Agilent Technologies

PCI Express Primer

September 26, 2002

presented by:

Dave Dzatko, MindShare

Perry Keller, Agilent Technologies

Introduction

- PCI Express is a new I/O interconnect standard for a wide range of systems
- This presentation will provide:
  - A basic understanding of key terms and concepts
    - A detailed tutorial would take much more time
    - For more detailed training visit WWW.Mindshare.com

Agenda

- PCI Express Feature Set
- PCI Express Overview
  - Platform Examples
- Architecture
  - Software Layer
  - Transaction Layer
  - Data Link Layer
  - Physical Layer
- Mechanical Concepts
- Performance

PCI Express Feature Set

- **Support for multiple market segments**
  - **Desktop, Mobile, Server, Workstation, Communications and Embedded**
- **Low Cost/High Volume**
  - **Full serial = Low pin count**
  - **Cost at or below current PCI at the system level**
  - **Utilize high volume Si, boards, and connectors**

PCI Express Feature Set

- **Compatible SW Model**
  - **Boot Existing OS's without modification**
  - **PCI Compatible Configuration, Power Management and Device Driver Interfaces**
- **Scalable Performance**
  - **Via frequency and width**
  - **High Bandwidth per pin**
  - **Low Latencies**
  - **Suitable for applications over next 10 years**

PCI Express Feature Set

- Support for Multiple Connection Types
  - Chip-to-chip, board to board via connectors, docking stations, new form factors
- Platform may include multiple switches with switch to switch interconnect
- 256 endpoints per domain

PCI Express Feature Set

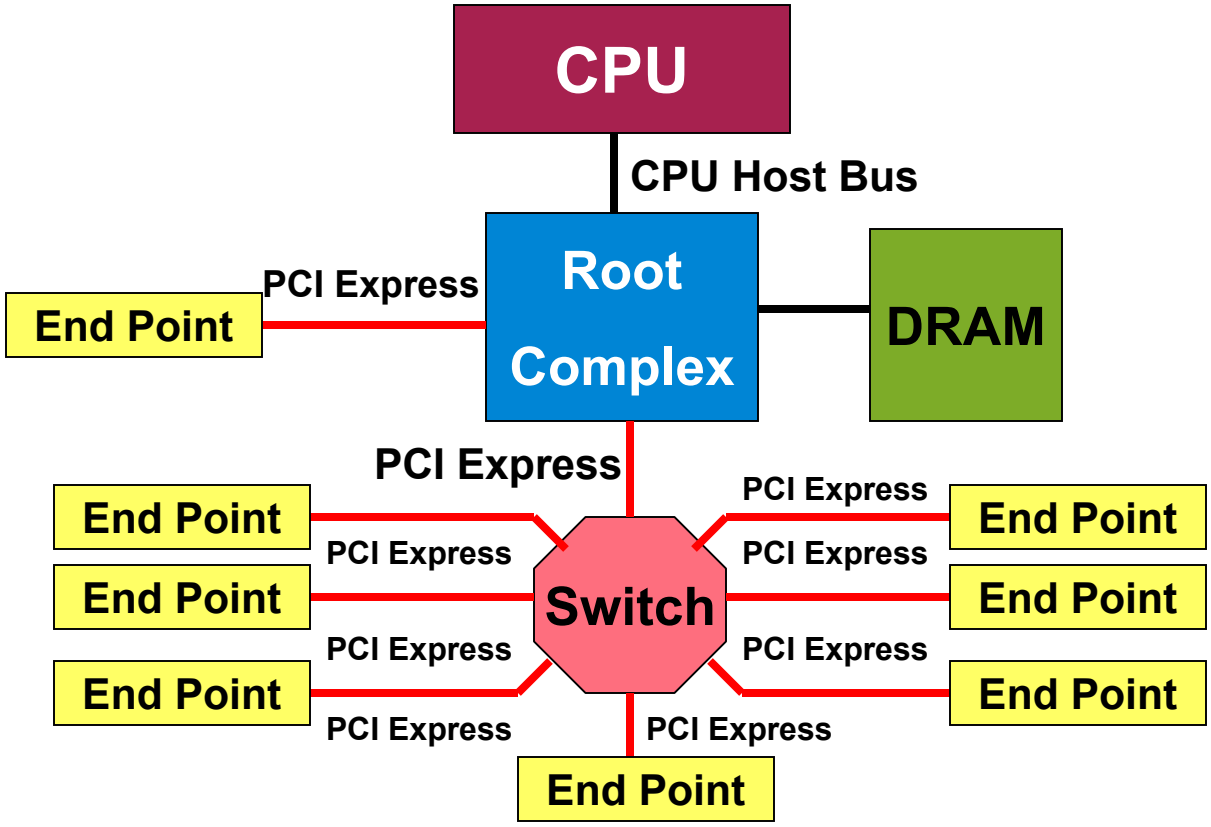
- Advanced Features
  - Virtual Channels and Traffic Class for different data types
    - Time-dependent Video and Audio Streaming
  - Isochrony
    - Predictable Latencies
  - Native Hot Plug/Surprise Unplug
  - Advanced Power Management
  - Data Integrity and Error Handling (RAS)
    - Rich set of error logging and reporting
  - Base mechanism enables communication and embedded applications
    - Advanced Switching Communications Extensions
  - Enhanced configuration
    - 4096 Bytes per function
      - First 256 Bytes alias to traditional PCI Index/data register space

PCI Express Overview

- High Speed, Low Pin Count, Point-to-point
  - Connections could be traces, connectors and cables
- Split transaction protocol with attributed packets for priority delivery

PCI Express System Diagram

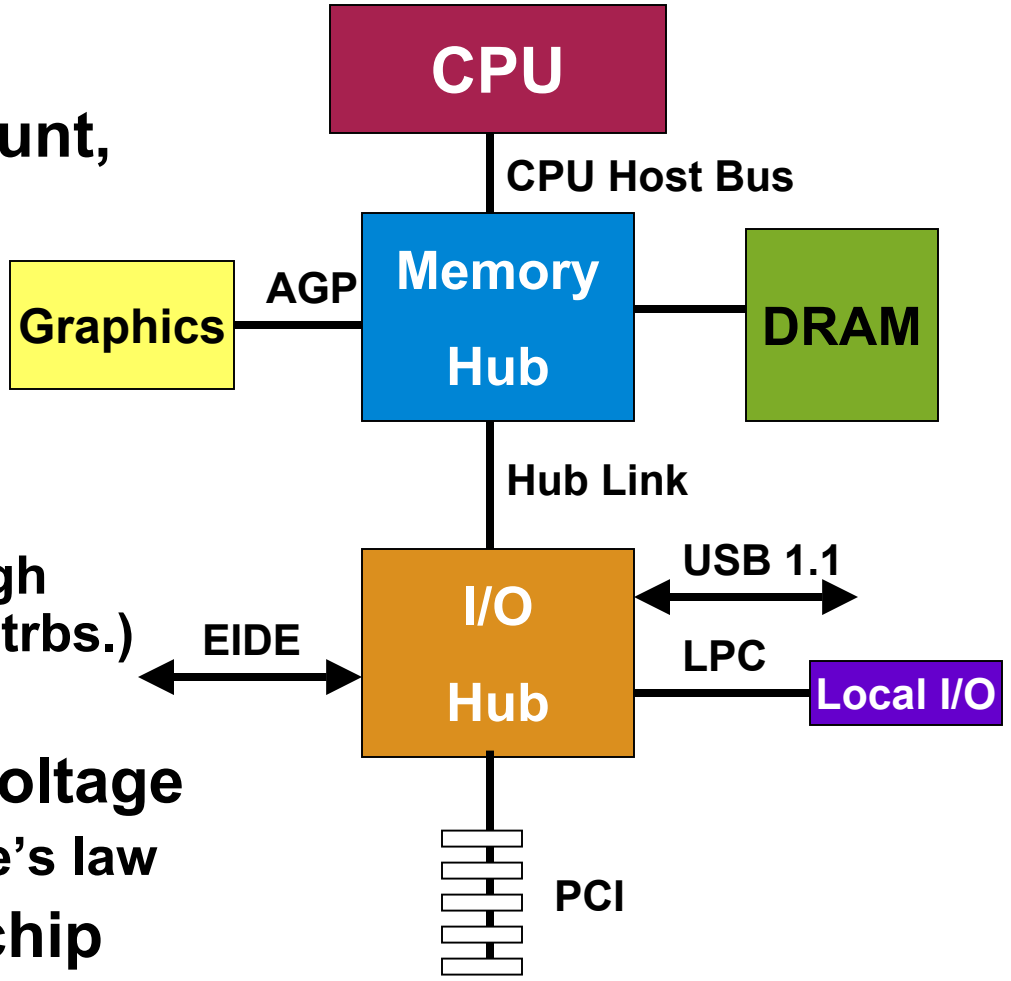
- Root Complex and switch with multiple endpoints
  - Switch replaces multi-drop bus
  - Provides fan-out
  - May provide peer-to-peer communication
  - Could be integrated into the chipset



Agilent Technologies

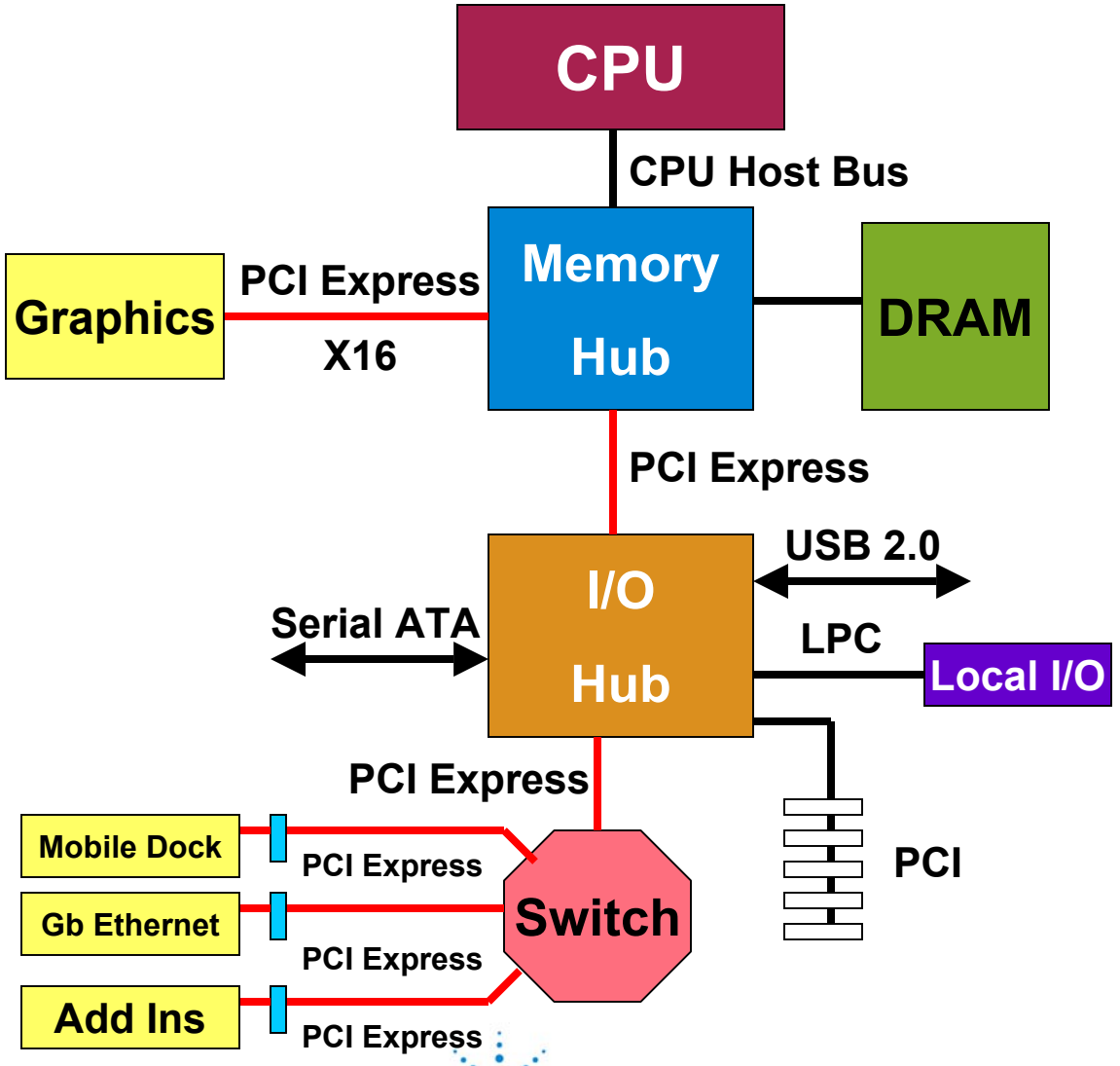
Current System Architecture

- Desktop/Mobile
  - Multiple, high pin count, parallel busses
    - Expensive
  - Difficult FR4 layout
    - Clock skew issues
    - Limited distance
    - Signal integrity at high frequencies (1 GHz strbs.)
  - Hard to scale up in frequency/down in voltage
    - Doesn't follow Moore's law
  - Proprietary chip-to-chip interfaces



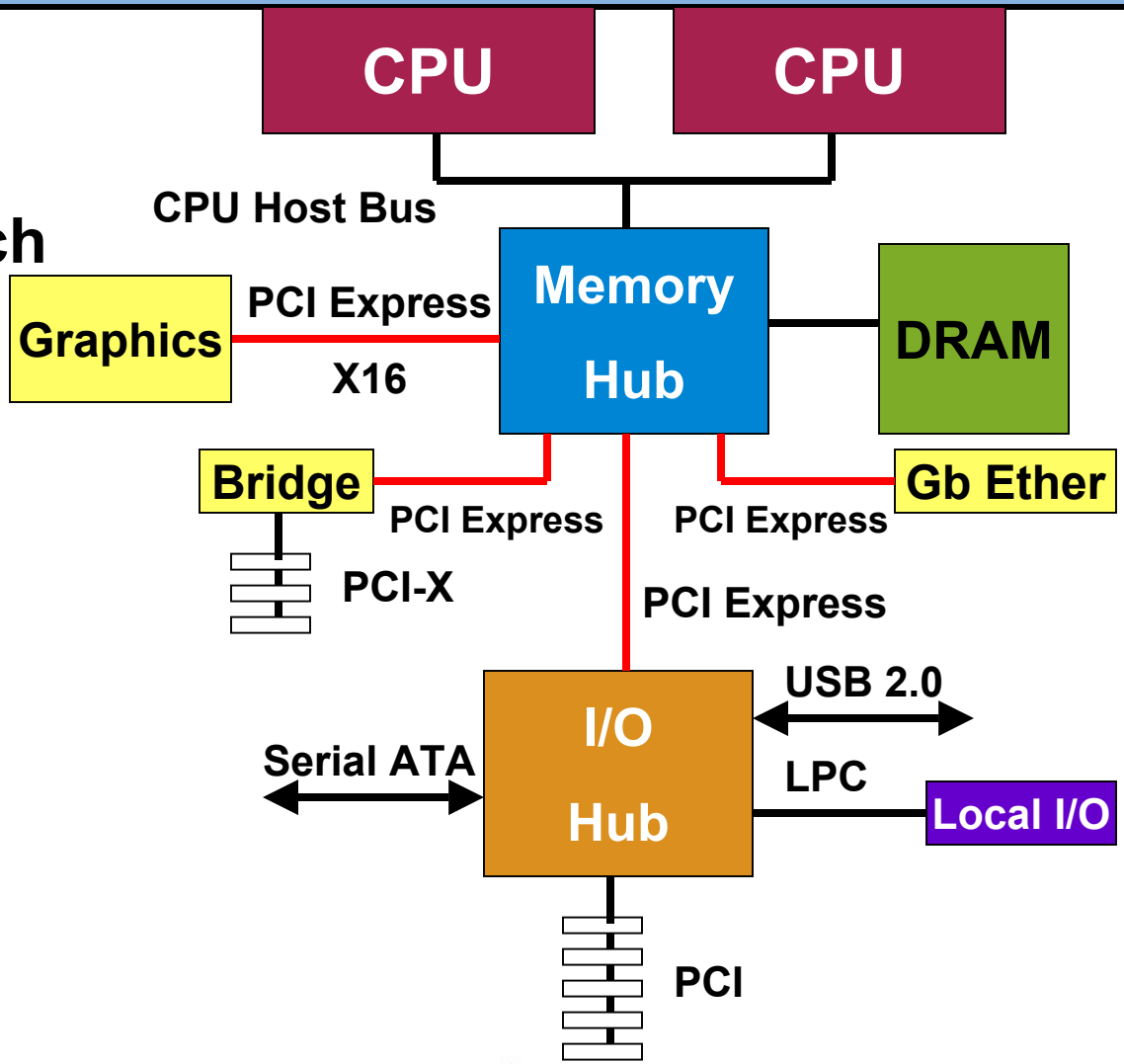
Platform Examples

- Desktop/Mobile
  - PCI Express graphics
  - PCI Express for chip-to-chip
  - USB 2.0
  - Serial ATA
  - PCI Express switch
  - PCI Express connectors



Platform Examples

- **Workstation**
  - Two CPU's
  - Integrated switch in memory hub
    - High BW I/O near memory
  - High BW graphics
  - Support for PCI-X via a PCI Express bridge

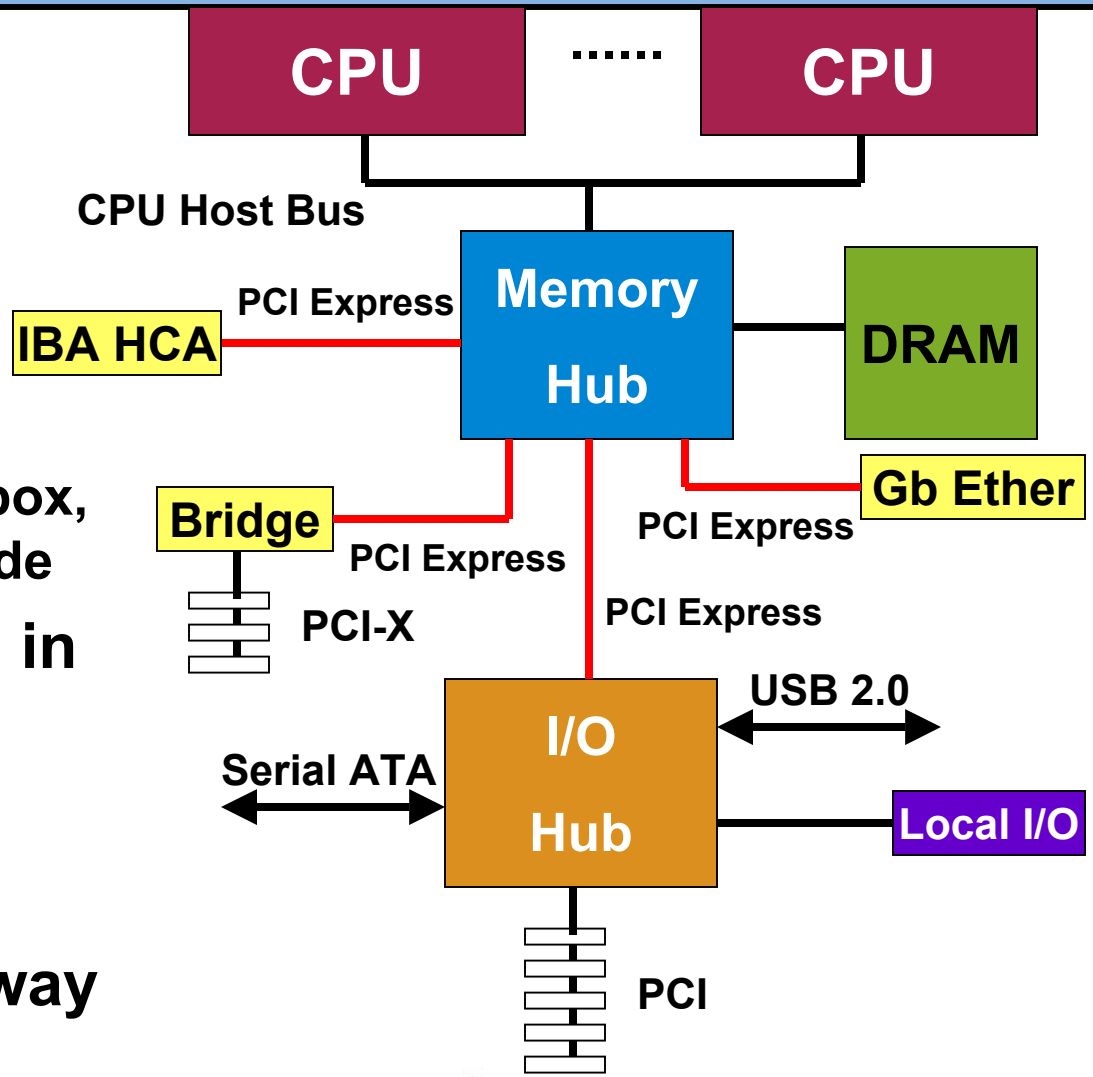


MINDSHARE

Agilent Technologies

Platform Examples

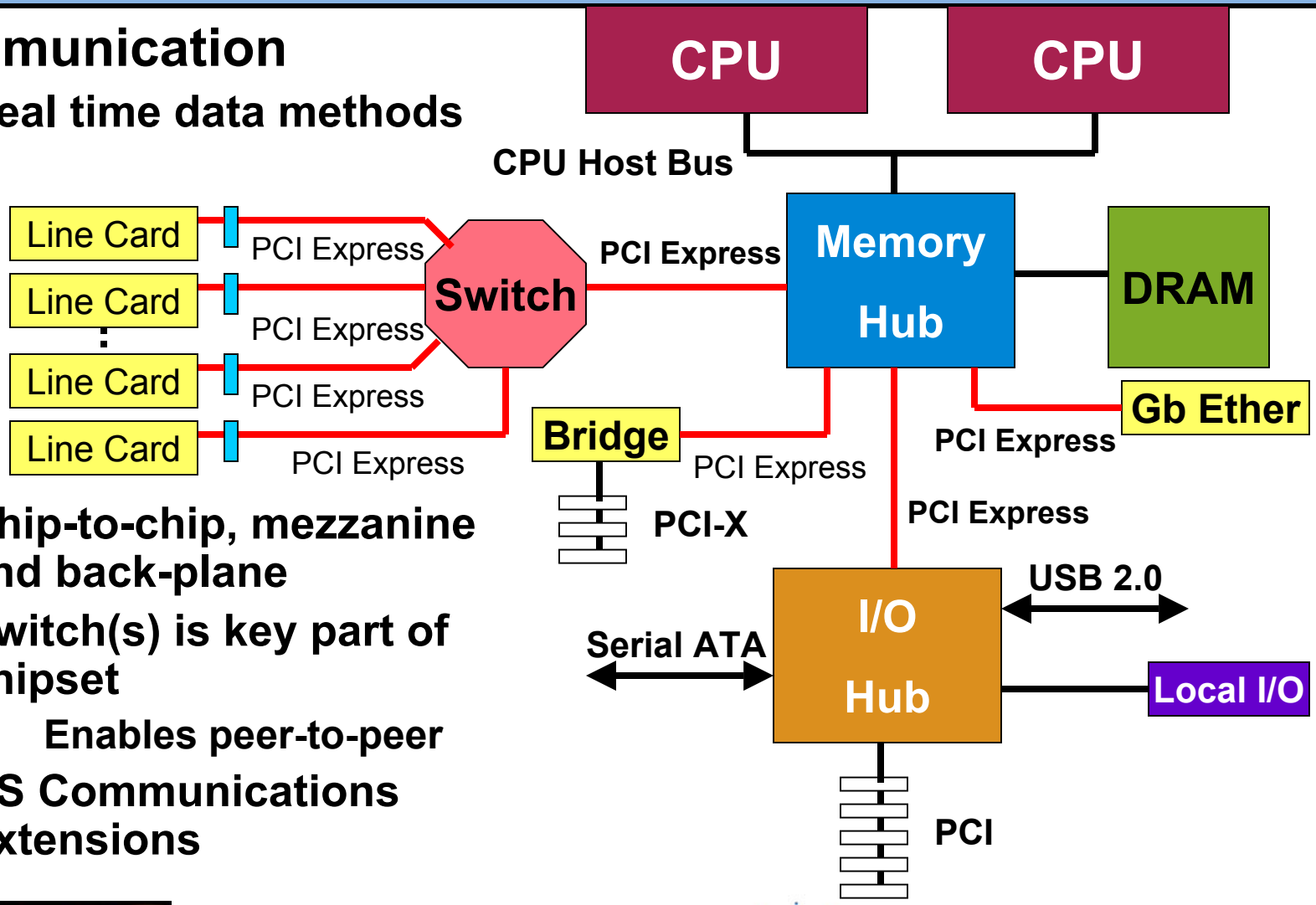
- Server
  - Multiple CPU's
  - Support for IBA HCA
    - Switched Fabric
    - IBA outside the box, PCI Express inside
  - Integrated switch in memory hub
  - Multiple PCI-X bridges
  - PCI doesn't go away



Agilent Technologies

Platform Examples

- Communication
  - Real time data methods
  - Chip-to-chip, mezzanine and back-plane
  - Switch(s) is key part of chipset
    - Enables peer-to-peer
  - AS Communications Extensions



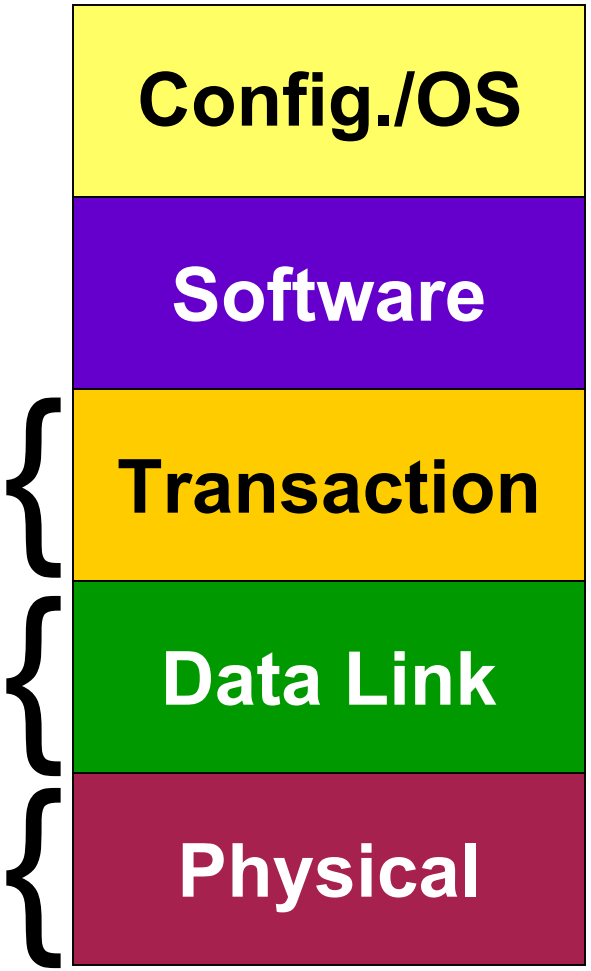
Q & A

- First question and answer session
  - One at the end, as well

Q & A

PCI Express Architecture

- Layered architecture enables:
  - Modularity
  - Scalability
  - Re-use
- Core architecture focuses on Transaction, DLL, and Phy
  - Plus mechanical under Phy



Configuration/Operating System

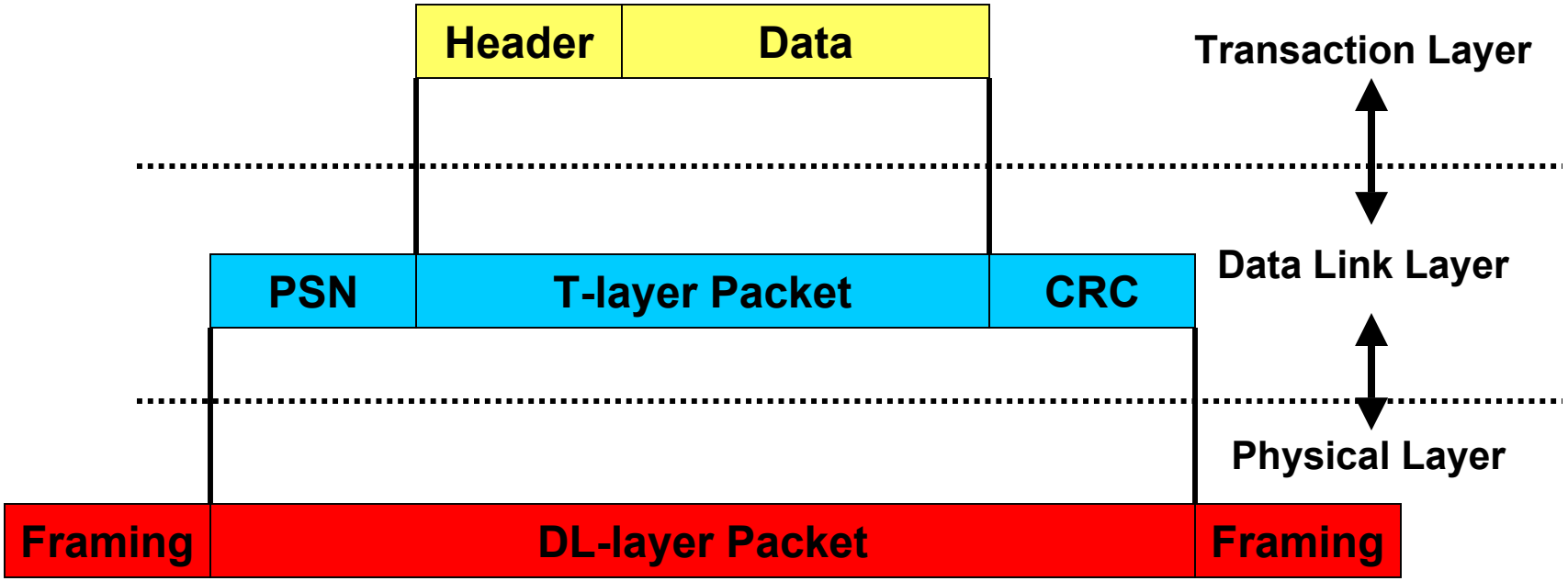
- **Compatibility with existing SW important for PCI Express**
  - **PCI already has a robust initialization model (plug and play) where SW can discover hardware and allocate resources**
  - **PCI Configuration space concept unchanged for PCI Express**
- **All existing OS's will boot on PCI Express platforms without modification**

Software Layer (Device Drivers)

- Run-time SW model supported by PCI is supported for PCI Express
  - Load/store architecture within a flat address space
  - All existing SW executes unchanged
    - Smooth integration of PCI Express into future platforms
- New SW must be written to use new capabilities

Transaction Layer

- Receives read and write requests from the SW layer
- Creates request packets for the Data Link Layer



Transaction Layer

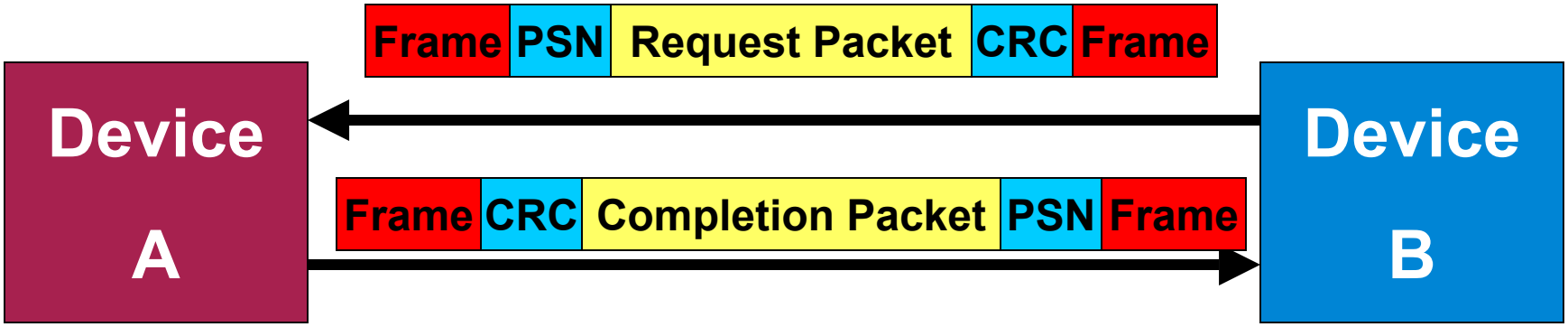
- All requests are implemented as split transactions
  - Some request packets require a response packet
- Transaction layer receives responses from the data link layer and matches these to the original SW request
  - Each response has a unique identifier that enables it to be directed to the correct originator

Transaction Layer

- Packet formats support 32 bit and extended 64 bit addressing
- PCI-X like producer/consumer ordering
- Transaction Layer end-to-end CRC's
  - For packets corrupted inside a switch
- Packets also have attributes that may be used for routing, including:
  - No Snoop
  - Relaxed Ordering
  - Priority

Basic PCI Express Transaction

- Split transaction model
  - Posted and non-posted requests
  - Non-posted requests require completions coming back
- Memory, I/O, Configuration, or Message Transactions

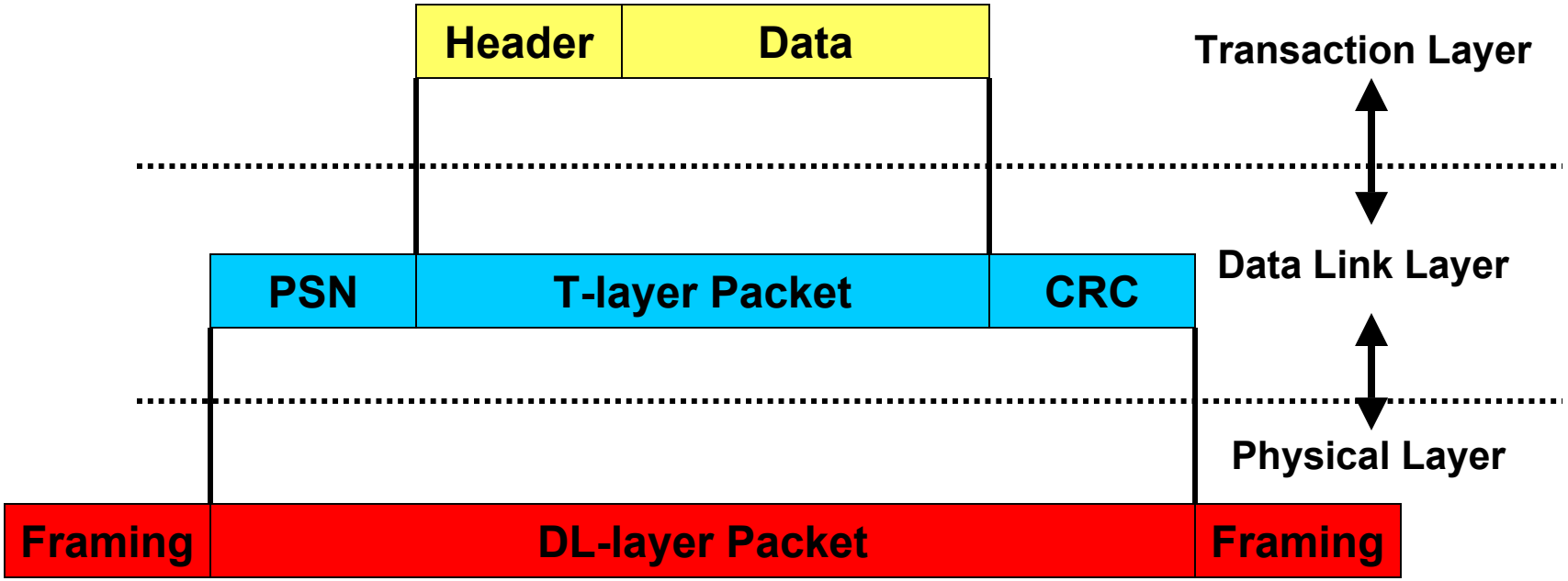


Transaction Layer

- MSI is primary method of delivery for PCI Express interrupts
- Message Space used to support all prior sideband signals
  - PME#
  - Reset#
- Messaging also used for special bus cycles
  - Interrupt Acknowledge
- Use of messaging as virtual wires eliminates costly pins

Data Link Layer

- Responsible for ensuring reliable delivery of the packet across the PCI Express link
  - Data integrity covered by a sequence number and a CRC

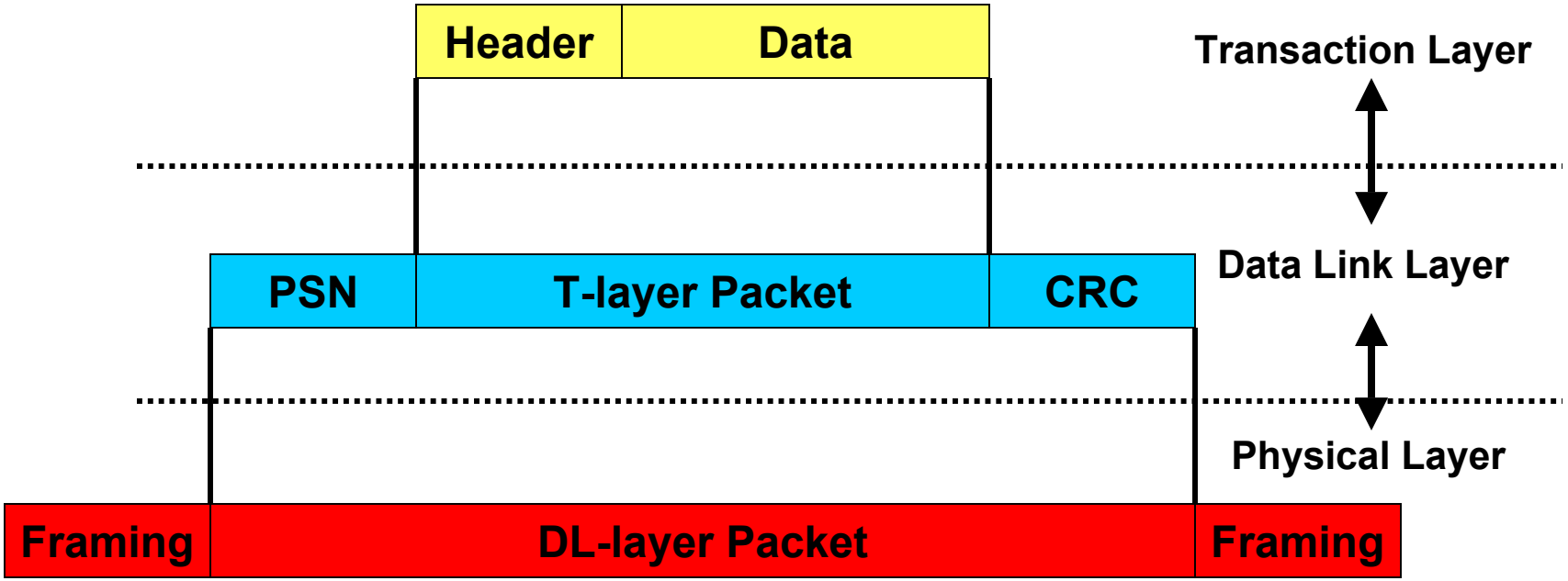


Data Link Layer

- Most packets initiated by the transaction layer
- DLL implements credit-based flow control
  - Ensures packet are only transmitted when a buffer is known to be available on the other end
  - Eliminates costly retries
- Data Link Layer automatically retries a packet signaled as corrupt
- 32-bit CRC generation and checking for TLP's
  - 16 bit CRC generation and checking for DLP's
- Lost packets detected via Packet Sequence Numbers

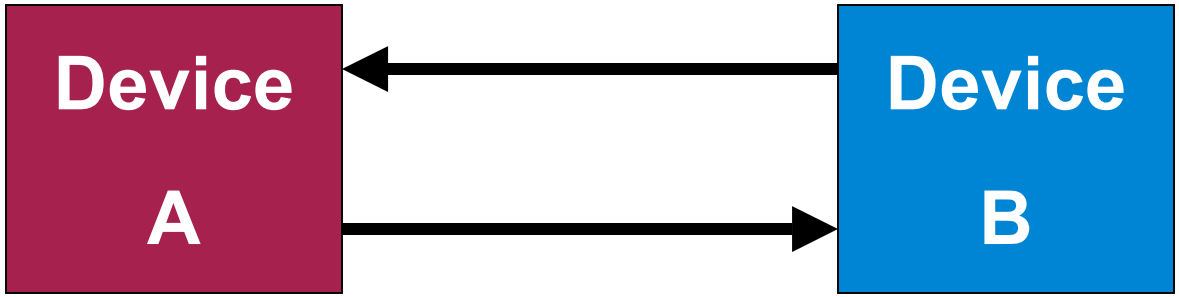
Physical Layer

- Adds framing characters to DLL packets prior to transmission



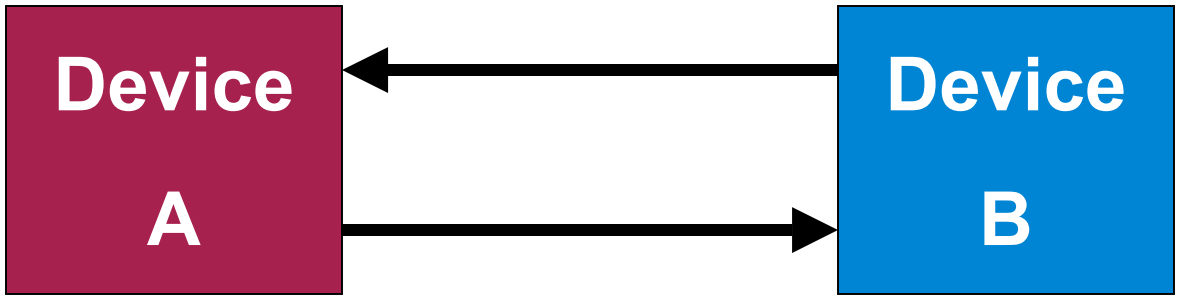
Physical Layer

- Each link contains two, low-voltage differential pairs
  - Transmit and receive; dual simplex
- Data clock is embedded using 8b/10b encoding
  - Same as IBA
- Initial frequency is 2.5 Gb/s
  - Expected to increase to 5 and 10 Gb/s

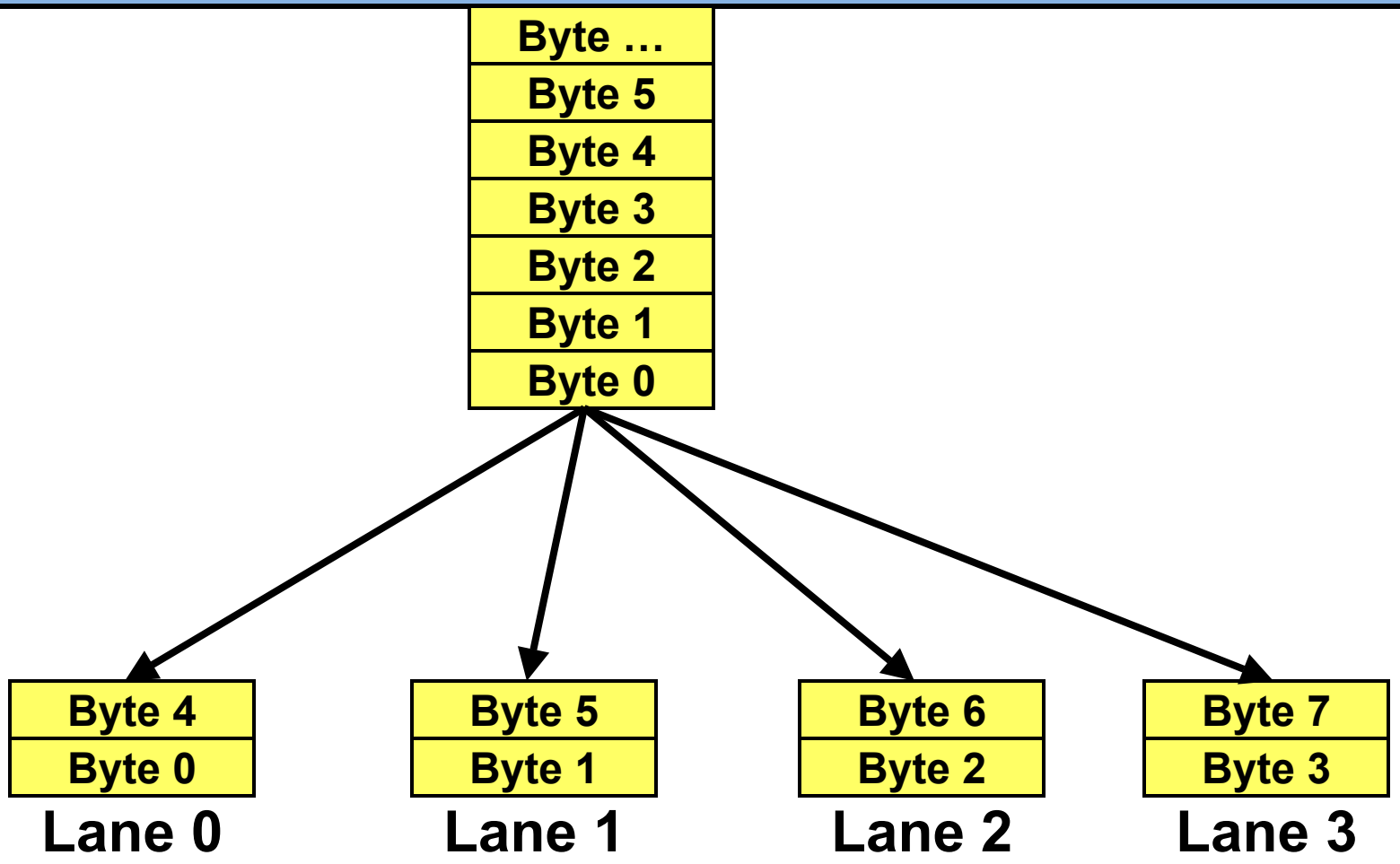


Physical Layer

- Bandwidth is scaled by adding signal pins to form multiple lanes
  - X1, X2, X4, X8, X12, X16, and X32 lane widths supported
- Byte striping shown on the next slide for a X4 link

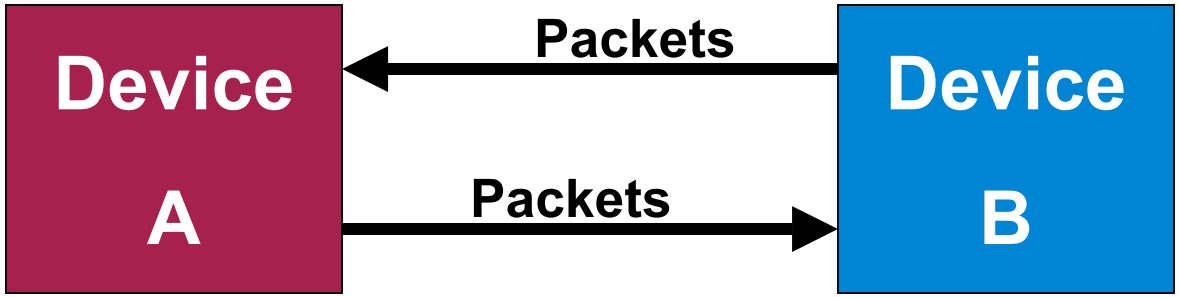


Byte Striping on a X4 Link



Physical Layer

- Each PCI Express link has auto-negotiated link width and frequency during initialization
  - Two devices on both ends come to a lowest common denominator with no firmware or OS involvement
- PCI Express expects future speed upgrades
  - Perhaps new encoding techniques and media
  - These changes would only effect the physical layer



Mechanical Concepts

- Evolutionary versus revolutionary
  - Evolutionary to allow co-existence with standard PCI
  - Revolutionary to encourage new system partitioning

Evolutionary Design

- For initial implementations
  - PCI Express add-in cards and connectors
  - Coexisting with PCI cousins
  - PCI form factors
    - Full size and half-height cards
  - No changes required to existing desk top/server form factors
  - Cost effective migration from PCI

Revolutionary Design

- Enabling separate compute brick and end user I/O device
  - Hot swappable module
  - Cable connection
- Small form factor Mobile Communications connector
  - Covers build/configure to order
  - Similar in concept to Mini PCI
    - Drastically reduced pin count and smaller
- Embedded Communication Backplane

Performance

Link Width	X1	X4	X8	X16	X32
Aggregate BW (GB/s)	0.5	2	4	8	16

- Theoretical Throughput
- A sampling of link widths
- Assumes 2.5 GB/s signaling
- Includes 8b/10b encoding overhead
- Aggregate means both directions simultaneously
- Higher than what is currently available today

Spec. Update

- 1.0 Final Specification released 7/23/02
- 2002/2003 Industry enabling efforts
- First Si 2003

Summary

- PCI Express is the PCI for the next decade
  - Architecture allows for smooth migration in the future
  - Stable and scalable
- Broad industry support and acceptance
  - Multiple market segments
- High volume shipments in 2003
  - No changes to system and SW infrastructure

Validation of PCI-Express Systems

Perry Keller

Computer I/O Bus Analysis Mgr.

Agilent Technologies

Design Validation PGU R&D

Validation of PCI Express Systems

16700 Logic Analyzer

8720 VNA + ATN 4002

86100 DCA

8113/4A Pulse Generator

81250 ParBERT

Config./OS

Software

Transaction

Data Link

Physical

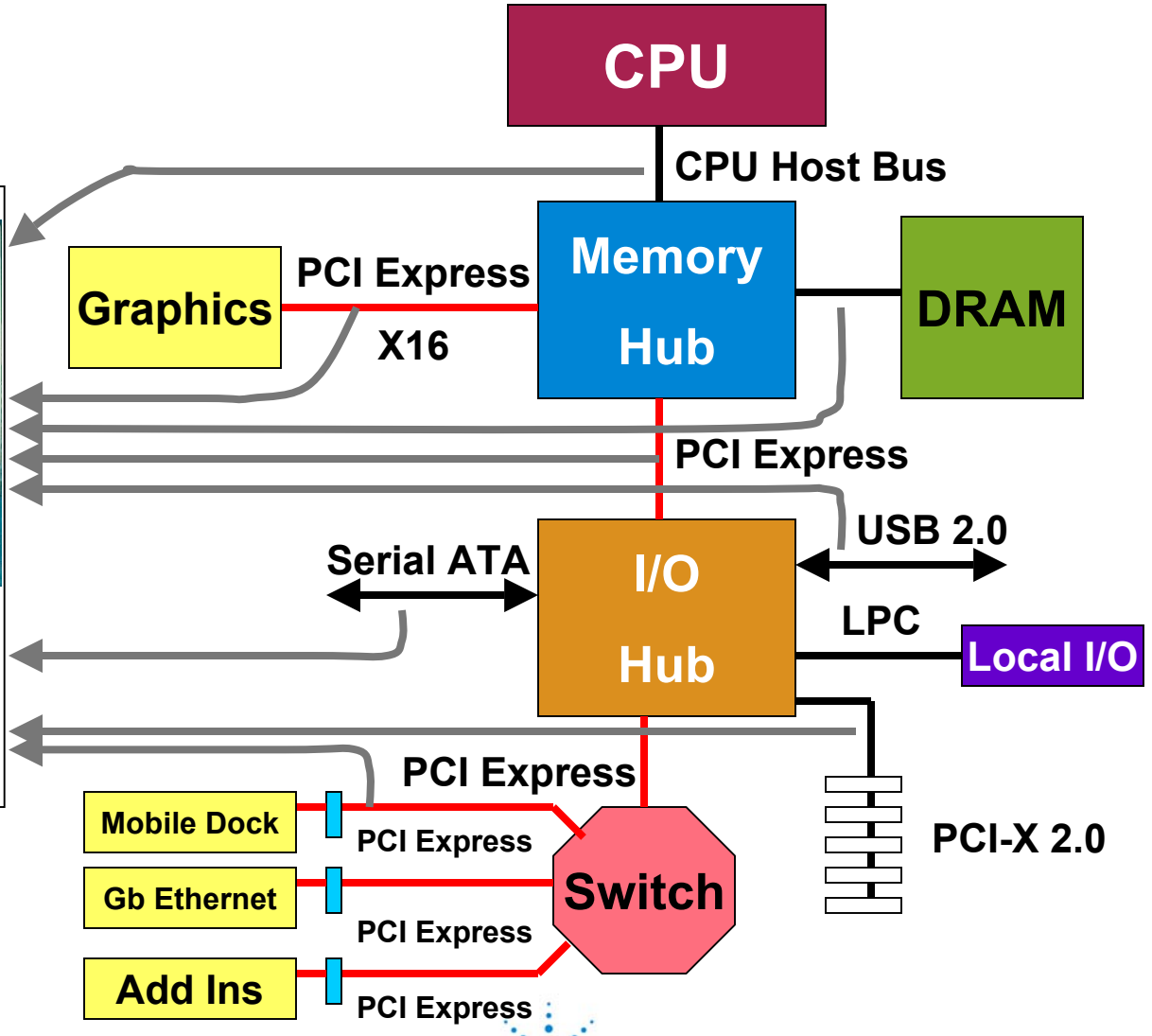
Agilent Technologies

Requirements for System Validation

- Non-intrusively probe PCI Express slots AND chip-to-chip links
- Support all widths (1x thru 32x) at 1st Silicon
- Packet trace and trigger
- Support for all operating modes
  - Squelch, 10B/8B, Link training, TLP, DLLP
  - Spread spectrum and data scrambling
- Cross-bus correlated analysis and sequenced event triggering across all busses

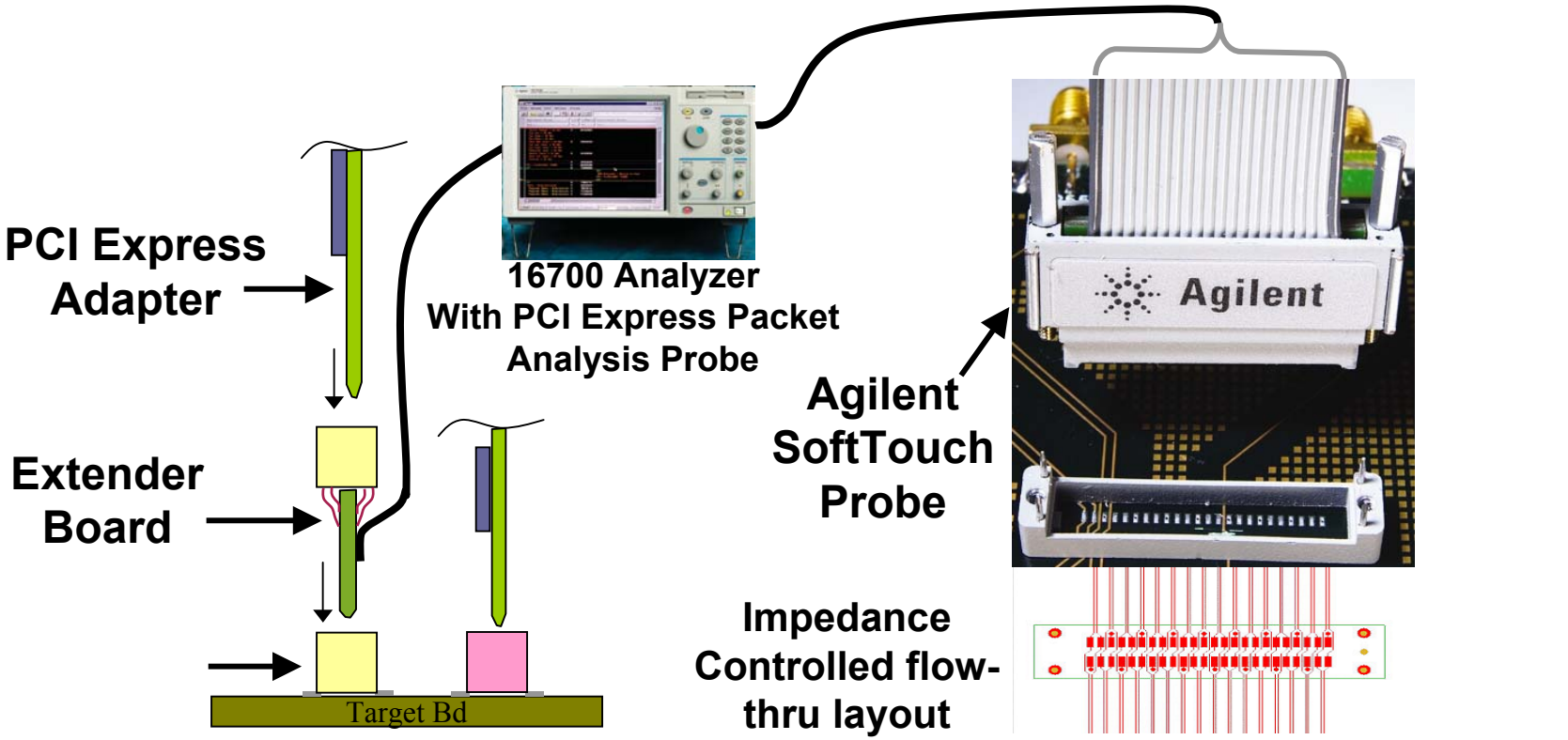
PCI Express Validation with 16700

16700 Logic Analyzer
With PCI Express Packet
Analysis and Other Bus
Probes



Agilent Technologies

Step 1: Probing the Link

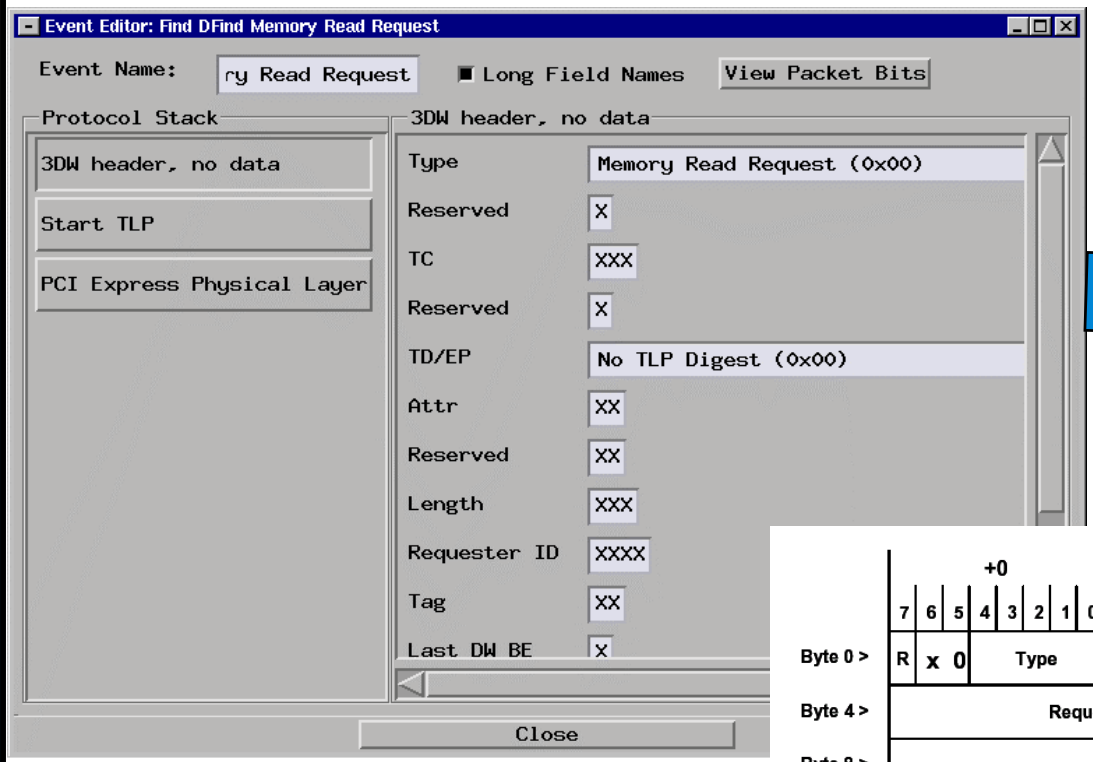


PCI-Express Slot Probing

Chip-Chip Connectorless Probing

Passive Probing Preserves all Bus Behavior

Step 2: Packet Triggering



Describe your trigger At Packet Level

Analyzer handles bit order, 10B/8B, scrambling, etc.

	+0								+1								+2								+3							
	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0
Byte 0 >	R	x	0	Type					R	TC							Reserved								Attr		R		Length				
Byte 4 >	Requester ID																Tag								Last DW BE				1st DW BE			
Byte 8 >	Address[31:2]																														T D	E P

Focus on Packets Instead of Format Details

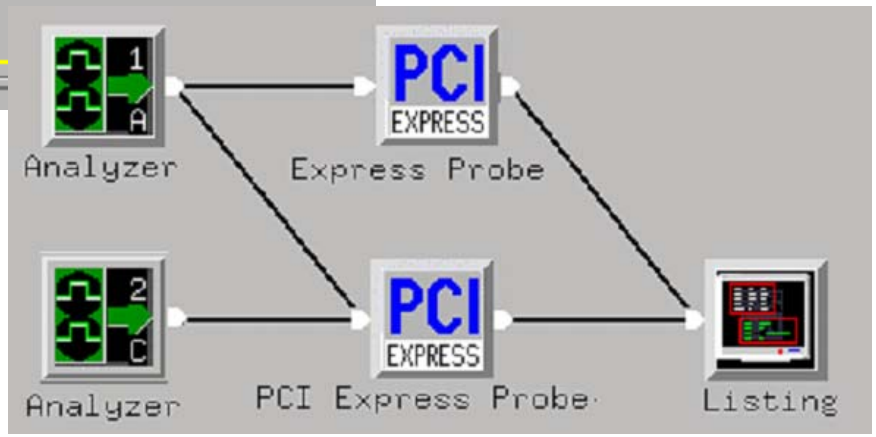
Agilent Technologies

Step 3: Cross Bus and Packet Sequence

Request

Followed by

Response



Combine Events From Any Bus in Any Order

Agilent Technologies

Step 4: Packet Trace and Decode

Packet Decode

Text

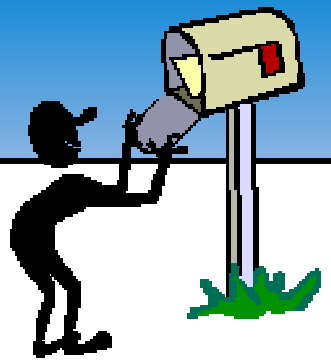
```
Start TLP
Reserved = 0 Hex
Packet Sequence Number = 000 Hex
Reserved = 0 Binary
Global Format = 00 Binary (3DW header, no
  Type = 04 Hex (Configuration Read Type 0
  Reserved = 0 Binary
  TC = 000 Binary
  Reserved = 0 Hex
  TD/EP = 0 Hex (No TLP Digest)
```

Summary

- PCI Express enables performance and cost improvements for the next decade
  - Architecture allows for smooth migration in the future
  - Stable and scalable
- Broad industry support and acceptance
  - Multiple market segments
- High volume shipments in 2003
  - No changes to system and SW infrastructure
- Design validation technology has advanced to support PCI Express 1st silicon turnon

Q & A

FREE Agilent Email Updates



Subscribe Today!

Choose the information YOU want.
Change your preferences or unsubscribe anytime.

Keep up to date on:

Services and Support Information
- Firmware updates
- Manuals
- Education and training courses
- Calibration
- Additional services

Events and Announcement
- New product announcement
- Technology information
- Application and product notes
- Seminars and Tradeshows
- eSeminars

Go To:

www.agilent.com/find/emailupdates