Cheat Sheet #66 zTidBits zEnterprise z196 I/O Subsystem

*The z/Architecture specifies an I/O subsystem to which all I/O processing is offloaded.
- This is a *significant contributor* to the performance and availability of the system, and it strongly distinctive with the architectures of other servers.
- The z196 I/O subsystem direction is evolutionary, expanding on development from the z10 EC.
- It is based on I/O cages and I/O drawers, a *new* companion to I/O cages.
- I/O cages and I/O drawers house I/O cards, which are connected to the CPC through I/O buses.
- The I/O subsystem is supported by an I/O bus identical to the z10 EC one, and includes the InfiniBand infrastructure (replacing the self-timed interconnect features found in the prior System z servers).
- This infrastructure is designed to reduce overhead and latency and provide increased data throughput.

*As with its predecessors, the z196 server has **a dedicated subsystem** to manage all I/O ope
- This is known as the **channel subsystem (CSS),** it is composed of:

HSA: Hardware System Area (HSA) is a reserved part of the system memory containing the I/O configuration, and it is used by I/O processor or SAPs. On the z196 a fixed amount of 16 GB is reserved, which is not part of the customer purchased memory. This provides for greater configuration flexibility and higher availability by eliminating some planned and pre-planned outages.

SAP: System Assist Processor (SAP) is a specialized processor that uses the installed PU cores. Its role is to offload I/O operations and manage channels and the I/O operations queues. It relieves the other PUs of all I/O tasks, and therefore allowing them to be dedicated to application logic. An adequate number of SAP processors is automatically defined, depending on the number of installed books. These are part of the base configuration of the server. See #17 zTidBits (Free Engines - SAPs)
Note: SAPs are never part of an LPAR configuration.

SAP was known previously as an I/O Processor (IOP).

Model	Books	Installed PUs	Standard SAPs
M15	1	20 (1 x 20)	3
M32	2	40 (2 x 20)	6
M49	3	60 (3 x 20)	9
M66	4	80 (4 x 20)	12
M80	4	96 (4 x 24)	14

NOTE: Additional SAPs can be characterized depending on workload, although in most runtime environments the standard number will usually suffice.

A SAP can execute well over 100,000 I/O OPS / sec.

Three subchannel sets are available to z196, while two subchannel sets are on z10 and z9.

Channels: Channels are small processors that communicate with the I/O control units (CUs). They manage the data transfer between memory and the external devices. Channels are contained in the I/O card features (i.e. FICON, ESCON, OSA, etc.).

Channel path: Channel paths are the means by which the channel subsystem communicates with the I/O devices. Due to I/O virtualization, multiple independent channel paths can be established on a single channel, allowing sharing of the channel between multiple logical partitions, with each partition having a unique channel path.

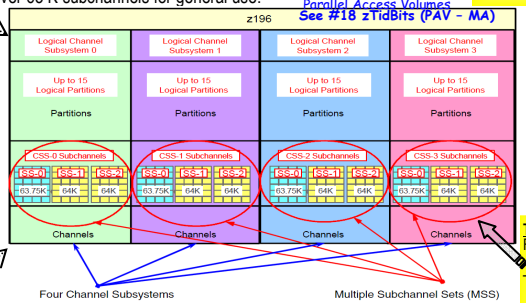
Subchannels: Subchannels appear to a program as a logical device and contain the information required to perform an I/O operation. One subchannel exists for each I/O device addressable by the channel subsystem. A third subchannel set is *new* with the z196.

NOTE: Do not confuse the multiple subchannel set (MSS) functionality with multiple channel subsystems .In most cases, a subchannel represents an addressable device. For example, a disk control unit with 30 drives uses 30 subchannels (for base addresses), and so forth. An addressable device is associated with a device number and the device number is commonly (but incorrectly) known as the device address. Subchannel numbers (including their implied path information to a device) are limited to four hexadecimal digits by the architecture (0x0000 to 0xFFFF). Four hexadecimal digits provide 64 K addresses, known as ***a set***. IBM has reserved 256 subchannels, leaving over 63 K subchannels for general use.

The processor architecture allows for *sets* of subchannels (addresses), with a current implementation of three sets. Each set provides 64 K addresses. Subchannel set 0, the first set, still reserves 256 subchannels for IBM use. Each of subchannel sets 1 and 2 provides the full range of 64 K subchannels. In principle, subchannels in either set could be used for any device-addressing purpose. However, the current z/OS implementation restricts Subchannel sets 1 and 2 to disk *alias* subchannels. Subchannel set 0 may be used for base addresses and for alias addresses.

The alias are used by Parallel Access Volumes

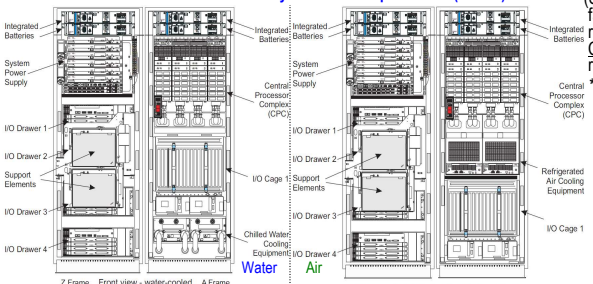
See #18 zTidBits (PAV - MA)



Four Channel Subsystems — Multiple Subchannel Sets (MSS)

*The z196 I/O subsystem design provides great flexibility, high availability, and excellent performance characteristics, as follows:
High bandwidth: The z196 uses InfiniBand as the internal interconnect protocol to drive ESCON and FICON channels, OSA ports, and ISC-3 coupling links. As a connect protocol, InfiniBand supports InfiniBand coupling (PSIFB3) with a link rate of up to 6 GBps.
Connectivity options :T he z196 can be connected to an extensive range of interfaces such as ESCON, FICON/Fibre Channel Protocol for storage area network connectivity, 10 Gigabit Ethernet, Gigabit Ethernet, and 1000BASE-T Ethernet for local area network connectivity, and ISC-3 coupling links.
Concurrent I/O upgrade: You may concurrently add I/O cards to the server if an unused I/O slot position is available. Additional I/O cages can be installed in advance to provide greater capacity for concurrent upgrades.
Concurrent I/O drawer upgrade: Additional I/O drawers can be installed concurrently without preplanning.
Dynamic I/O configuration Dynamic I/O configuration supports the dynamic addition, removal, or modification of channel path, contro units, and I/O devices without a planned outage.
Pluggable optics The FICON Express8 and FICON Express4 features have Small Form Factor Pluggable (SFP) optics to permit each channel to be individually serviced in the event of a fiber optic module failure. The traffic on the other channels on the same feature can continue to flow if a channel requires servicing.
Concurrent I/O card maintenance: Each I/O card plugged in an I/O cage supports concurrent card replacement in case of a repair action.

The z196 is a follow-on to the System z10 Enterprise Class (z10 EC).



*z196 also offers a new I/O infrastructure element, a companion to the I/O cage of previous systems, the I/O drawer.

I/O drawer: I/O drawers provide increased I/O granularity and capacity flexibility and can be concurrently added and removed in the field, an advantage over I/O cages, which also eases pre-planning. The z196 server can have up to four I/O drawers, two on the A frame & two on the Z frame. I/O drawers were *first offered* with the z10 BC and can accommodate up to eight I/O features, in any combination.

NOTE: The I/O drawer is five EIA units high and supports up to eight I/O feature cards. Combinations of /O drawers and I/O cages can be installed to support up to 72 I/O feature cards (with two I/O drawers and two I/O cages), or 84 I/O feature cards with RPQ 8P2506, supporting a third I/O cage. Installing an I/O drawer is **nondisruptive** and can be done concurrently in a running system. Each drawer supports *two I/O domains* (A and B) for a total of eight I/O card slots. Each I/O domain uses an IFB-MP card in the I/O drawer and a copper cable to connect to a Host Channel Adapter (HCA) fanout in the CPC cage.

I/O cage: The z196 has a CPC cage and, optionally, one I/O cage in the A frame. The Z frame can accommodate two additional I/O cages, bringing the total for the system to three. Adding the third I/O cage requires an RPQ. Each I/O cage can accommodate up to 28 I/O features, in any combination.

The fixed size of the HSA (16GB) eliminates planning for future expansion of the HSA because HCD/IOCP always reserves::
- Four channel subsystems (CSSs)
- Fifteen logical partitions in each CSS for a total of 60 logical partitions
- Subchannel set 0 with 63.75 K devices in each CSS
- Subchannel set 1 with 64 K devices in each CSS
- Subchannel set 2 with 64 K devices in each CSS

The HSA has sufficient reserved space allowing for dynamic I/O reconfiguration changes to the maximum capability of the processor.

The z196, z10 EC, and z9 EC support up to 4 CSSs, while the z10 BC and z9 BC support up to 2 CSSs.

Number of I/O feature cards	Number of I/O drawers	Number of I/O cages
1-8	1	0
9-16	2	0
17-24	3	0
25-32	4	0
33-36	1	1
37-44	2	1
45-52	3	1
53-60	4	1
45-56	0	2
57-64	1	2
65-72	2	2
73-84 (RPQ required)	0	3

ESCON channels The Enterprise Systems Connect channels support connectivity to ESCON disks, tapes, and printer devices. Historically, they represent the first use of optical I/O technology on the mainframe. They are much slower than FICON channels. FICON Express8 is the preferred technology. The maximum number of supported ESCON features is 16 (up to 240 ports) on the z196. RPQ 8P2505 will allow you to go beyond 16 features if necessary.

NOTE: The zEnterprise 196 is planned to be the last high end server to offer ordering of ESCON channels on new builds, migration offerings, upgrades, and System z exchange programs. Enterprises should begin migrating from ESCON to FICON..

Summary of supported features
- Up to 240 ESCON channels (up to 360 with an RPQ 8P2507)
- Up to 288 FICON Express4 channels (carried forward on upgrade only)
- Up to 288 FICON Express8 channels (up to 336 with an RPQ 8P2506)
- Up to 48 OSA-Express2 ports (when carried forward on upgrade only)
- Up to 96 OSA-Express3 ports
- Up to 48 ISC-3 coupling links
- Up to 32 InfiniBand coupling links (12x InfiniBand, 1x InfiniBand)

See #23 zTidBits (Multiplex/DeMultiplex)i

Fibre Connection (FICON) channels follow the Fibre Channel (FC) standard and support data storage and access requirements as well as the latest FC technology in storage and access devices. FICON support the following protocols:
- Native FICON, Channel-to-Channel (CTC) connectivity, and zHPF traffic to FICON devices such as disks, tapes, and printers in z/OS, z/VM, z/VSE, z/TPF, and Linux on System z environments.
- Fibre Channel Protocol (FCP) in z/VM and Linux on System z environments support connectivity to disks and tapes through Fibre Channel switches and directors. z/VSE supports FCP for SCSI disks only. The FCP channel can connect to FCP SAN fabrics and access FCP/SCSI devices.

NOTE: It is possible to choose any combination of the FICON Express8 and FICON Express4 features. Depending on the feature, auto-negotiated link data rates of 1, 2, 4, or 8 Gbps are supported (1, 2, and 4 for FICON Express4; 2, 4, and 8 for FICON Express 8). FICON Express8 provides significant improvements in start I/Os and data throughput.
Statement of Direction: The z196 will be the *last* server to support FICON Express4 features.

---

*The **Multiple Image Facility (MIF)** allows channel sharing among multiple LPARs and optionally shares any associated I/O devices configured to these shared channels.
-MIF also provides a way to limit the LPARs that can access a reconfigurable channel, spanned channel, or a shared channel to enhance security.
-With multiple LCSSs, the CSS provides an independent set of I/O controls for each logical channel subsystem called a CSS image.
-Each LPAR is configured to a separate CSS image in order to allow the I/O activity associated with each LPAR to be processed independently as if each LPAR had a separate CSS.
For example, each CSS image provides a separate channel image and associated channel path controls for each shared channel and separate subchannel images for each shared device that is configured to a shared channel.
-With MIF and multiple channel subsystems, shared and spanned channel paths can provide extensive control unit and I/O device sharing. MIF allows all, some, or none of the control units attached to channels to be shared by multiple LPARs and multiple CSSs.
-Sharing can be limited by the access and candidate list controls at the Channel Path ID (CHPID) level and then can be further limited by controls at the I/O device level.
For example, if a control unit allows attachment to multiple channels (as is possible with a 3990 control unit), then it can be shared by multiple LPARs using one or more common shared channels or unique unshared channel paths.

*You can configure a channel path as: 1. An unshared dedicated channel path to a single LPAR, 2. An unshared reconfigurable channel path that can be configured to only one LPAR at a time it can be moved to another LPAR within the same LCSS, 3. A shared channel path that can be concurrently used by the system z images or coupling facility partitions within the same LCSS to which it is configured.

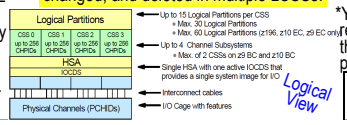
I/O Domain: *Each* I/O domain has up to 4 I/O feature cards of any type (ESCON, FICON, ISC or OSA). The I/O cards are connected to the IFB-MP card through the backplane board. Each I/O cage supports up to seven I/O domains and a total of 28 I/O card slots. .

InfiniBand is an industry-standard specification that defines a first-order Interconnection technology, which is used to interconnect servers, com-munications infrastructure equipment, storage, and embedded systems. InfiniBand is a fabric architecture that leverages switched, point-to-point channels with data transfers of up to 120 Gbps, both in chassis backplane applications and through copper and optical fiber connections.

The CSS structure offers the following:
Four logical channel subsystem (LCSSs)
– Each LCSS can have up to 256 channels defined.
– Each LCSS can be configured with one to 15 LPARs (cannot exceed 60 LPARs per system).

Spanned channels are shared among LPARs across Logical Channel Subsystems (LCSS). Spanning channels is the ability for the channels to be configured to multiple LCSSs, and be transparently shared by any or all of the configured LPARs without regard to the LCSS o which the partition is configured.

Channel paths, control units, and devices can be dynamically added, changed, and deleted in multiple LCSSs.



*InfiniBand (IFB) is exploited by the z196 server.
Internally, in the server, cables from the CPC cage to the I/O cages and I/O drawers carry the IFB protocol.
- For external usage, Parallel Sysplex InfiniBand (PSIFB) links are available which can *completely replace* the ISC-3 and ICB-4 offerings available on previous servers. They are used to interconnect System z servers in a Parallel Sysplex.

***Coupling links** are used in the z196 Parallel Sysplex cluster configurations of System z servers.The links provide high-speed bidirectional communication between members of the sysplex.The z196 supports internal coupling links for memory-to-memory transfers, 12x InfiniBand for distances up to 150 meters (492 feet), and InterSystem Channel-3 (ISC-3) and InfiniBand for unrepeated distances up to 10 km (6.2 miles).
***HiperSockets** function is an integrated function of the z196 that provides users with attachments to up to 32 high-speed *virtual* local area networks with minimal system and network overhead.
HiperSockets is a function of the virtualization Licensed Internal Code (LIC) and performs memory-to-memory data transfers in a totally secure way. HiperSockets eliminates having to utilize I/O subsystem operations and having to traverse an external network connection to communicate between logical partitions in the same z196 server. HiperSockets offers significant value in server consolidation by connecting virtual servers.
Statement of Direction: The z196 will be the last server to offer ISC-3 features.