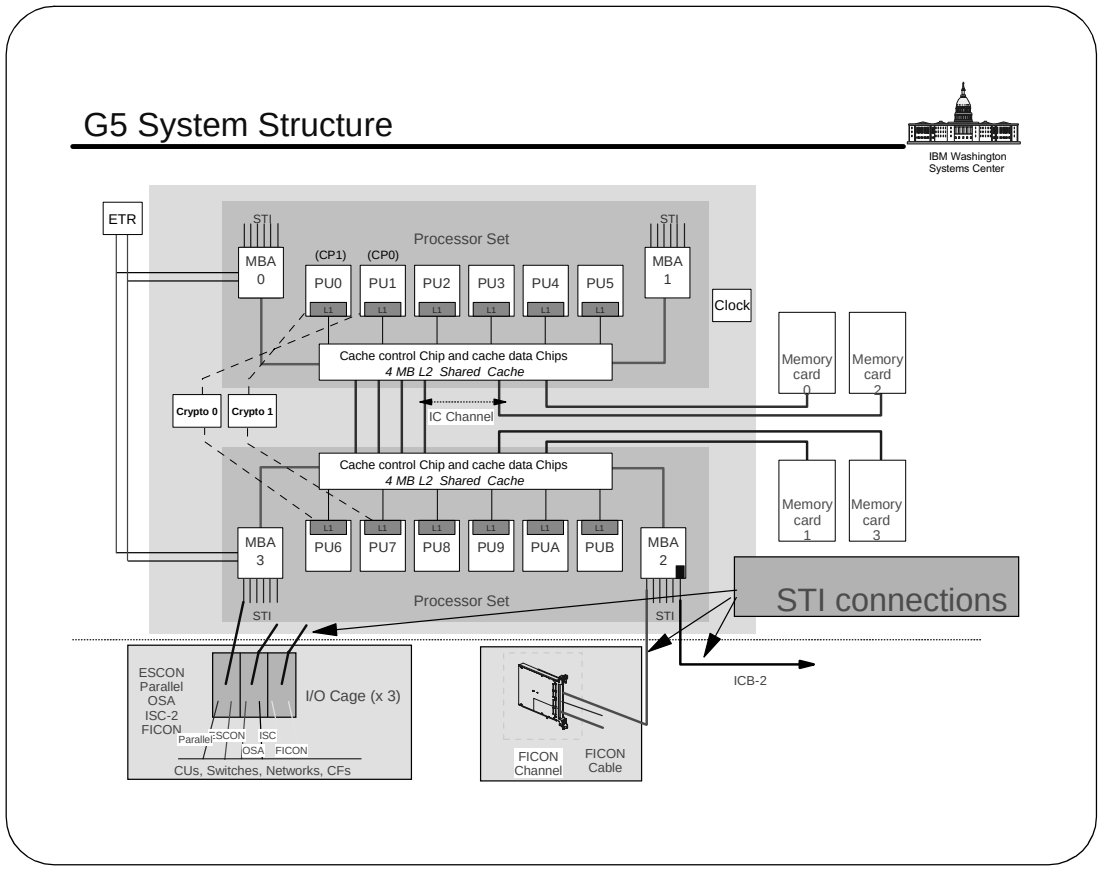


FIGURE 1. G5 System Structure

In Figure 1, please note that an ICB-2 feature originates from an STI port and can run at the STI speed of 333 MB/s. The ISC-2 channels originate from ISC daughter cards which are installed in the IO cage of the machine. FYI, this type of link can run at the speed of 1 Gb/s. Even though this figure represents a G5 machine, it is applicable to a G6 for the purposes of this discussion.

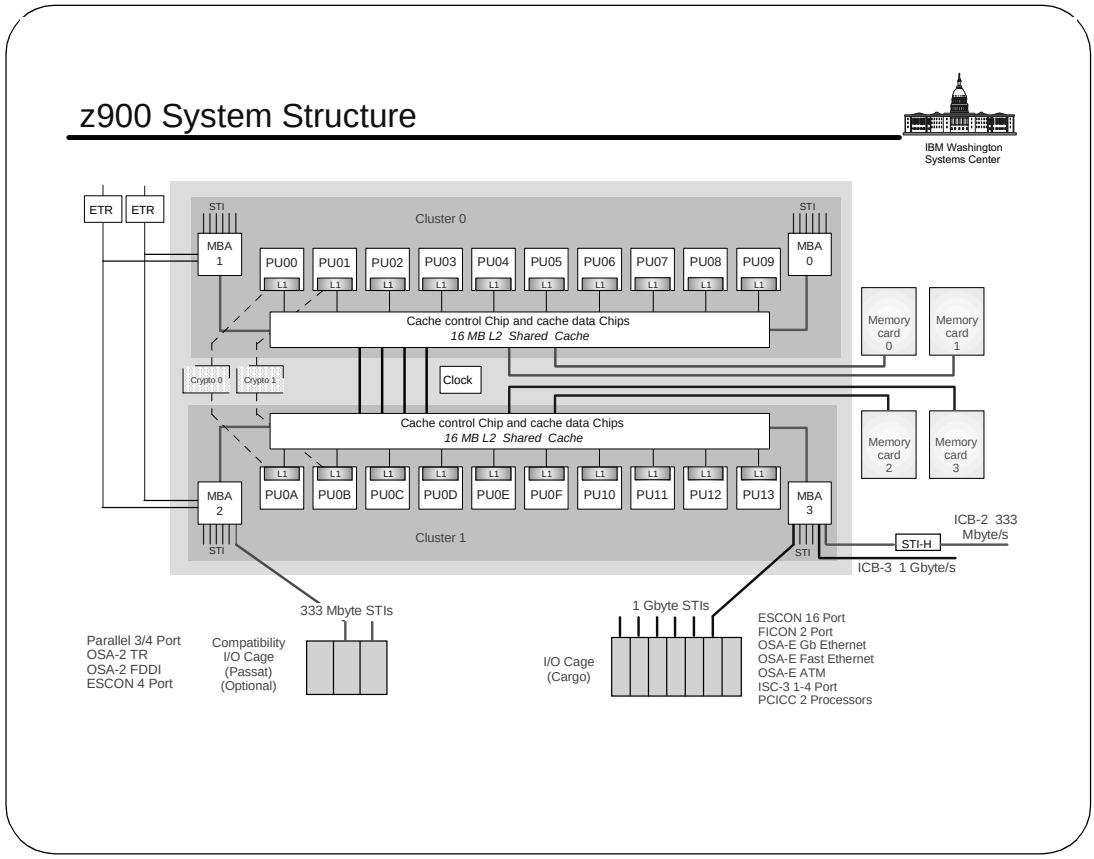


FIGURE 2. z900 System Structure

In Figure 2, you should see a couple of things. First of all, the ICB-3 originates from an STI port and can run at the STI speed of 1GB/s. Secondly, the ICB-2 feature must originate from what is referred to as an STI-H card. Basically, this card takes the native STI speed of 1GB/s and steps it down to run at the legacy STI speed of 333 MB/s. An STI-H card will take a single 1 GB/s STI and break it down into four 333 MB/s STI's. The point of this discussion is that an ICB-2 and an ICB-3 on a z900 use different hardware.

Also note that the ISC-3 links originate from an ISC-3 daughter card that is installed in the IO cage. Code is loaded into the card to allow the port to run in either peer mode (2 Gb/s) or compatibility mode (1 Gb/s).

3.0 Advantages of Peer Mode Links

Basically there are two primary advantages of peer mode links over and above the legacy links:

1. Increased speed of links and greater link utilization.

2. Reduced number of links required

The following sections will attempt to illustrate the second advantage. I have already indicated the speed differences but let's summarize here anyway:

1. **ISC-2**: 1Gb/s
2. **ISC-3**: 2Gb/s (peer mode)
3. **ICB-2**: 333MB/s
4. **ICB-3**:1GB/s (peer mode)

By far the fastest links will be the Internal Coupling Links which primarily use the CP structure of the machine. With ICs, there is never any reason to use external links between Operating System Partitions and the Coupling Facility Partitions on the same machine.

4.0 Using Peer Mode Links

Let's begin the discussion of how to use peer mode by first reviewing how one would set up and use a link in the legacy mode. Figure 3 is intended to represent a typical configuration with legacy links and Figure 4 is an illustration of how the IOCP would be coded by HCD. As you know, a legacy sender channel (CFS,CBS, or ICS) can be shared by multiple Operating System Logical Partitions. However, the other end of the link, the receiver side (CFR,CBR, or ICR), must be defined as dedicated or reconfigurable to a Coupling Facility Logical Partition.

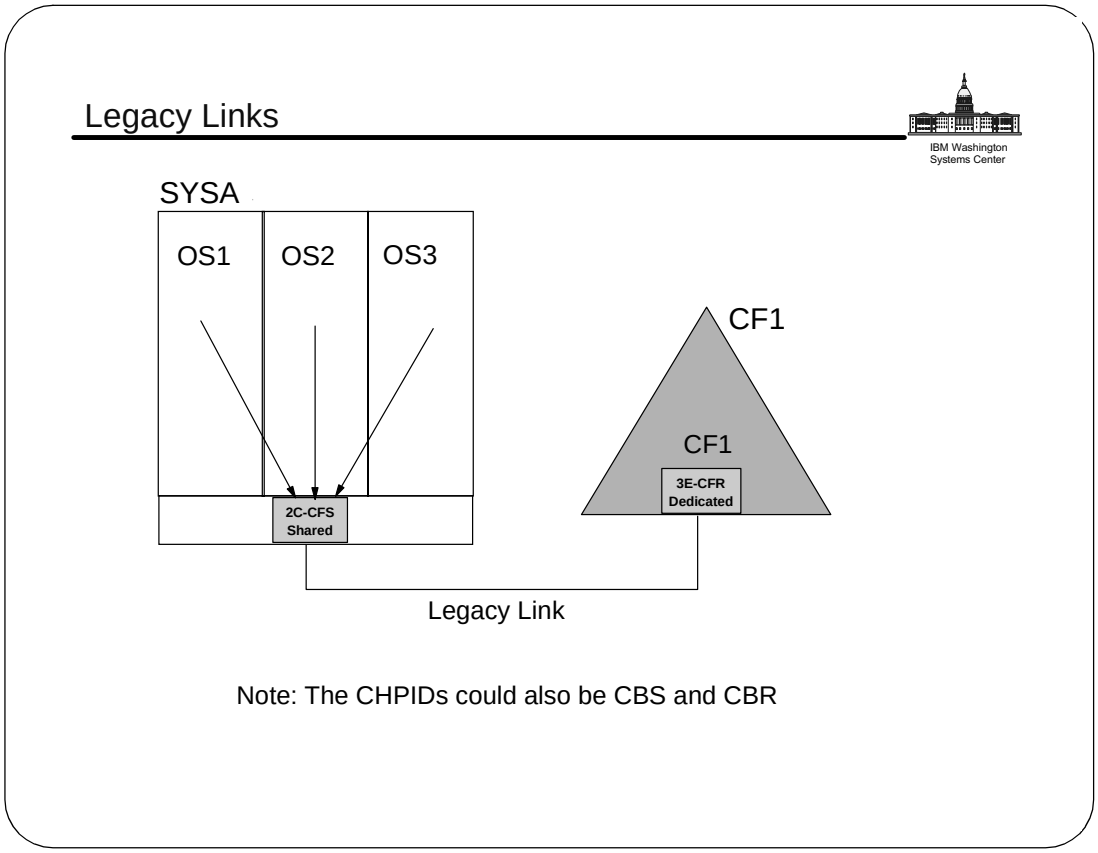


FIGURE 3. Legacy Link Configuration

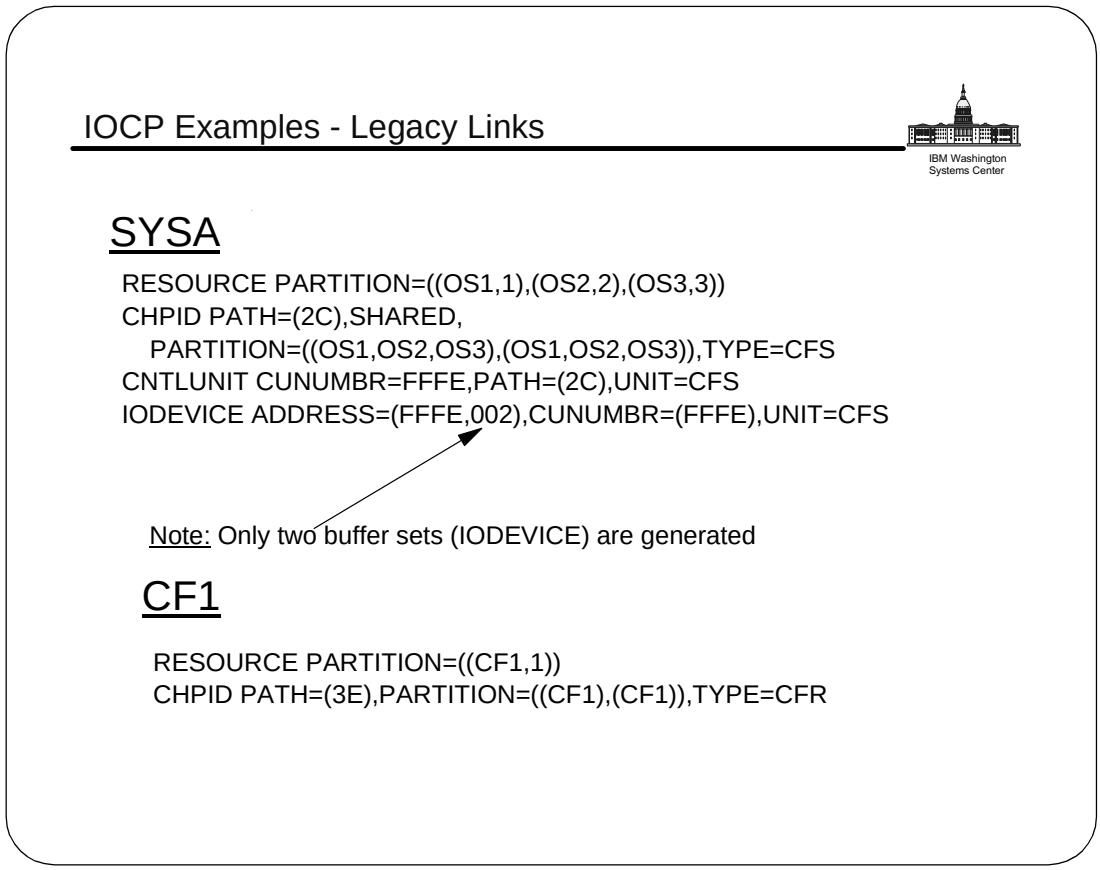


FIGURE 4. Legacy Link IOCP coding

Regarding the comments on the buffer sets, that is what the IODEVICE statement is set up to accommodate. Legacy links set up two buffer sets for use in transferring data across the links. As we will later see, peer mode increases the number of buffers which allows a better link utilization due to the larger number of buffers.

The following discussion will show how the same configuration might be set up and coded in a simple peer mode configuration.

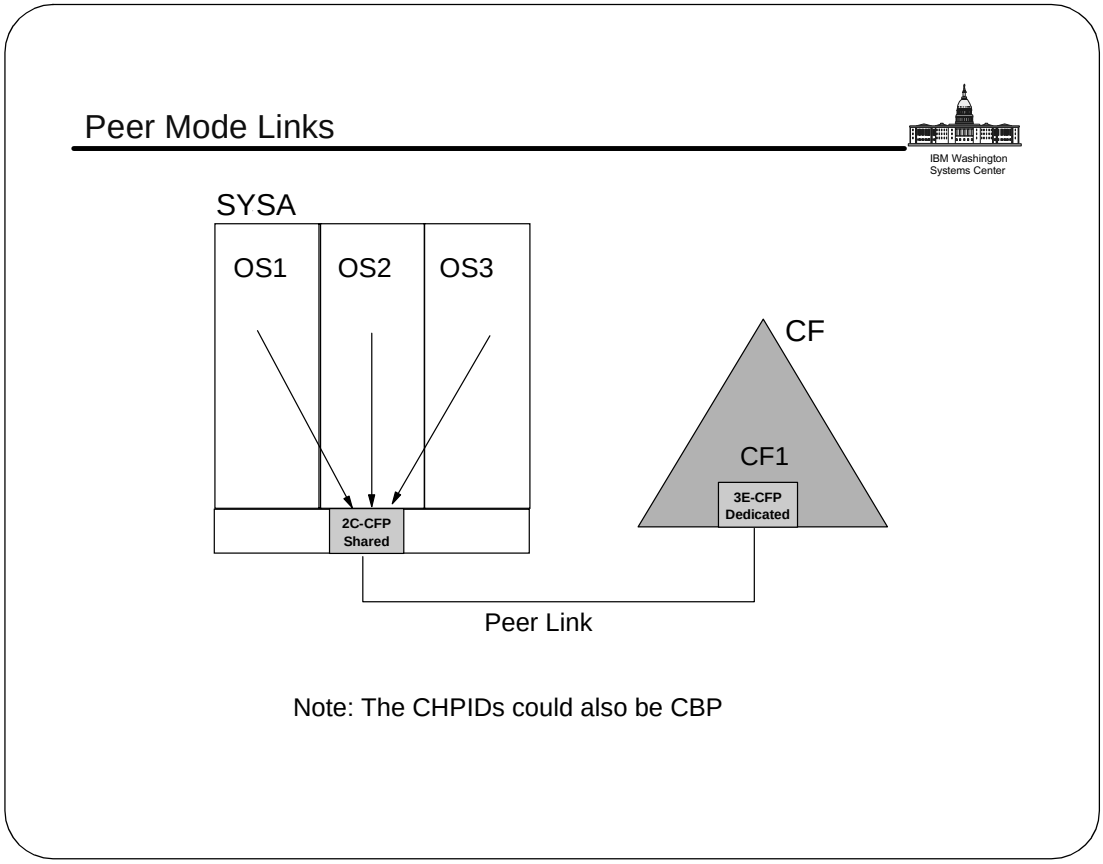


FIGURE 5. Peer Mode Link Example 1

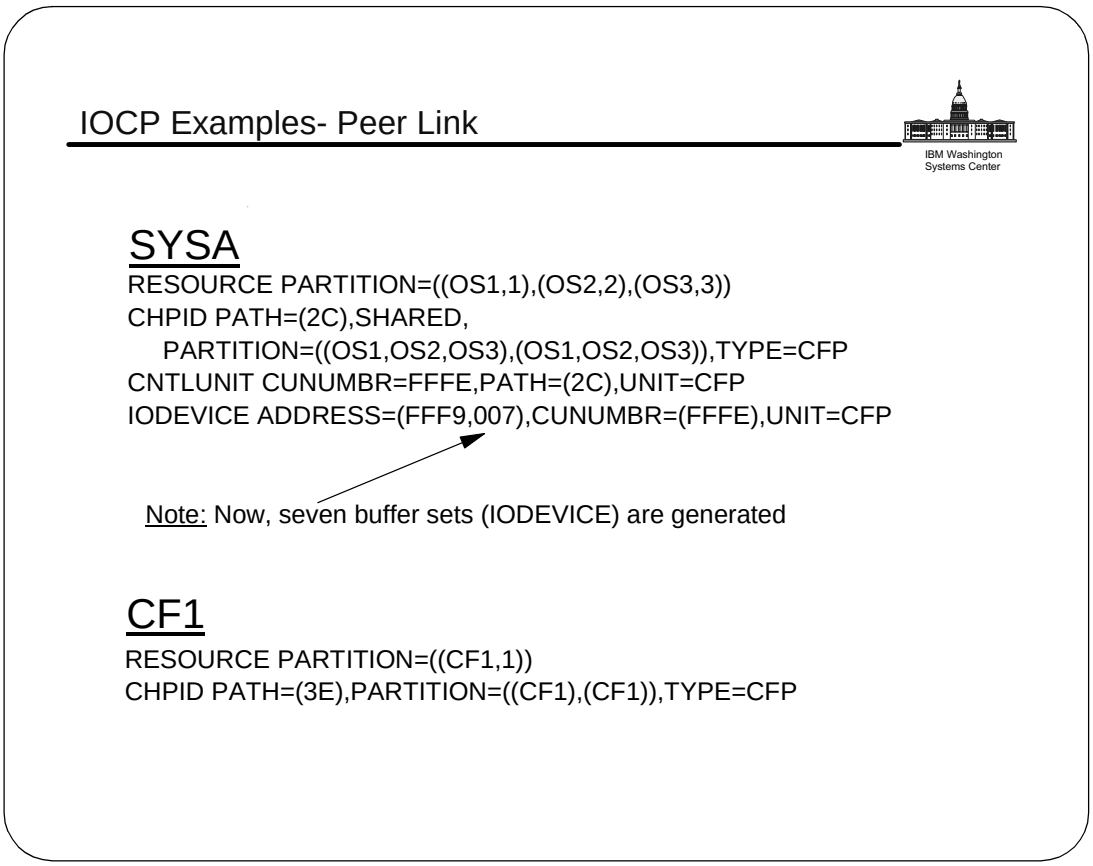


FIGURE 6. Peer Mode Link Example 1- IOCP

There is not a lot changed here except perhaps the advantages that will now be achieved due to the increased throughput capability of the peer mode links. Also noted, in order to accommodate the increased speed capability, additional buffer sets are allocated in order to handle the anticipated increase in traffic. With the speed of the links increased and the larger number of buffers, overall link utilization will be improved with peer mode links.

Just a point to note, this is a minimum configuration without much availability. Normally one would have redundant links to account for the case where one of the links should fail.

The next sections will illustrate a better use of peer mode links:

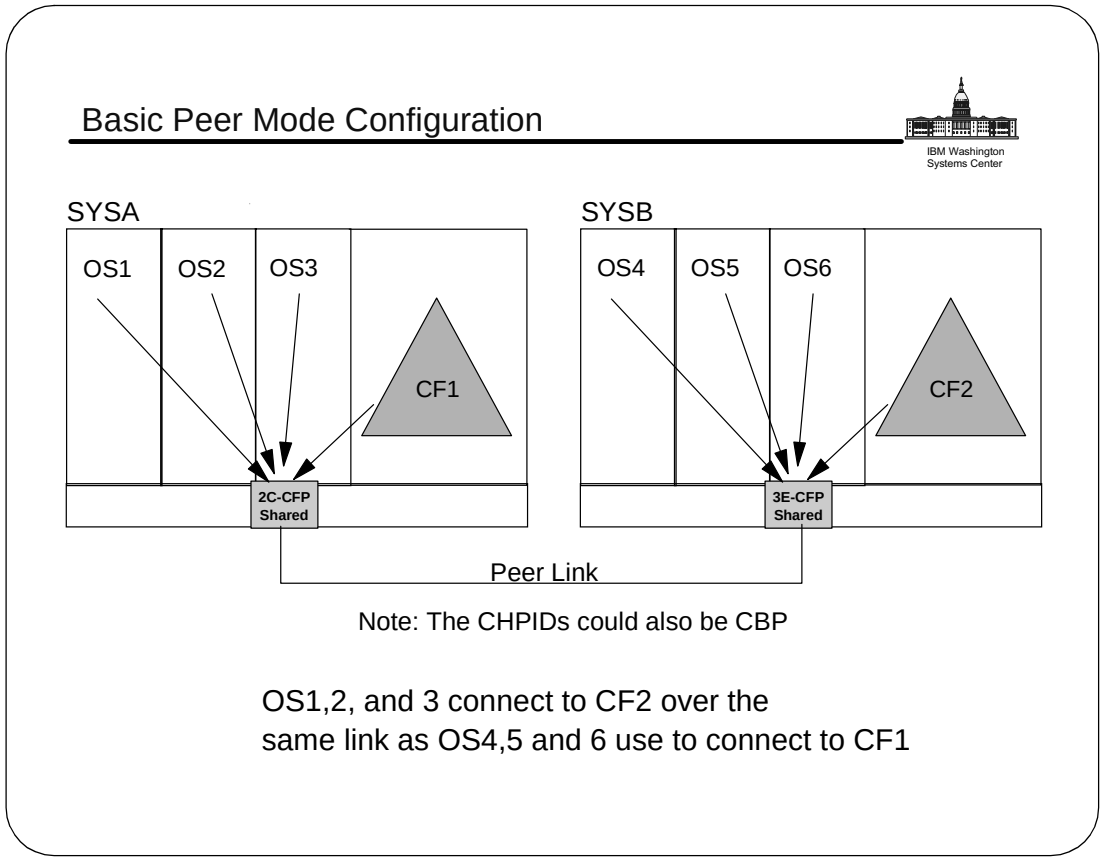


FIGURE 7. Peer Mode Link Example 2

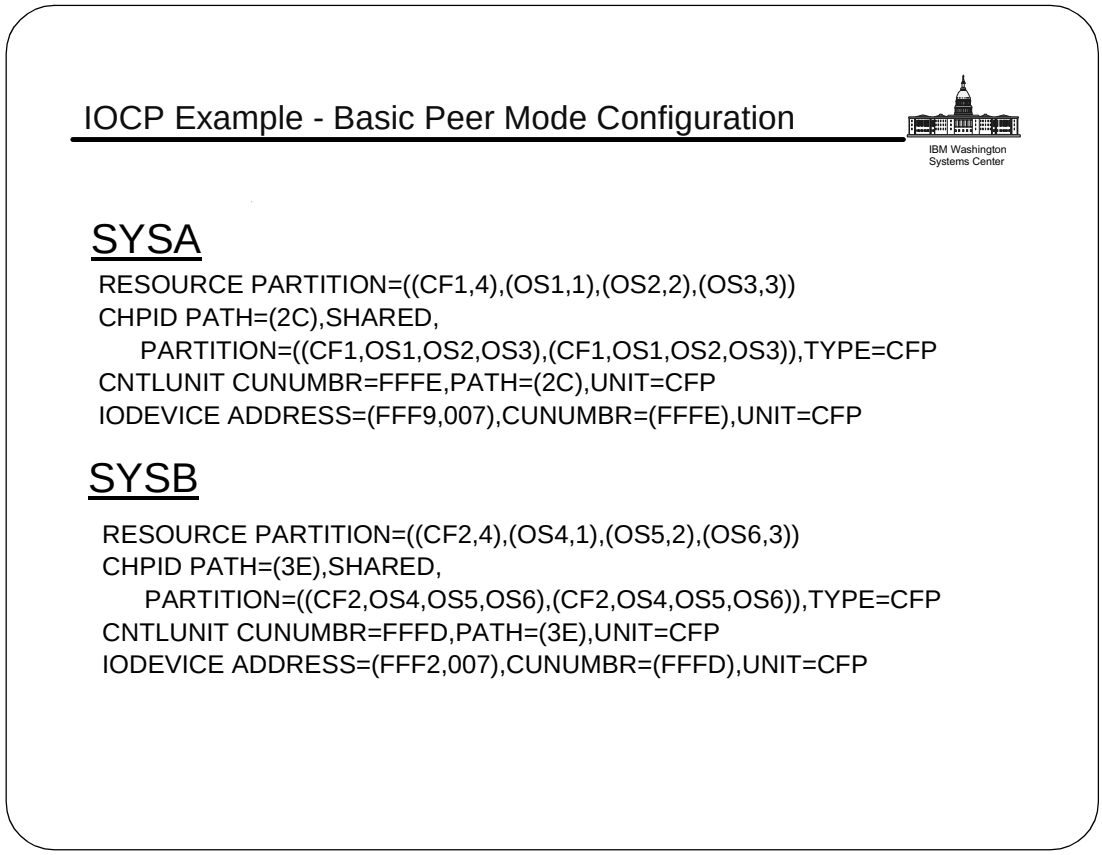


FIGURE 8. Peer Mode Link Example 2 - IOCP

This example should clearly show the value of peer mode and how the required number of links can be reduced. If this were a legacy type of configuration, there would need to be an additional link required for connectivity. One link would be used to allow the Operating System Partitions on SYSA to communicate with the Coupling Facility Partition on SYSB. A second link would be required to allow the Operating System Partitions on SYB to communicate with the Coupling Facility Partition on SYSA. If you have been dealing with links for awhile, you are aware that on a legacy link the sender side CHPID (CFS,CBS, ICS) can be shared by multiple Operating System Partitions whereas the receiver side CHPID (CFR,CBR, ICR) must be defined as dedicated or reconfigurable to a single Coupling Facility Partition. With peer mode, this restriction is relaxed. The CHPID associated with either end of the link is defined the same (CFP, CBP, ICP) and the CHPID can be shared by multiple Operating System Partitions and **one** Coupling Facility Partition. Thus, both ends of the peer mode link can be defined as shared with some exceptions. If you go back and look at Figure 5, note that the CHPID on the Coupling Facility is defined as dedicated. This is a standalone Coupling Facility (e.g. 2064-100). For this case, remember that it was indicated the CHPID for a peer mode link could be “shared by multiple Operating System Partitions and one Coupling Facility Partition”. On a standalone coupling facility, there are ONLY Coupling Facility Partitions and HCD will enforce that the CHPID for this case has to be dedicated or reconfigurable.

The reader may have noticed that there does not appear to be a way for the Operating System Partitions on either footprint to communicate with the Coupling Facility Partition on the same footprint. This is where the Internal Coupling Link channels come into play.

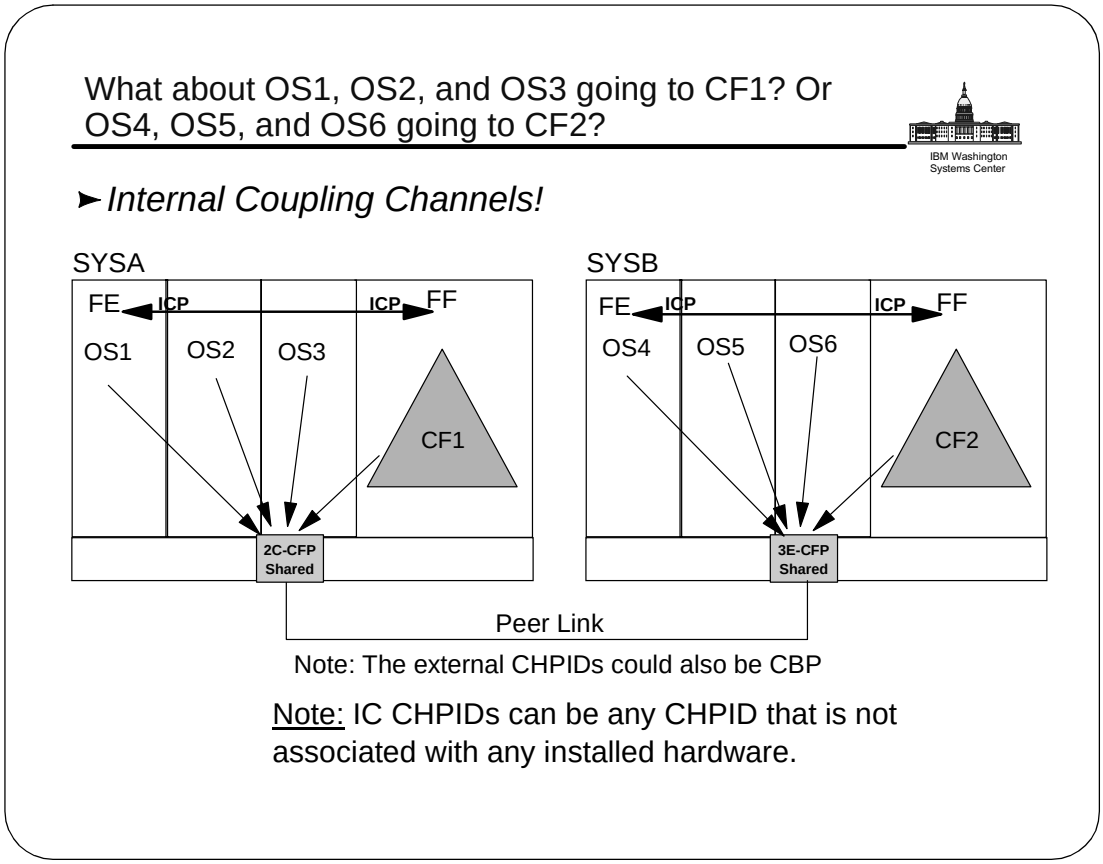


FIGURE 9. Peer Mode Link Example 3

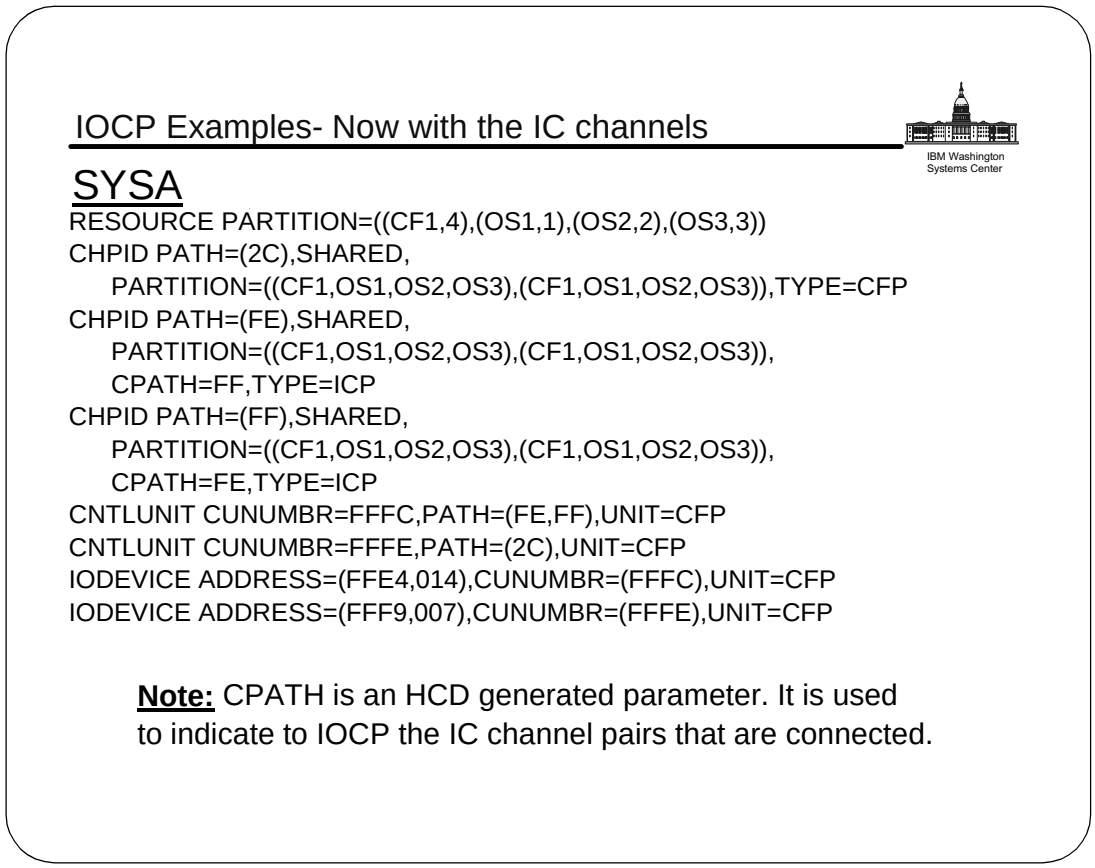


FIGURE 10. Peer Mode Link Example 3 - IOCP SYSA

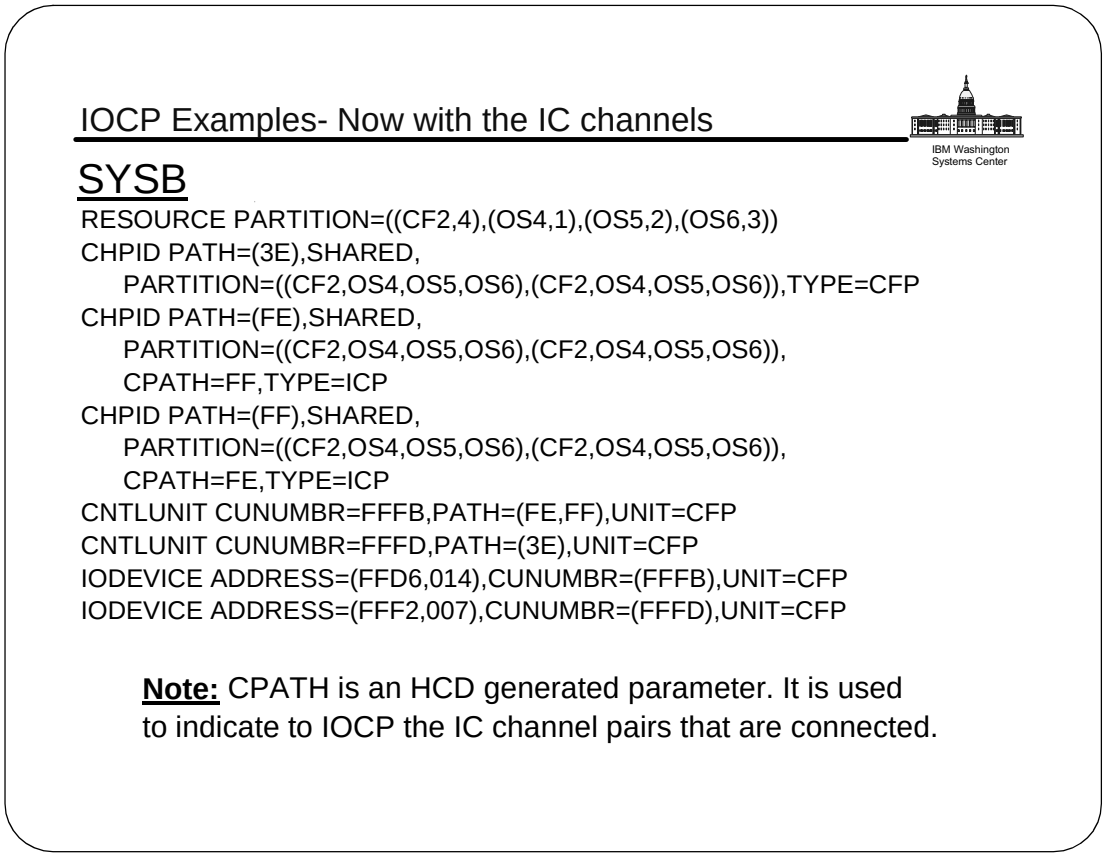


FIGURE 11. Peer Mode Link Example 3 - IOCP SYSB

Just a note on the CPATH parameter. As indicated, this is an HCD generated statement and you do not specify it in HCD. All it is used for is to indicate the CHPID pairs that are connected for an Internal Coupling Link.

An interesting item to look at is the way the IC channels are set up. Note that for the two CHPIDs used for the IC link, FE and FF, all of the partitions are shown as in the access lists. If these were legacy types of links (ICS, ICR), you would be restricted to having the Operating System partitions sharing the sender side CHPID (ICS) and the Coupling Facility partition would have exclusive access to the receiver side (ICR). That one way link has now been turned into a bidirectional link with peer mode. You could surely code the peer IC links in the same manner and construct as the ICS/ICR configuration (i.e. one side given to the Operating System partitions and the other to the Coupling Facility) since that is familiar, but you lose the value of peer mode and the ability to go both directions.

The following example shows how peer mode links can be established for good availability:

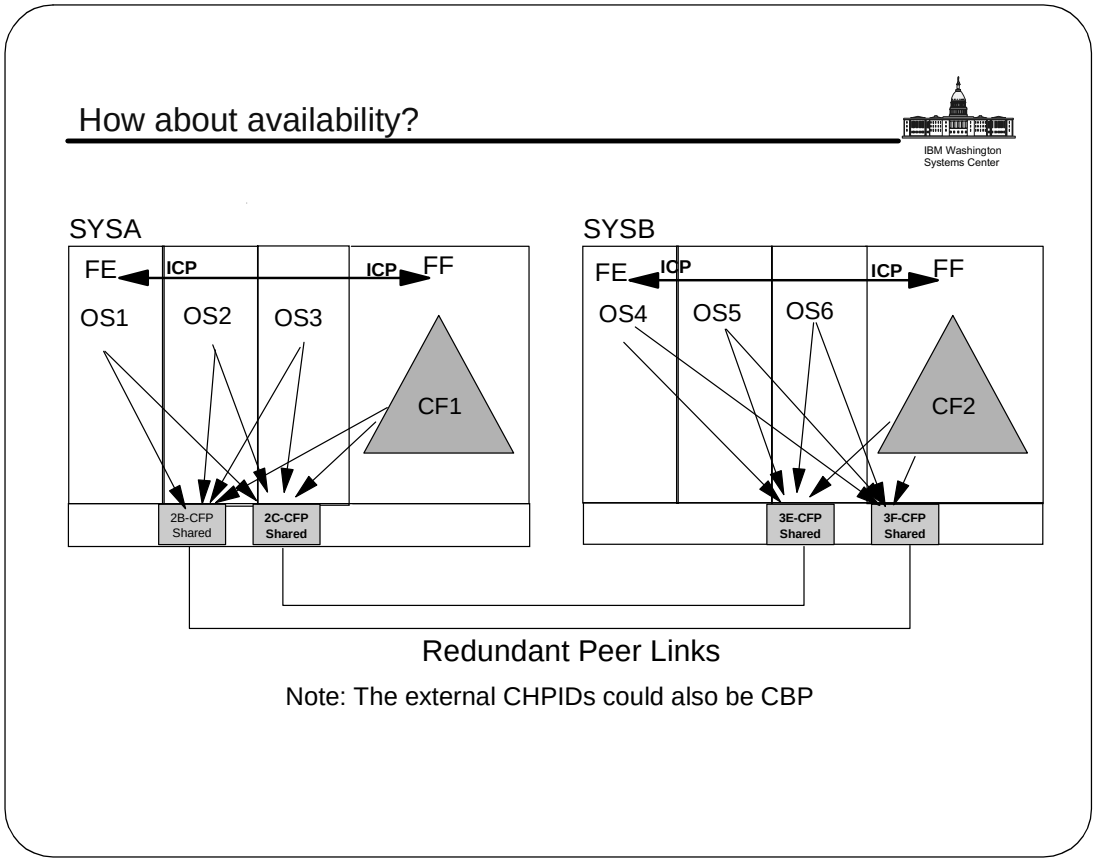


FIGURE 12. Peer Mode Links with Availability

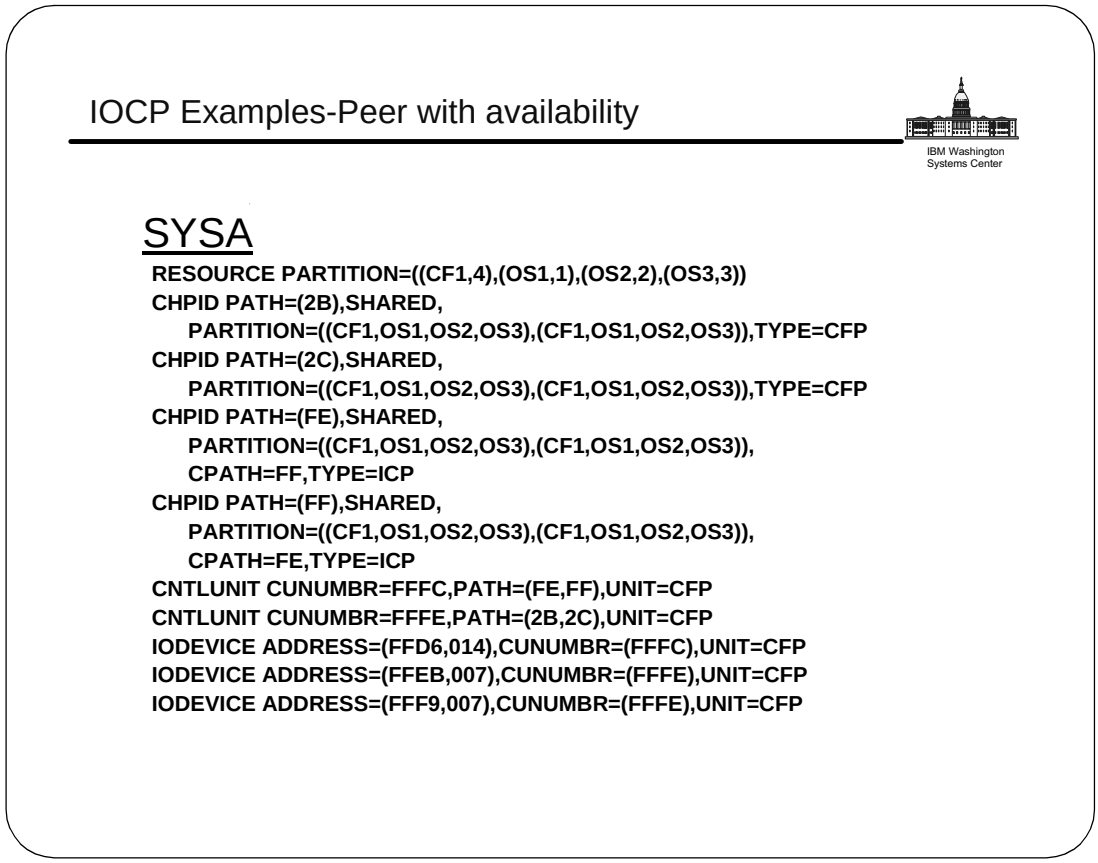


FIGURE 13. Peer Mode Links with Availability-IOCP SYSA

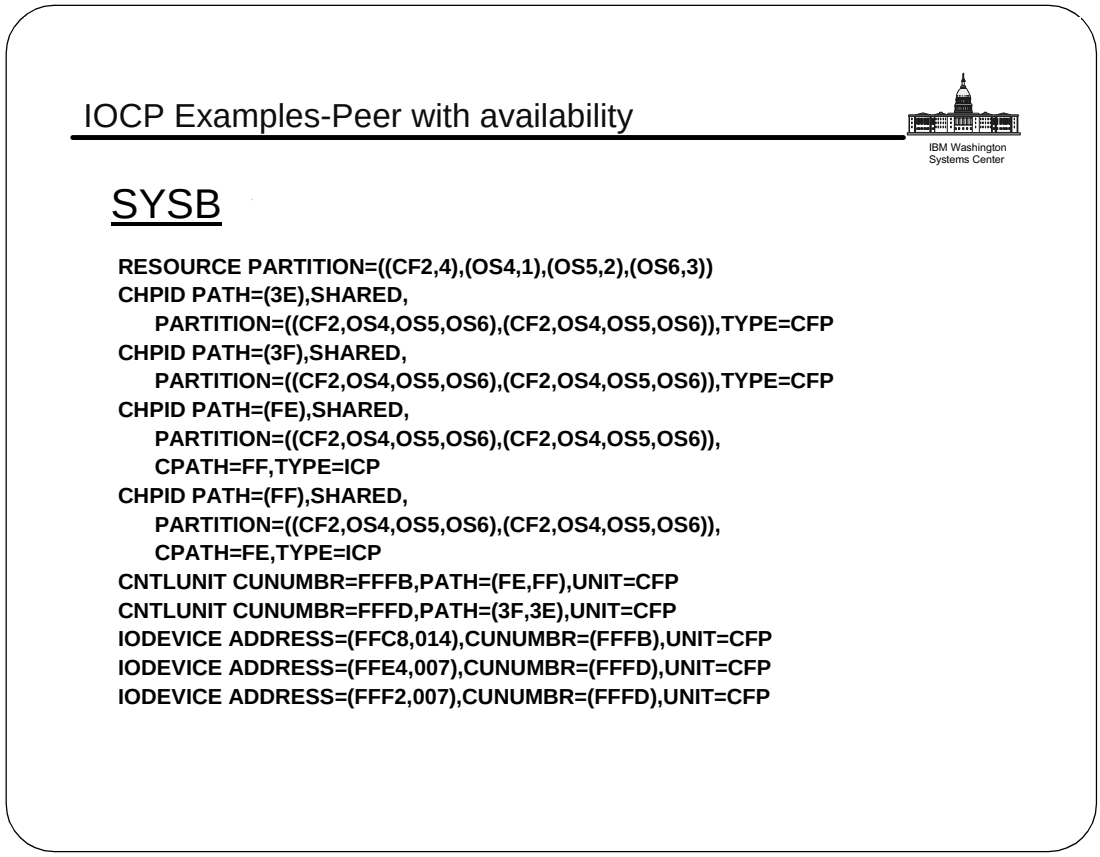


FIGURE 14. Peer Mode Links with Availability-IOCP SYSB

Now, there are a couple of things to notice. First of all, if this were a configuration using legacy links, one would need four links between the two machines for good availability. Secondly, you may have noticed that there are no redundant Internal Coupling links. Internal Coupling Links use the internal structure of the processor to connect and a failure of an Internal Coupling Link channel would mean that the internal workings of the processor are in a potentially non-functioning state. Thus, there is no value in coding redundant Internal Coupling Links for availability. You might, however, need more than one link for performance reasons. Secondly, notice what really changes in the IOCP. Only an additional set of IODEVICEs is required for the additional link added.

This is a good point to discuss the implications of Systems Managed CF Structure Duplexing. Systems Managed CF Structure Duplexing requires a connection between cooperating Coupling Facility Partitions. The peer mode links shown in the preceding examples can be used to satisfy this requirement since the participating Coupling Facility Partitions have access to a common link. That would not be the case for legacy links.

The following will focus on a nice use of Internal Coupling Links operating in peer mode. For this, let’s focus on the following example where it is desired to run two separate sysplexes in the same footprint using IC channels. One sysplex will consist of OS1, OS2, and CF1. The other will consist of OSA, OSB, and CFA. Now, it is only an assumption that

these are separate plexes. They could all be in the same plex if desired but this would obviously have some availability issues.

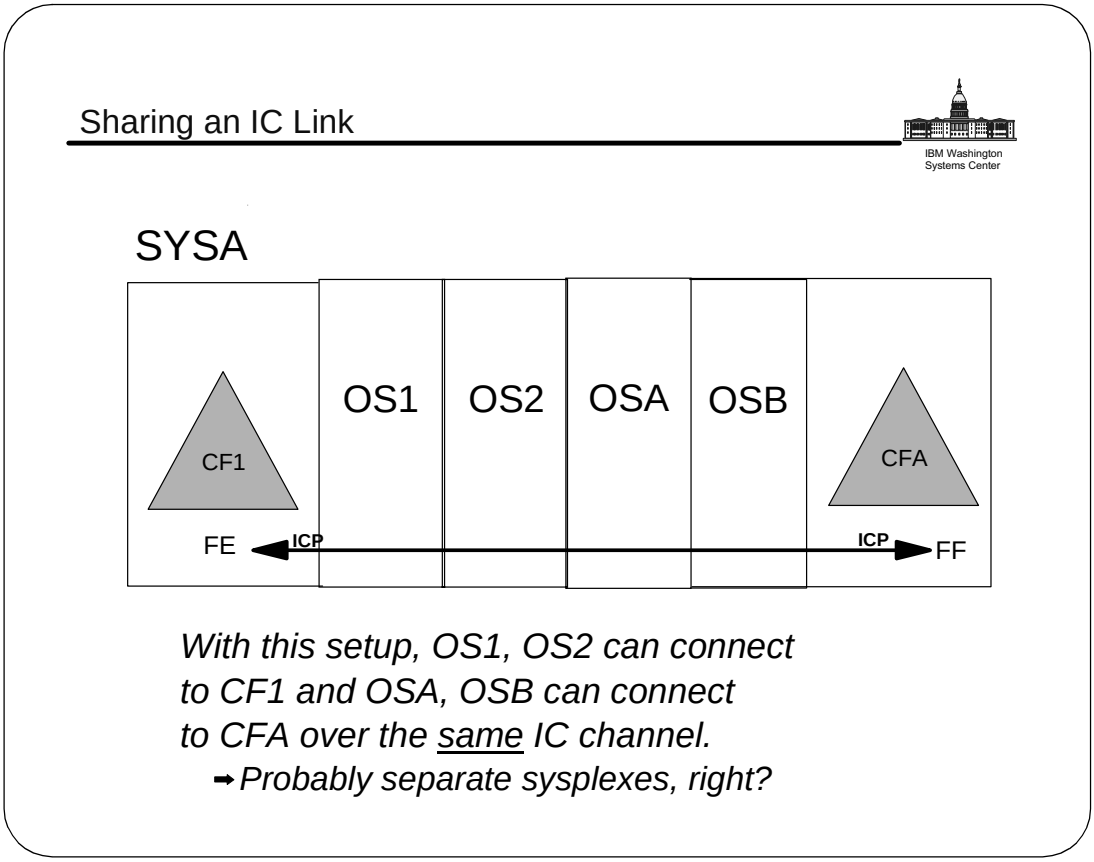


FIGURE 15. Peer Mode IC Links

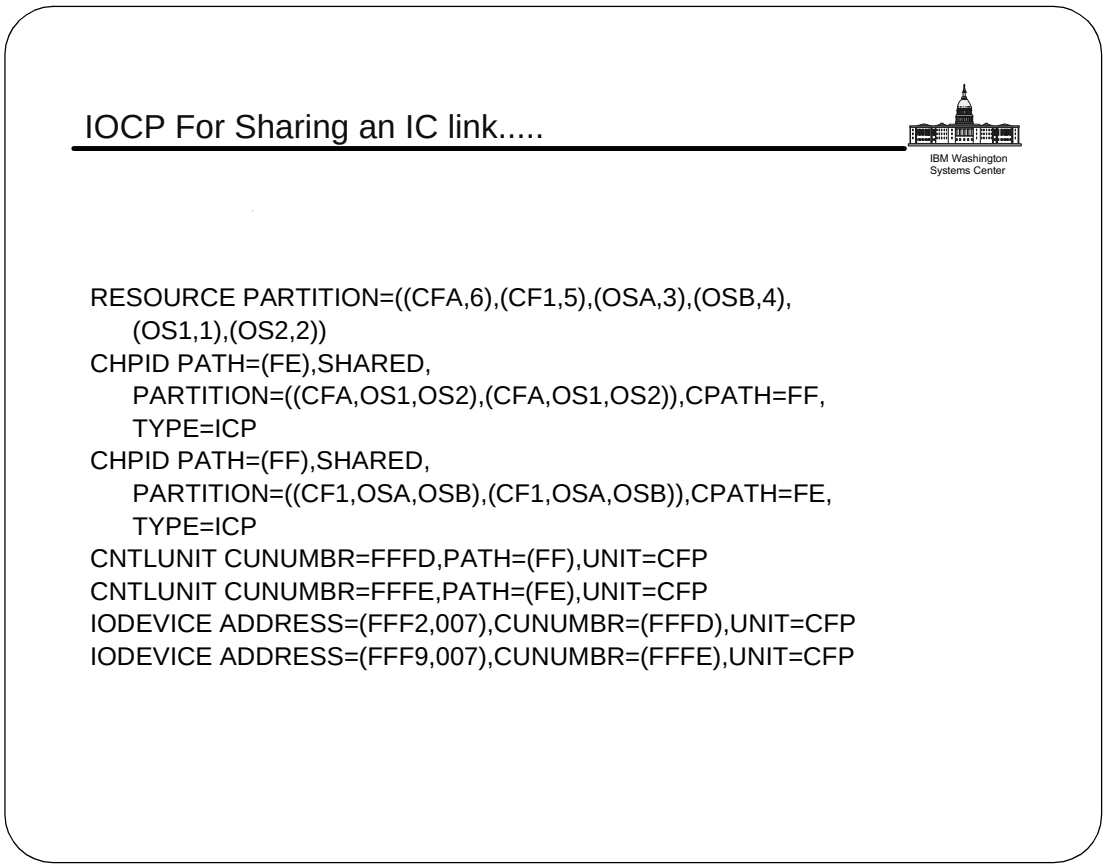


FIGURE 16. Peer Mode IC Links - IOCP

Why in the world would one want to do this? Aren't Internal Coupling Channels free? True, but recall that an Internal Coupling Channel requires two real CHPID addresses. These CHPIDS must come from the available 256 CHPIDs that are the limit on z900 and z800 machines. CHPID numbers can be a scarce commodity and one should use them sparingly.

In the example, I have only shown the structure inside the machine whereas normally one would have external links in the configuration as well.

5.0 Restrictions and Rules

The following figure illustrates the Restrictions and Rules associated with peer mode and I will follow this up with some examples to illustrate the points. For the examples, I will use the z900 but this also applies to the z800.

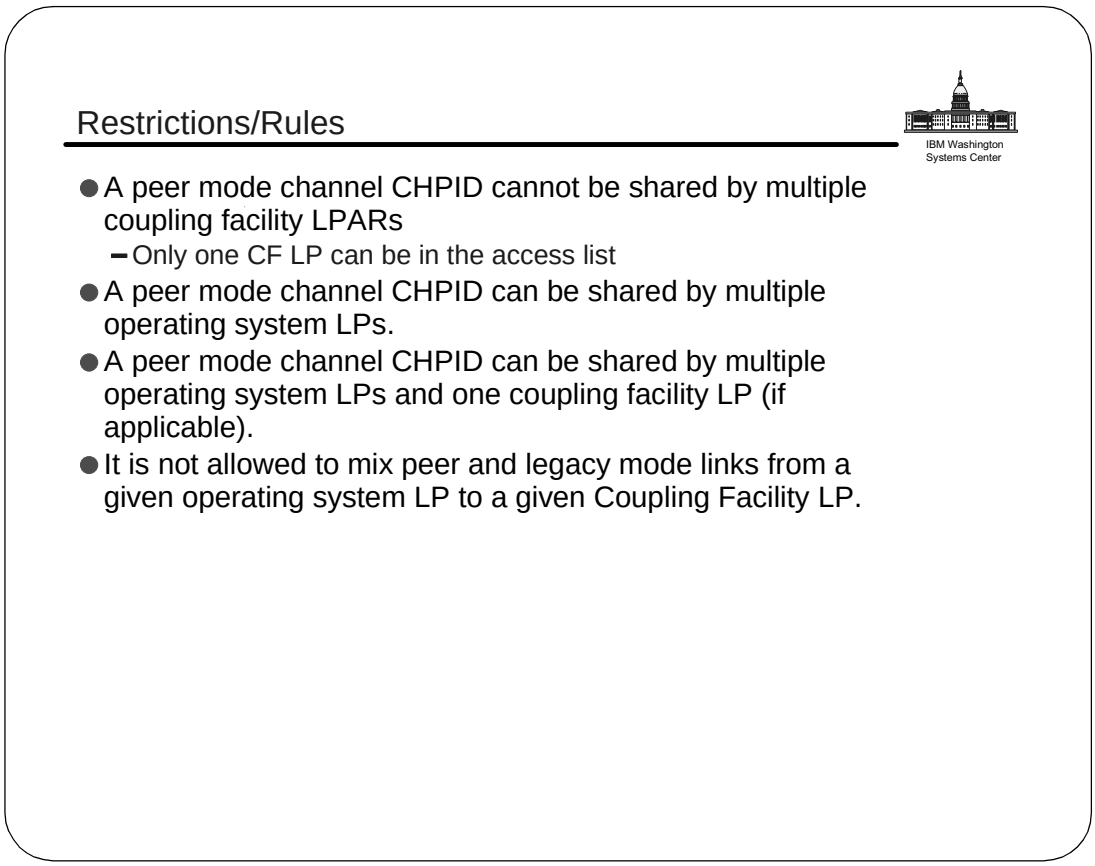


FIGURE 17. Peer Mode Restrictions/Rules

These seem clear enough but let’s illustrate with some examples.....

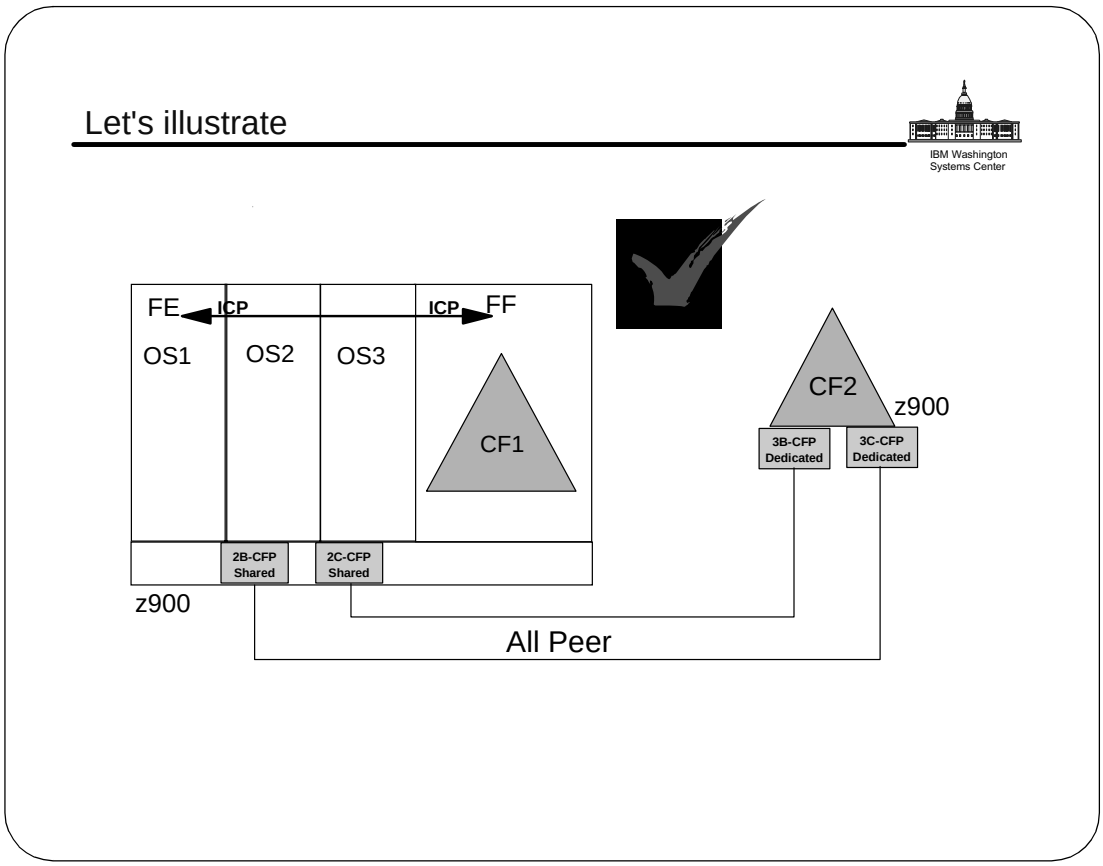


FIGURE 18. Restrictions and Rules Example 1

The above is an example of the correct use of peer mode links as has been discussed up to this point.

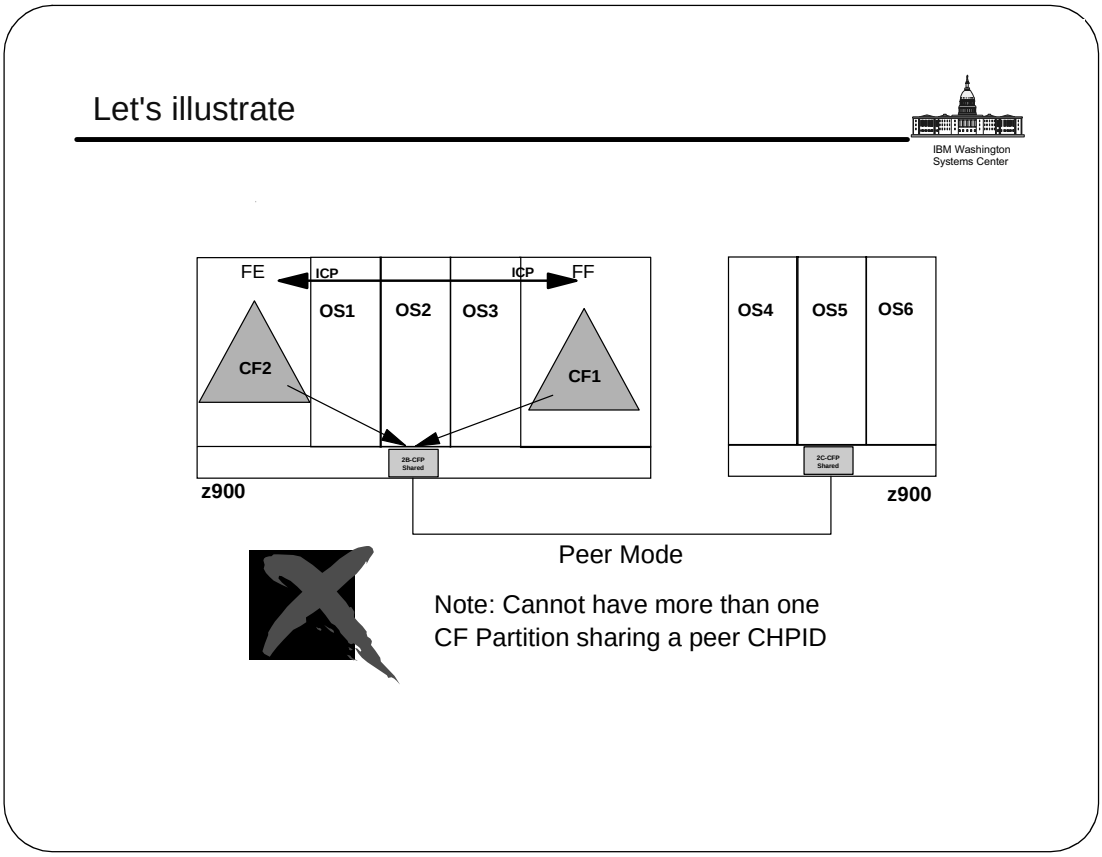


FIGURE 19. Restrictions and Rules Example 2

The problem in Figure 19 is that we are trying to allow two Coupling Facility Partitions to share the same CHPID. This is not allowed. To put this in HCD and IOCP terminology, you can only have one Coupling Facility Partition in the access list for a peer mode CHPID.

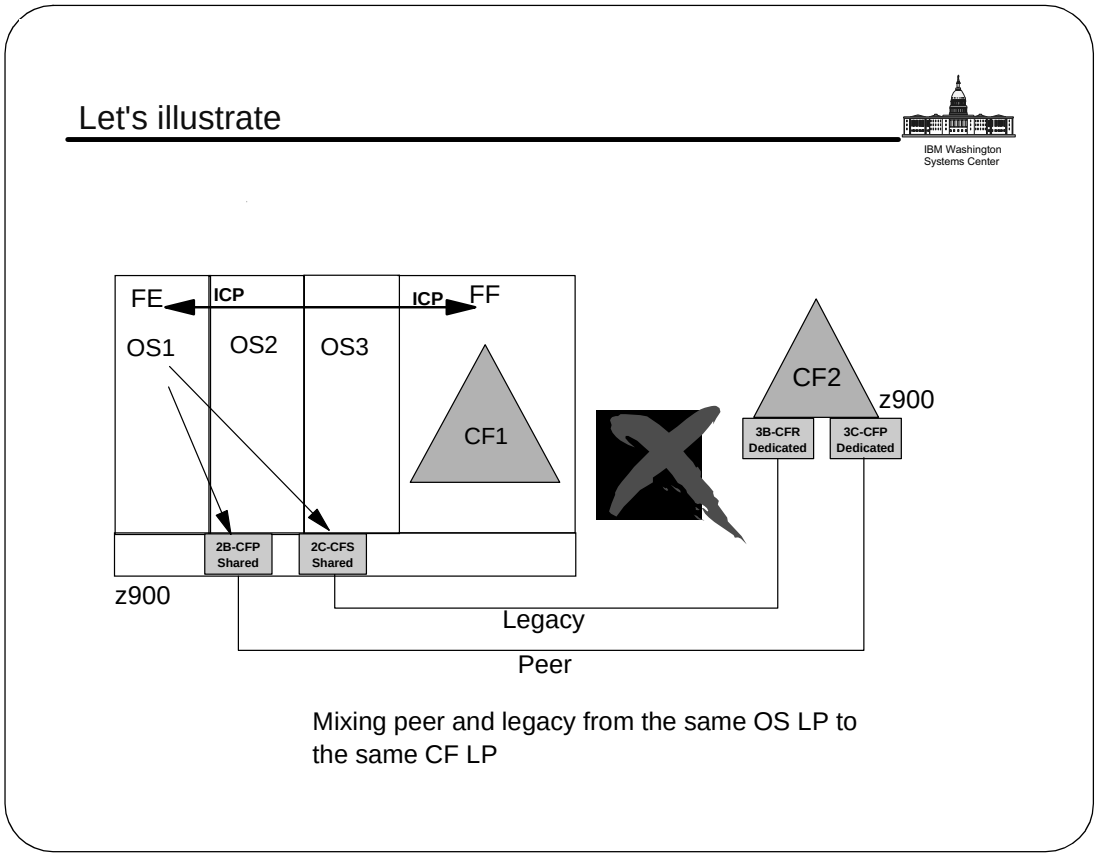


FIGURE 20. Restrictions and Rules Example 3

As noted, the problem in Figure 20 is that we are attempting to mix peer and legacy links between the same Operating System Partition and the same Coupling Facility Partition. The reason one cannot do this has to do with the structure of the control unit definitions that are generated.

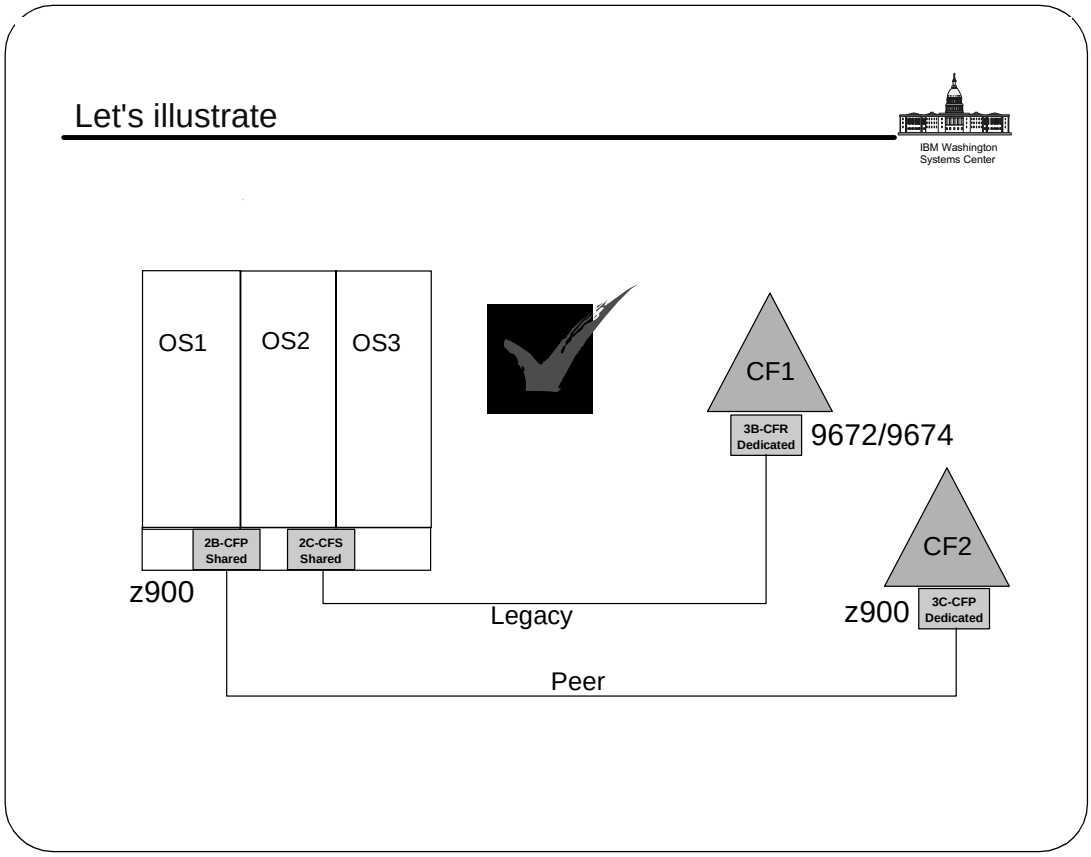


FIGURE 21. Restrictions and Rules Example 4

Figure 21 is not a case of mixing peer and legacy in a manner that is disallowed. Note that the target Coupling Facilities are different.

The following is a similar example of allowed mixing of peer and legacy links.

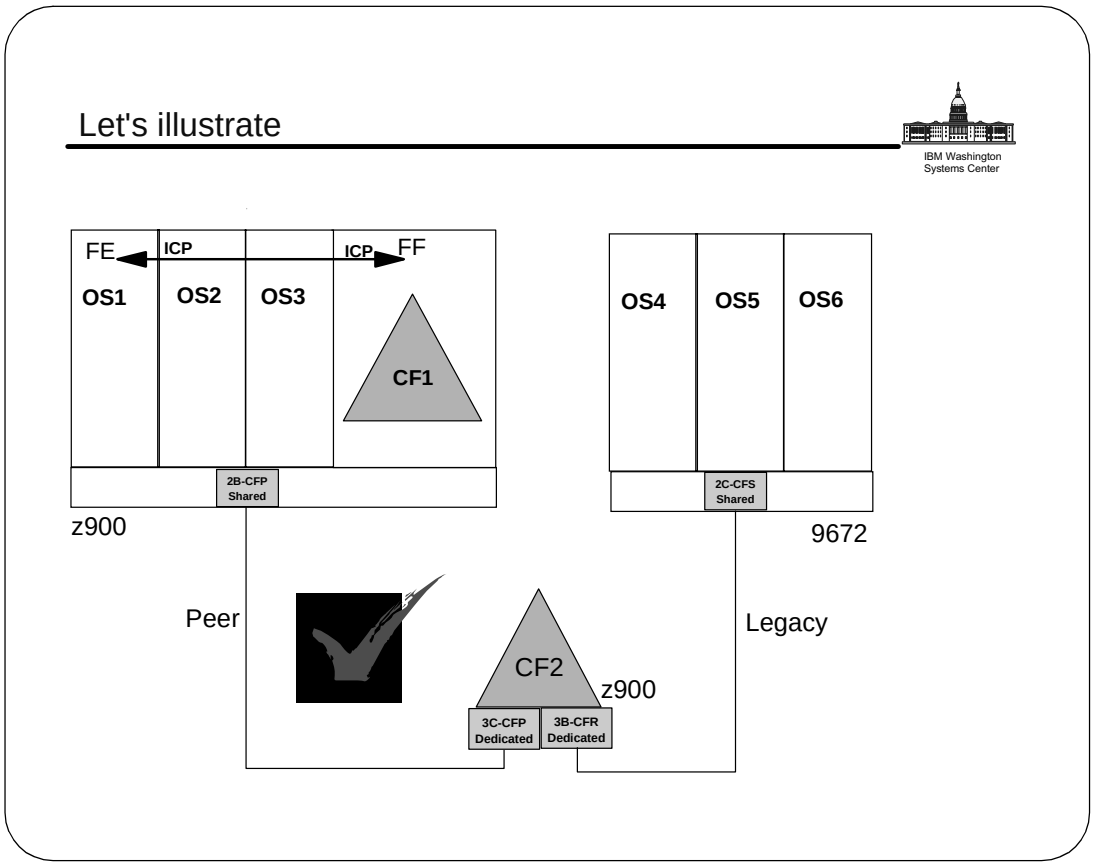


FIGURE 22. Restrictions and Rules Example 5

The situation is okay in Figure 22 since it is indeed allowed to have both peer and legacy links coming into a single Coupling Facility Logical Partition.....as long as they are not coming from the same Operating System Partition.

And, one more of the mixed peer and legacy which is fine....

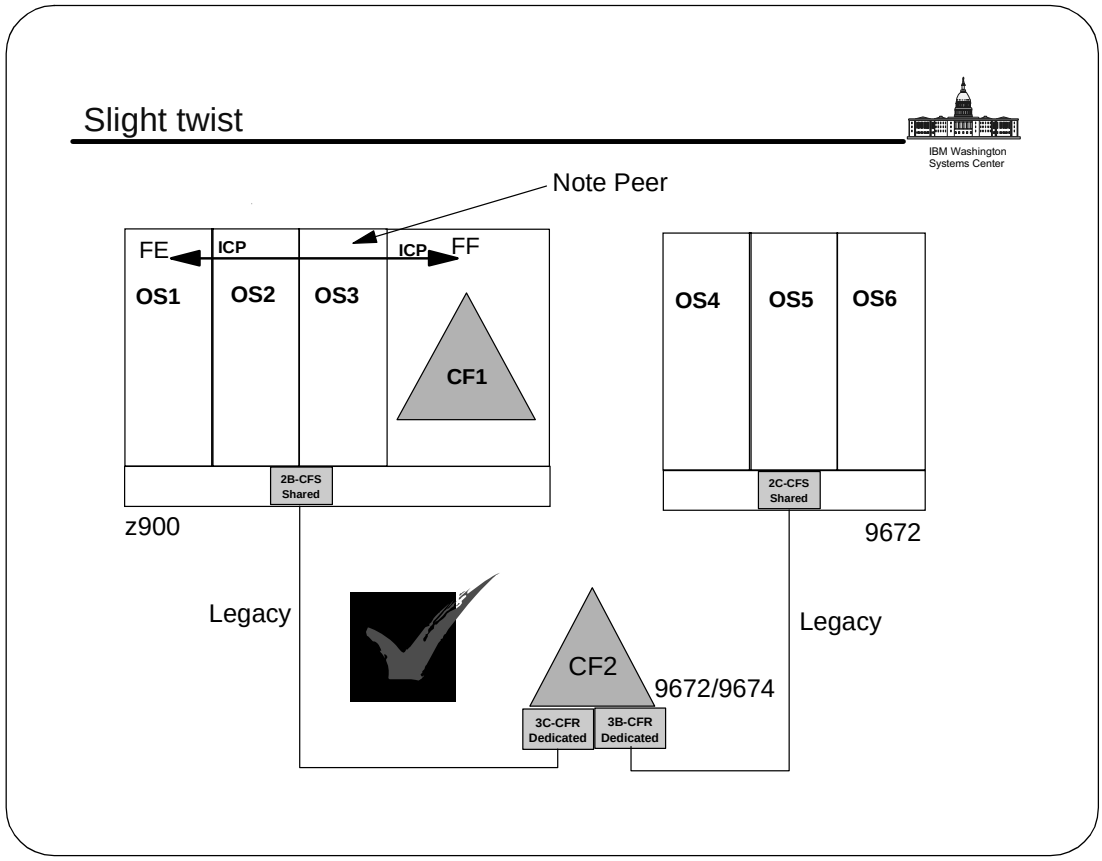


FIGURE 23. Restrictions and Rules Example 6

Again, Figure 23 shows a mixture of peer and legacy but it does not violate any of the rules mentioned before.

6.0 Summary and Conclusions

It is hoped that the previous discussions have helped de-mystify the setting up of peer mode coupling links and will help avoid some of the confusion. However, the Advanced Technical Support staff at the Washington Systems Center is always ready to help clarify and explain should there be any questions that might have been missed in the discussions.