10-123v1

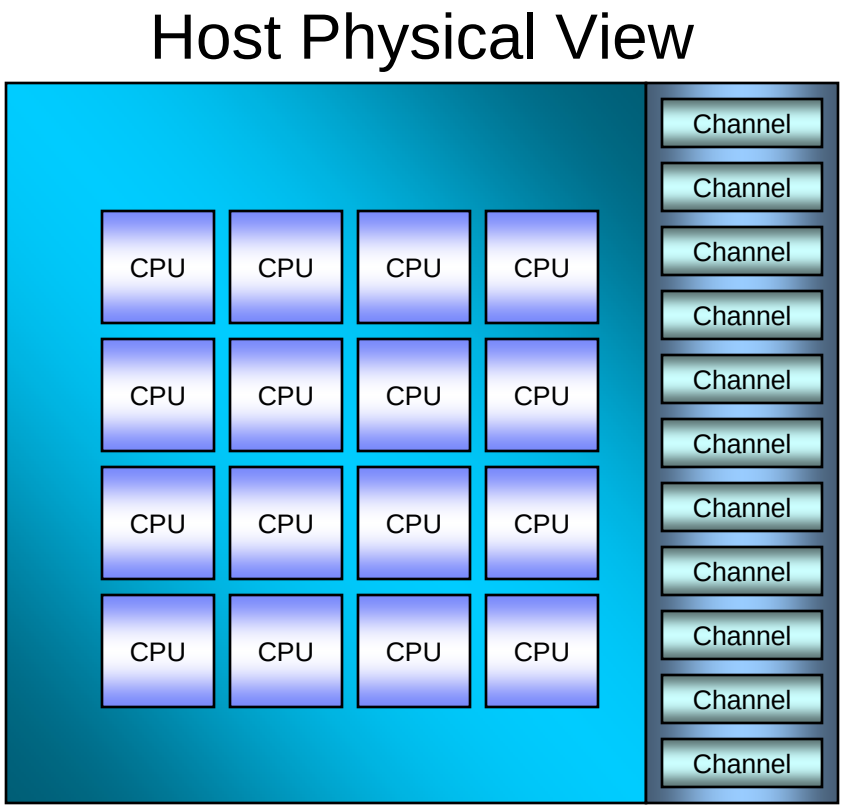
Living in a Mainframe World

Why timing is everything !

Lou Ricci

March 31, 2010

Virtualization is EVERYWHERE

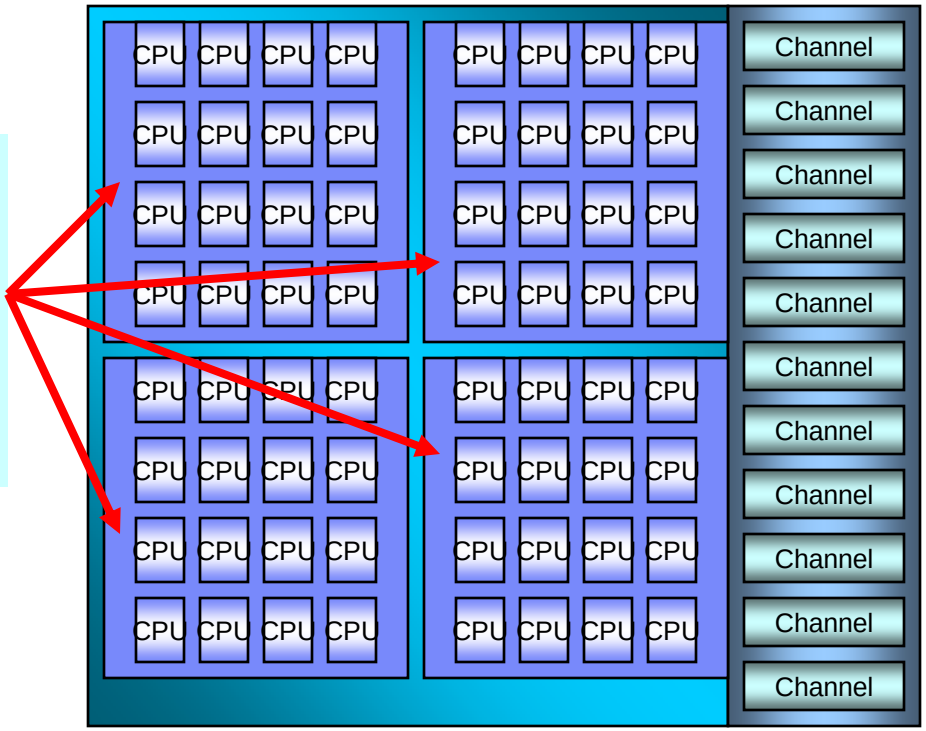


Virtualization is EVERYWHERE

Each purple box is a Logical Partition (LPAR)

Each LPAR has no knowledge of the others

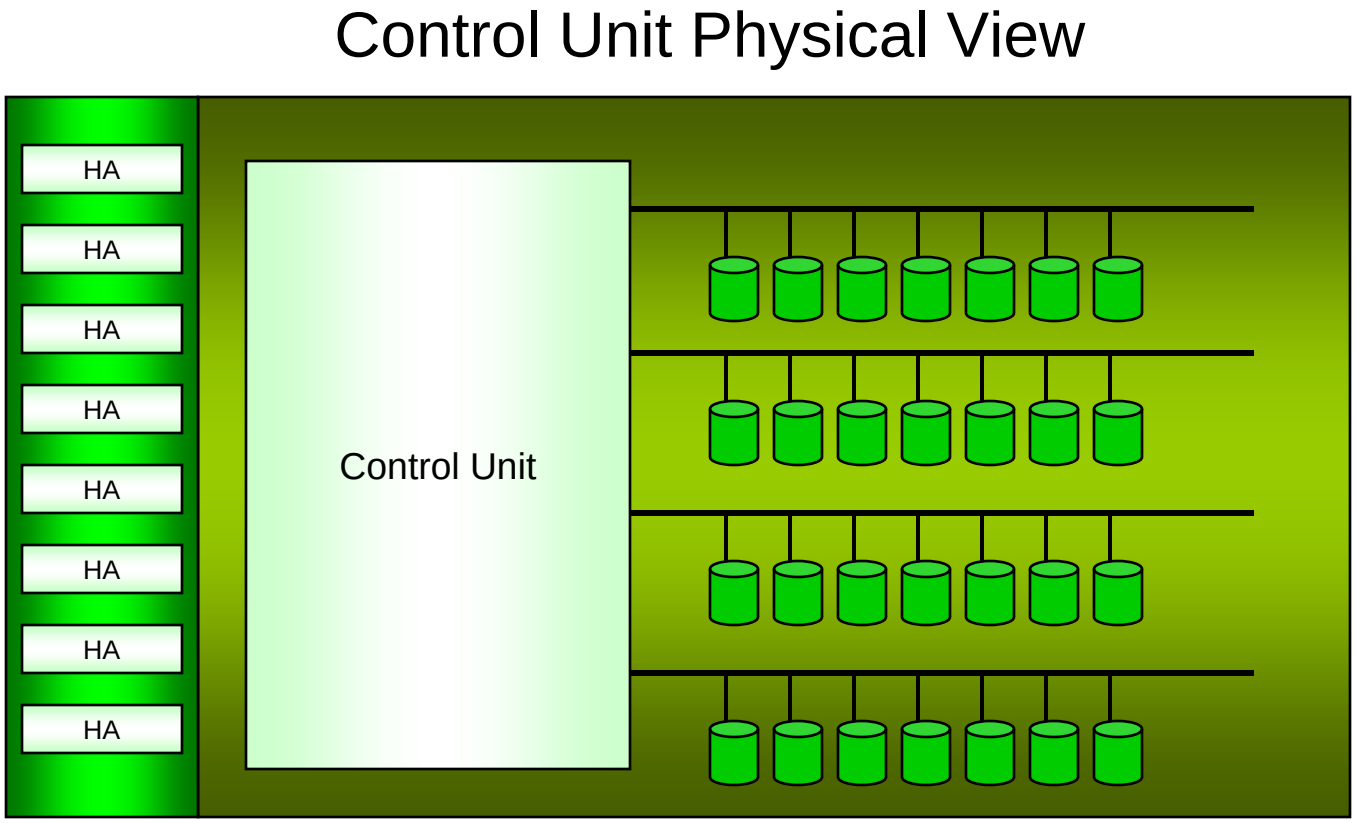
Host Virtual View



Up to 256 LPARs* can be instantiated within a physical system

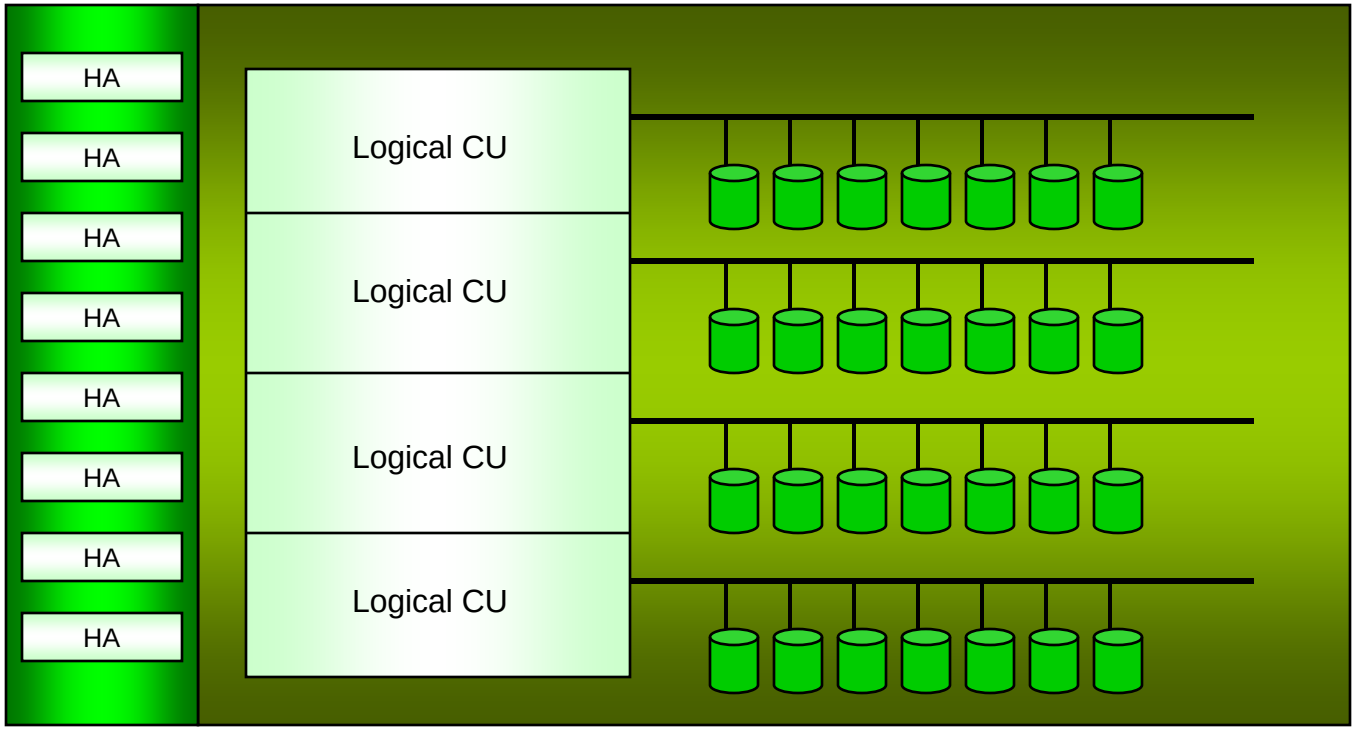
*Some implementations limit the number of Logical Partitions to a much smaller number

Virtualization is EVERYWHERE



Virtualization is EVERYWHERE

Control Unit Logical View



Up to 256 Logical Control Units (LCU)* can be instantiated within a physical Control Unit

Each LCU can address 256 Devices

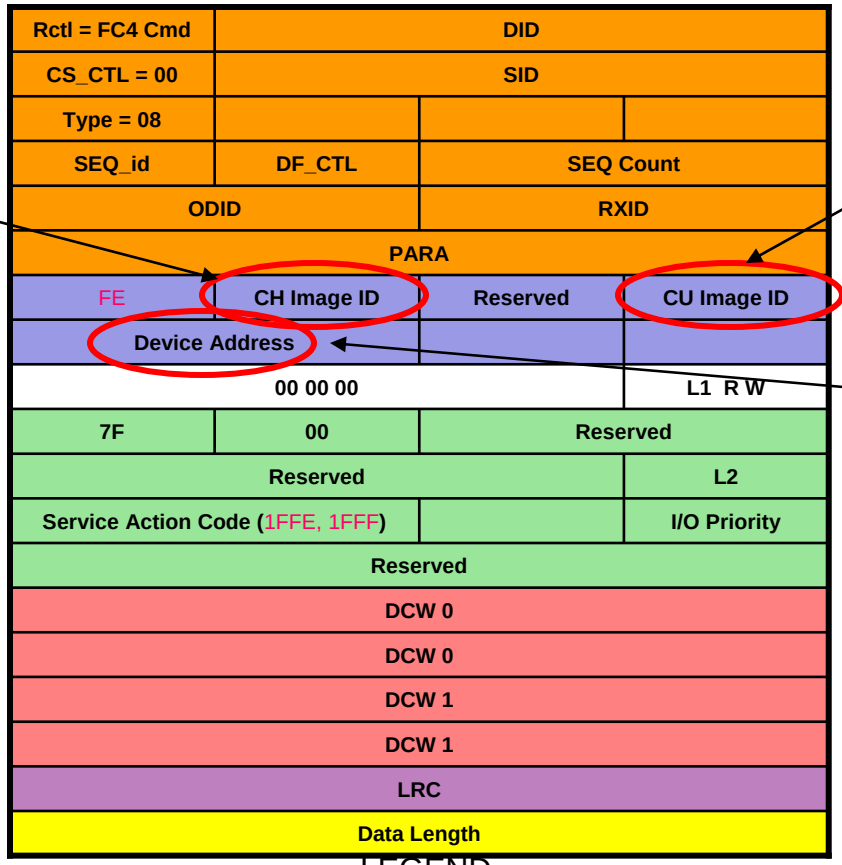
*Some implementations limit the number of LCUs and device to a much smaller number

Transport mode Cmd Frame

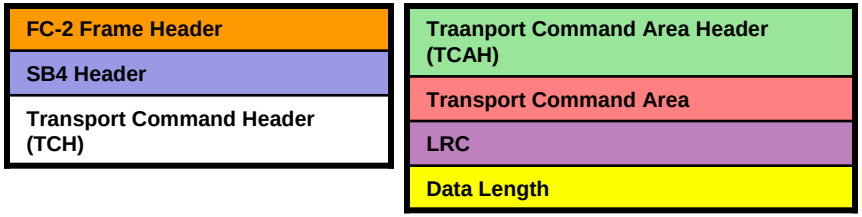
Identifies the LPAR

Identifies the Logical Control Unit

Identifies the Device within that LCU



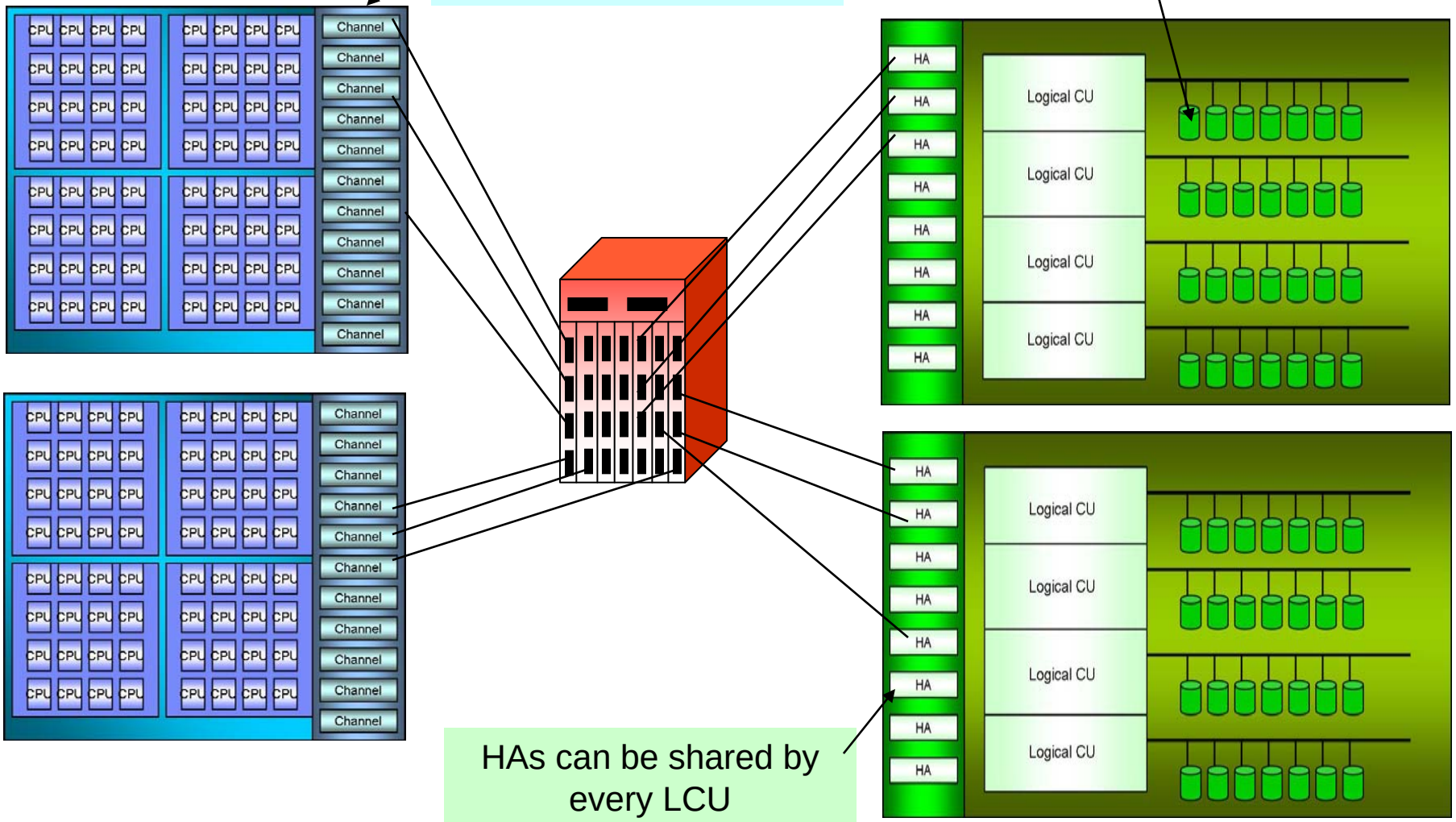
LEGEND



Share and Share Alike

Channels can be shared by every LPAR

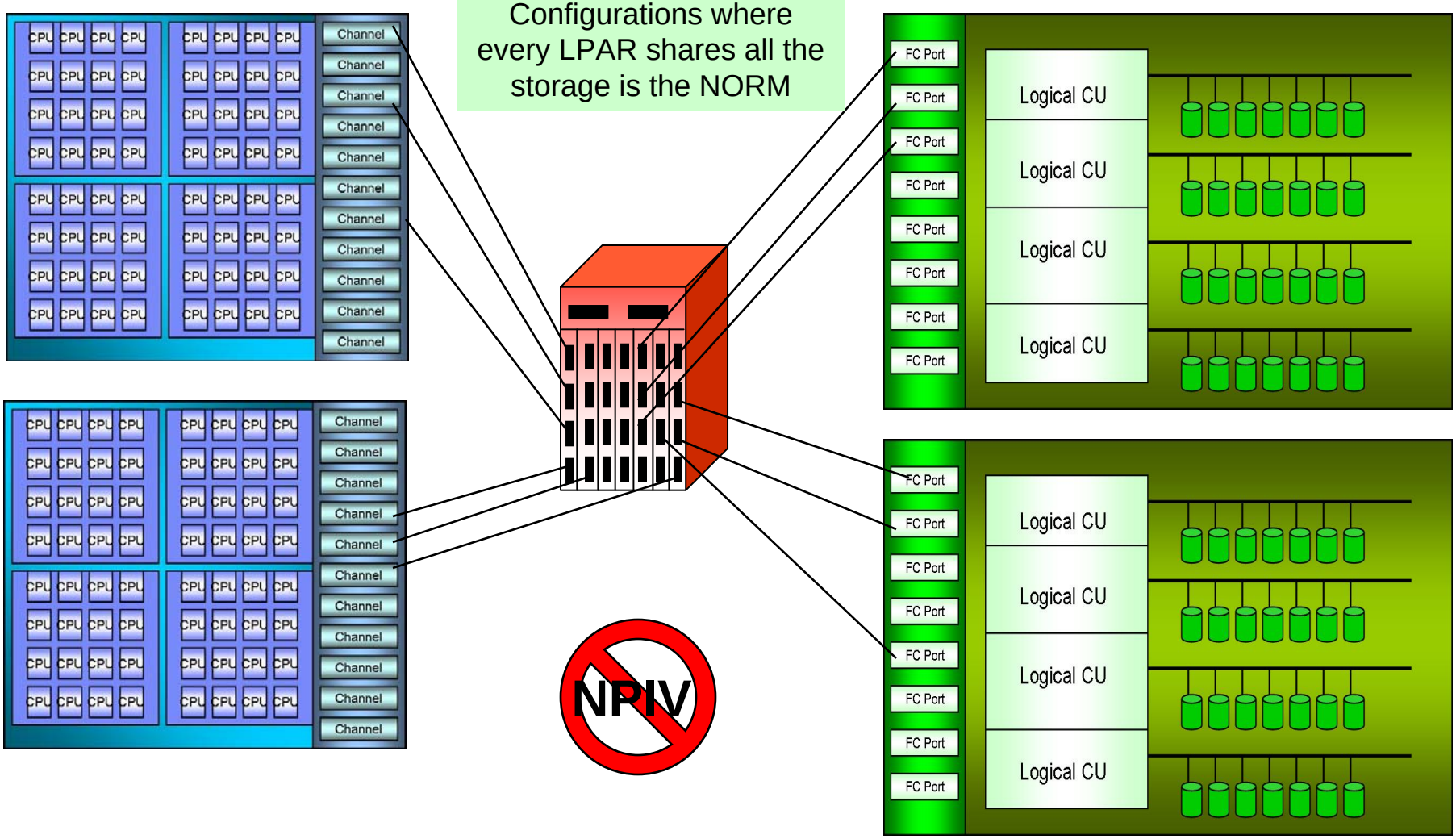
Devices can be shared by every LPAR



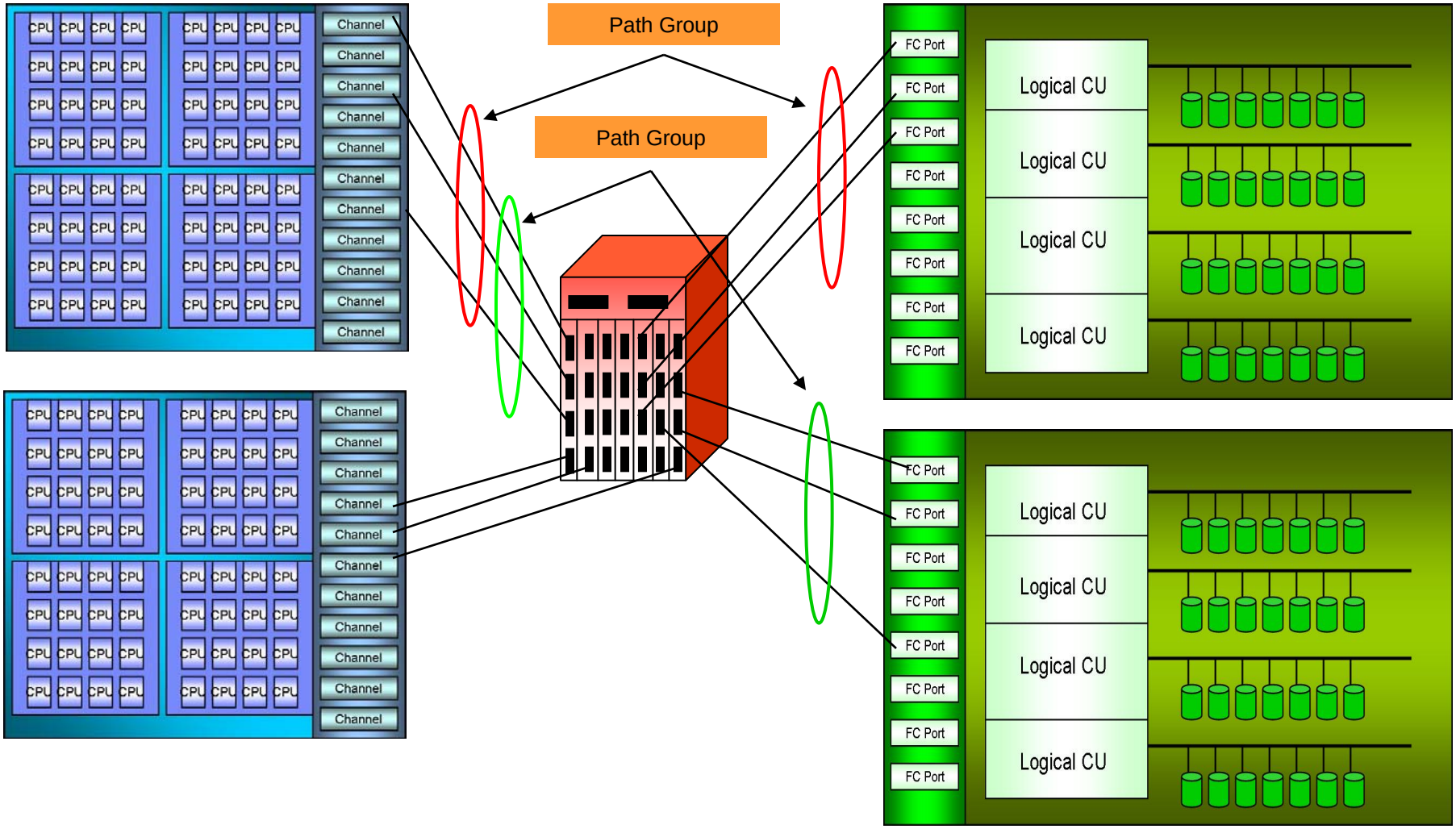
HAs can be shared by every LCU

Share and Share Alike

Configurations where every LPAR shares all the storage is the NORM



More Than One Way to Get From Here to There



More Than One Way to Get From Here to There

- **Each LPAR can have up to 8 physical paths to each device**
- **The OS “tells” the logical CU about the path group**
- **The logical CU.device maintains state information about the device on a LPAR.Logical CU.Device basis**
- **I/O operations can start on one physical path, and finish on a different one**
  - Dynamic Path Reconnection
  - Example: Read commands are sent from Channel 1 to CU/Device via HA a (this opens an exchange pair between Channel 1 and HA a
  - CU/Device detects a cache miss on the Nth read command
  - CU/Device sends “disconnect” status to the Channel 1 via HA on the existing exchange pair
  - CU stages data from device to cache
  - CU sends “reconnect” status to a Channel 2 from HA b on a new exchange
  - Read operation continues on the “new” physical path

Physical Paths are Virtualized Too

- **Each LPAR that wants to communicate with each logical CU must establish a "Logical Path"**
  - Done with SB-4 Link Level command "Establish Logical Path"
  - Creating or removing a Logical Path performs a "System Reset" at the LCU
    - Resets all state information, including the dynamic path information for the LPAR.Logical CU for ALL devices on that LCU
  - The first I/O to after a "System Reset" gets rejected
    - Sense information indicates "Resetting Event"

Resetting Events – Oh the Pain !!!

- **A link failure for more than LP_TOV removes all logical paths**
  - LP_TOV is typically 1.5 to 2 Seconds
- **Avoids data integrity problems when cables are pulled / swapped / replugged**
- **When OS encounters a "resetting event" it performs "path validation" for every device on the effected link**
  - Can result in Millions of I/Os in a very short time
  - Production I/O stops until "path validation" is complete
  - Millions of I/Os can swamp the fabric and/or the CU
  - Can result in measurable loss of revenue for the user

So Here's Why Timing is EVERYTHING

- **SB-4 allows for short duration link failures without loosing logical paths**
  - Avoids unnecessary path validation
  - LP_TOV picked low enough to prevent undetected cable swaps
- **Internal recovery actions in channels and/or HAs can eat much of the LP_TOV duration**
  - Light is down during internal recovery
- **Link Initialization time must be significantly shorter than LP_TOV**
  - Channels and HAs disable link speed auto negotiation after recovery
    - It too often takes too long !

Thank You