Introduction to PCI Express-

A New Serial Multi Gigabit Commodity Data Bus

Satish K Dhawan

Yale University

Abstract

New PCs have a new bus called PCI Express is a dual-simplex, point to point serial differential low voltage interconnects that will consolidate application requirements for use by multiple segments in the industrial world. The signaling rate is 2.5 Gbit per second, with 8/10 bit encoding to embed clock in the data stream. On the transmit side parallel data is shifted out serially and on the receive side serial data is shifted into registers for parallel data output. The receiver also recovers the embedded clock.

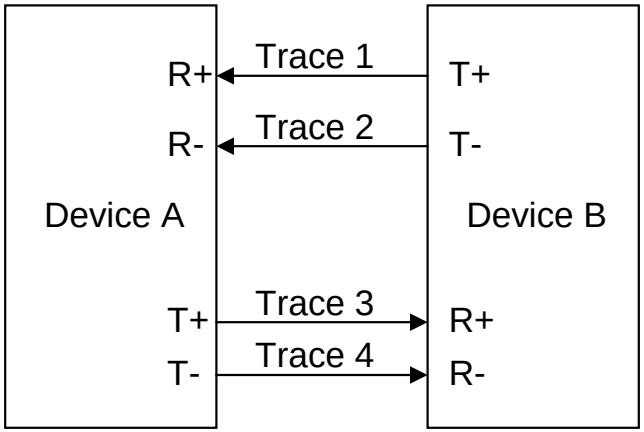
This bus can be used to connect module or boxes via twisted pair copper wires.

IO Bus comparison

Table 1: I/O Bus Bandwidth Comparison

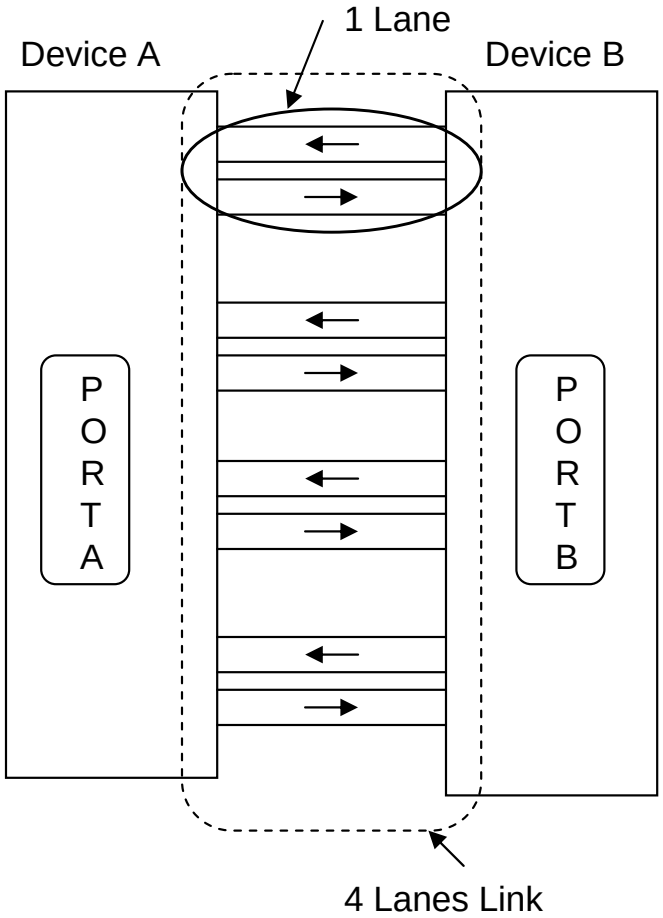
	I/O Bus	MHz	Bus Width	Rate	Transmission	Measured
			Bits	Mbytes/sec		@ Yale
PC Buses	Industry Standard Architecture (ISA)	8.3	8/16	8.3		
	Extended Industry Standard Architecture (EISA)	8.3	32	33		
	Peripheral Component Interconnect (PCI)	33	32	132		
	AGP4X			1,000		
	AGP8X			2,000		
Networks	Ethernet	10	1	1.25	CAT 5 Cable	
	USB 2.0	480 mb	1	55		40
Networks	Gigabit Ethernet	1250	1	125	CAT 5 Cable	65%
	PCI Express	2500	x1	250	Dual Simplex	
			x4	1,000	Dual Simplex	
	PCI Express Gen2	5000	x1	500	Dual Simplex	
			x4	2,000	Dual Simplex	
Networks	Infiniband	2500	x4	1,000	Dual Simplex	75%
Instrumentation						
	CAMAC	1	24	3		3
	FASTCAMAC Level 1	2.5	24	7.5		7.5
	FASTCAMAC Level 2	10	24/48	30/60		40
	VME	10	16-64	20-80 ??		
	Compact PCI	33	32	132		

X1 is 1 Lane: one Transmit and one Receive



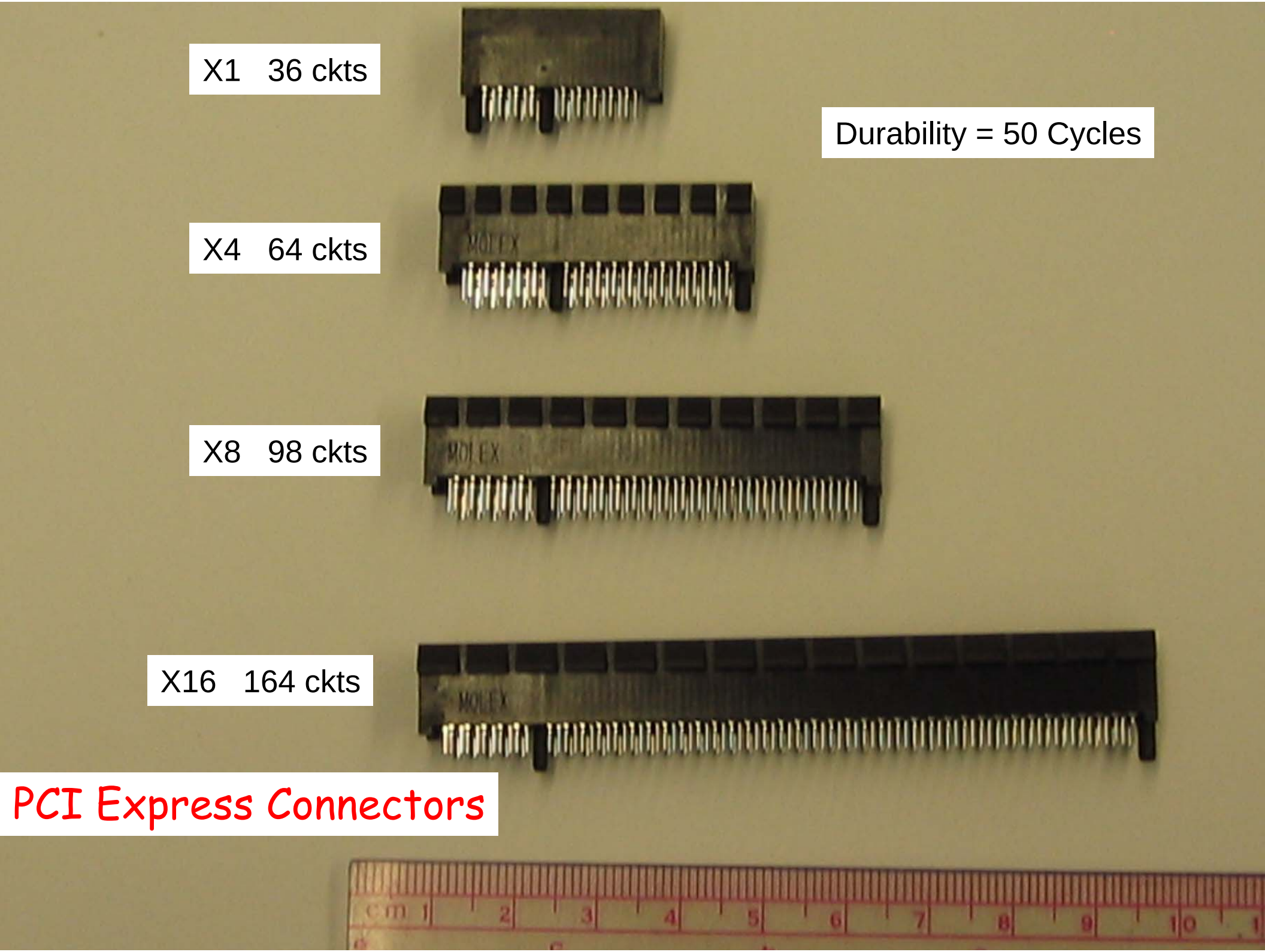
2 Twisted Pairs

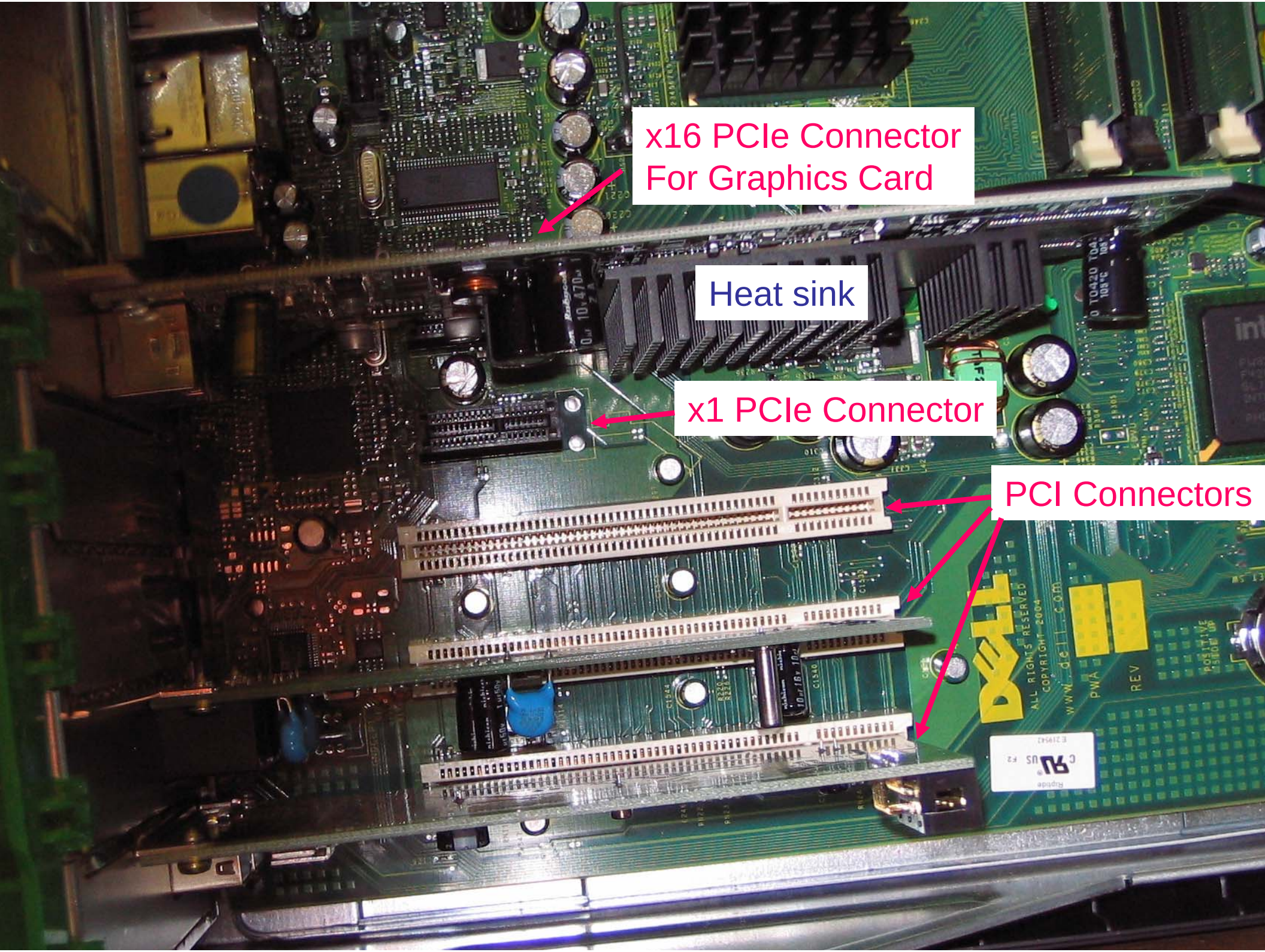
Point to Point Connections between Two PCIe Devices

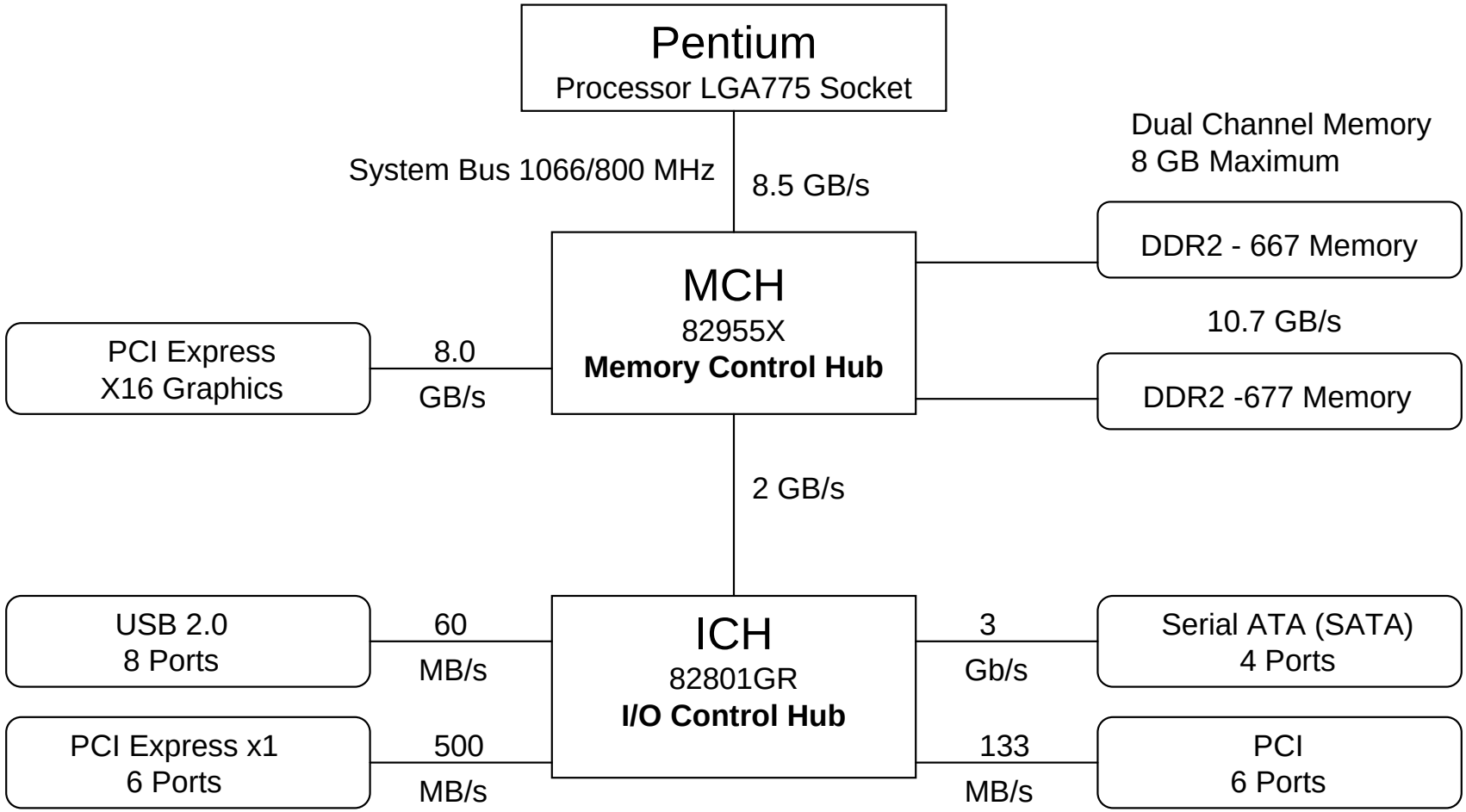


Electricals

- Low Voltage Differential Signaling
- AC coupling on Transmit Side
- 800 mV min for Initial Bit Transition
- 505 mV min for Subsequent Bit Transition
- Above is done to Reduce Inter-Symbol Interference
- 4 Layer FR4. Signal Layers on Outside
- 100 Ohms Traces

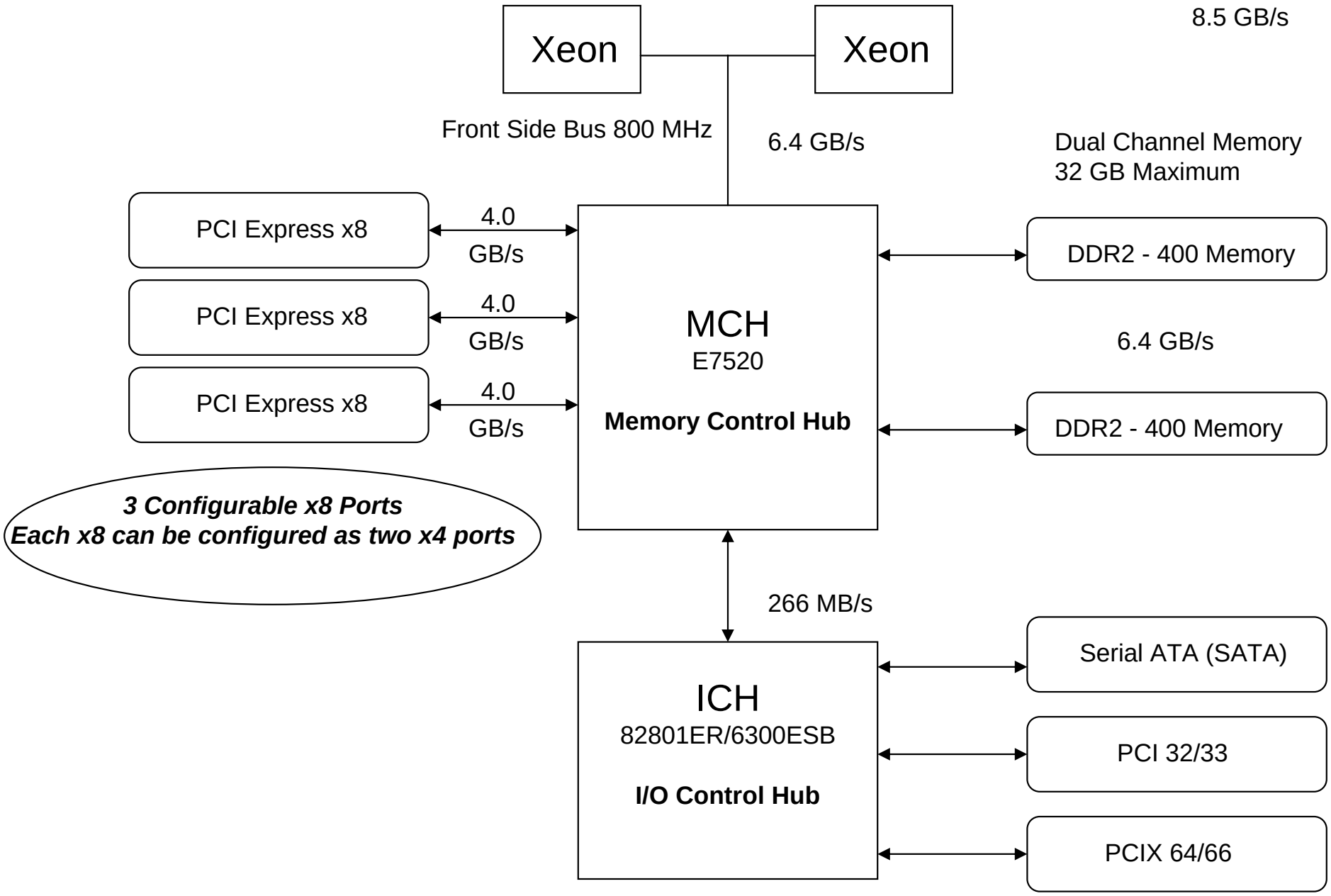

X1 36 ckts
Durability = 50 Cycles
X4 64 ckts
MOLEX
X8 98 ckts
MOLEX
X16 164 ckts
MOLEX
PCI Express Connectors


x16 PCIe Connector
For Graphics Card
Heat sink
x1 PCIe Connector
PCI Connectors

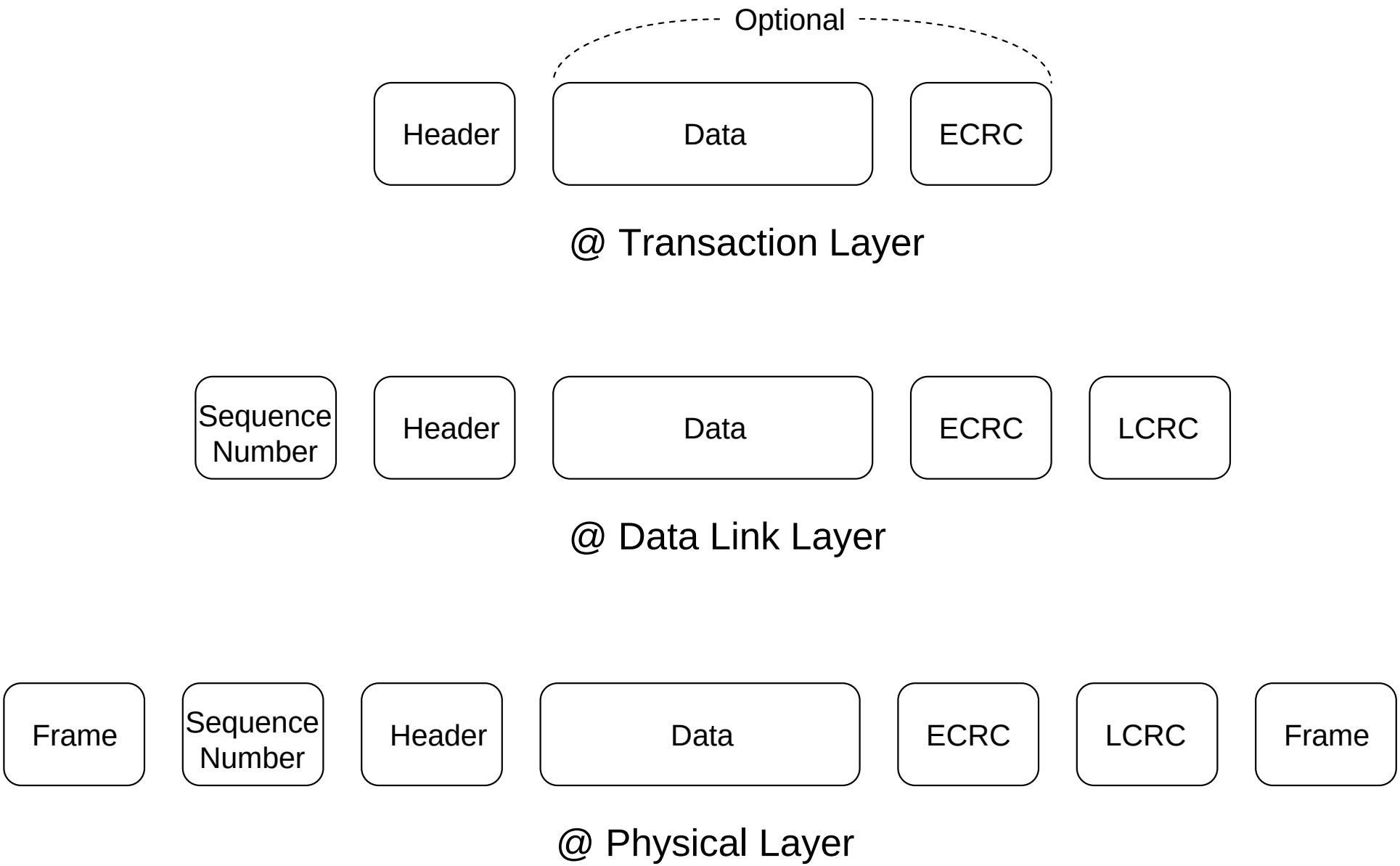


Intel 955X Express Chipset supports Intel Pentium processor Extreme Edition and all Pentium processors in LGA775 Socket

Desktop Chipset

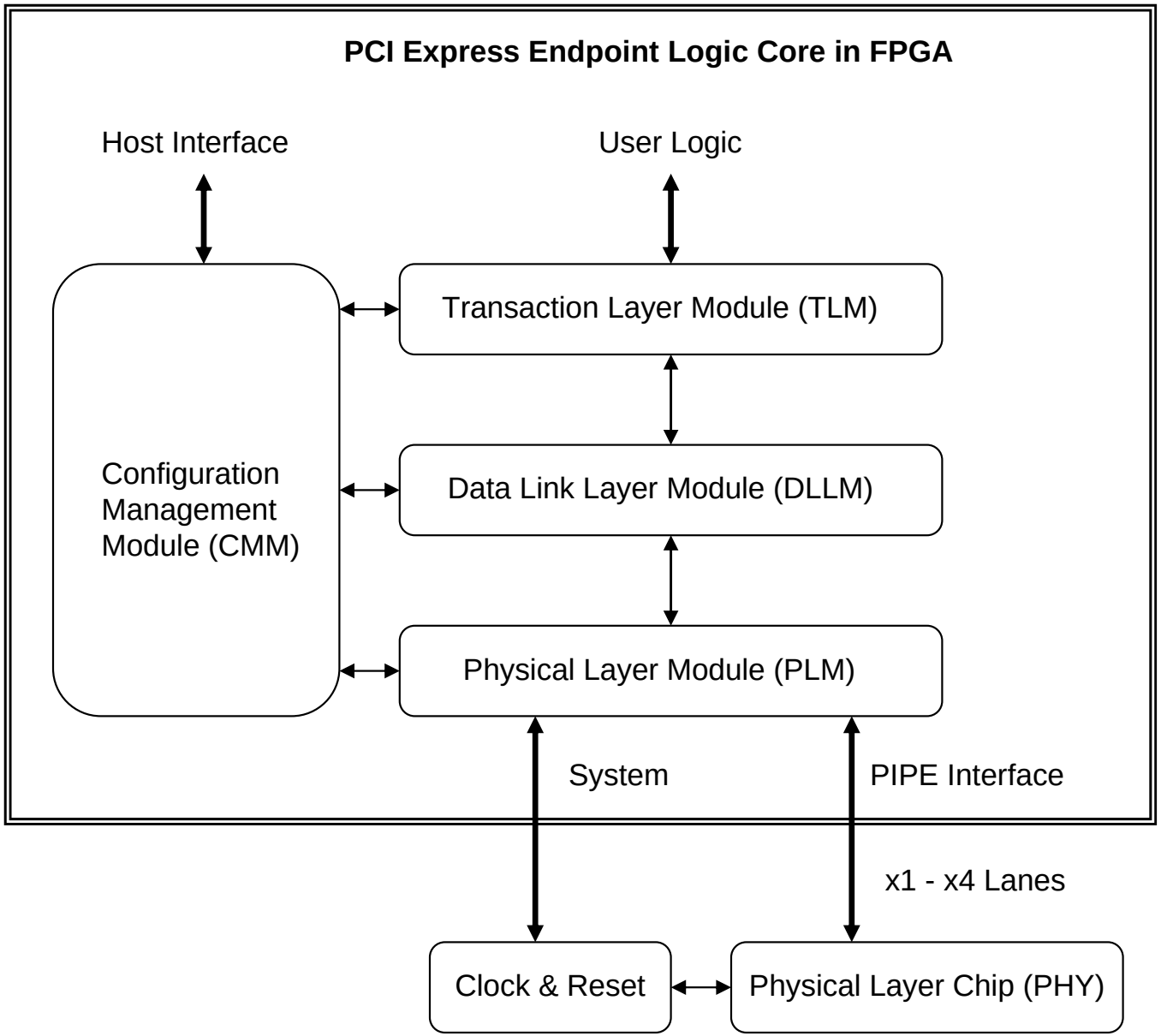


Dual Processor Chipset



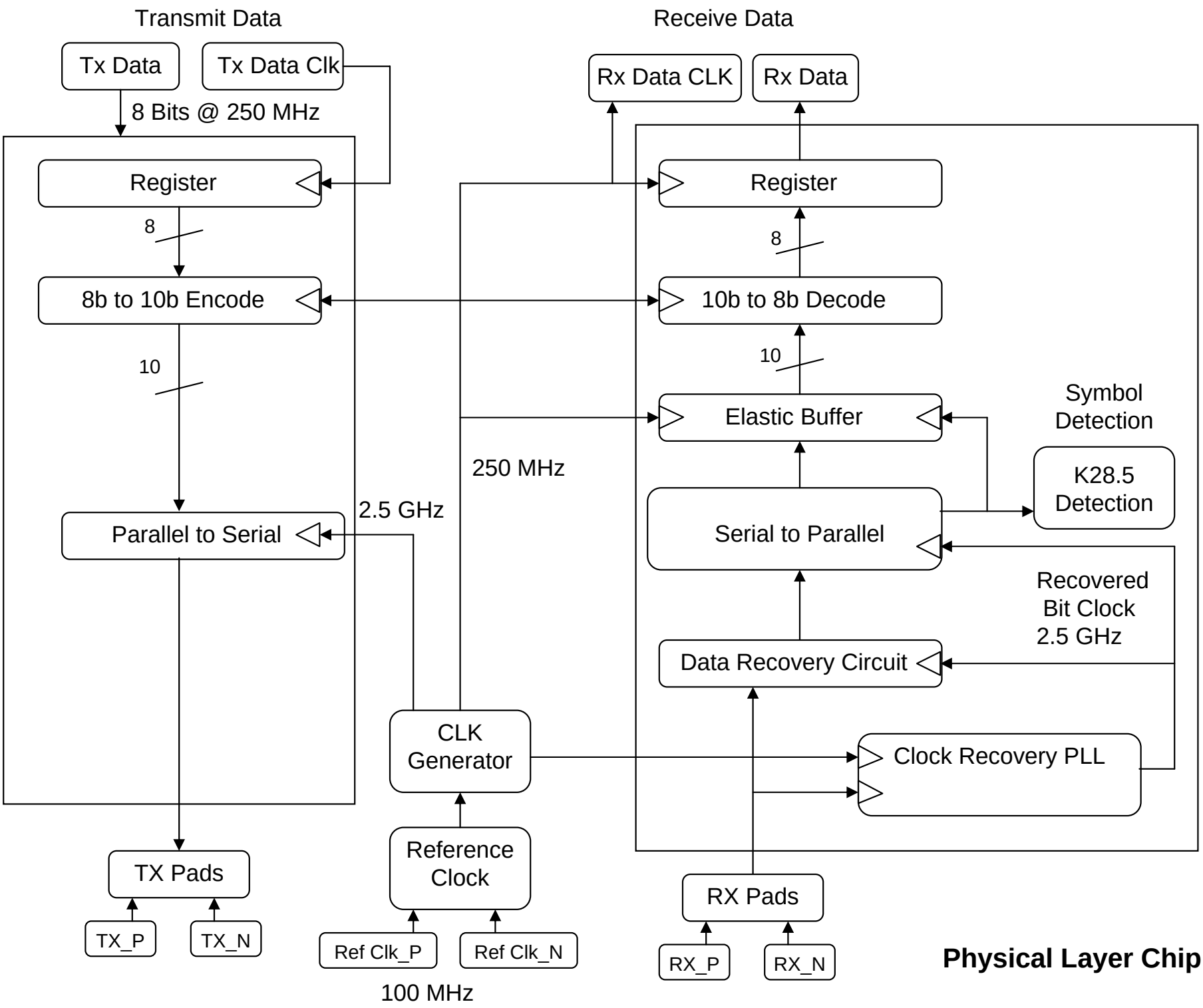
Data: Maximum 4K Bytes – **Typical implementation 256 bytes**, limited by high speed SRAM

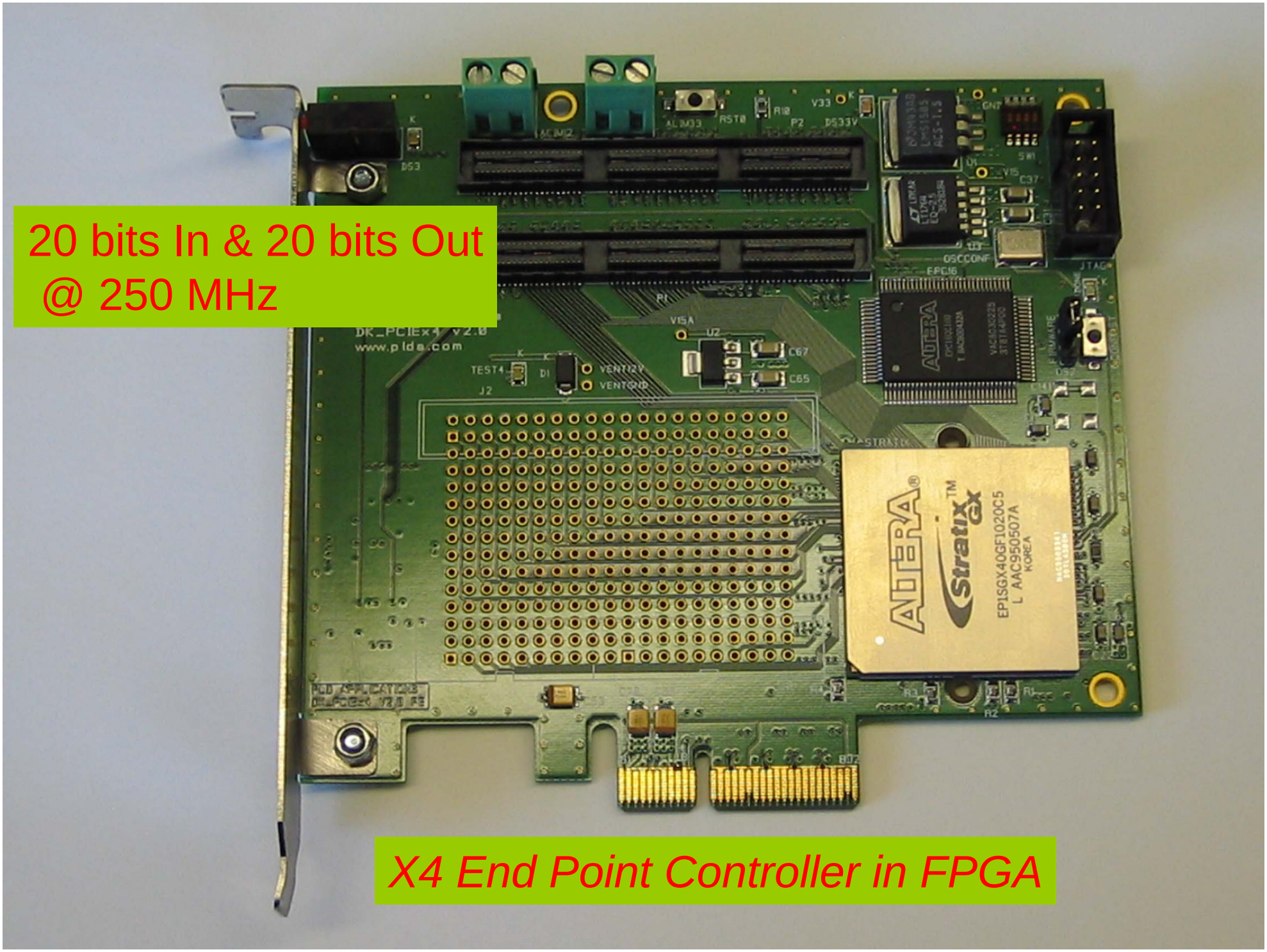
Transaction Buildup for a TLP through Architecture Layers



PCI Express Endpoint Controller

Intel Specified Interface - PIPE: (Physical Interface for PCI Express)


Transmit Data
Tx Data
Tx Data Clk
8 Bits @ 250 MHz
Register
8
8b to 10b Encode
10
Parallel to Serial
TX Pads
TX_P
TX_N
2.5 GHz
250 MHz
CLK Generator
Reference Clock
Ref Clk_P
Ref Clk_N
100 MHz
Receive Data
Rx Data CLK
Rx Data
Register
8
10b to 8b Decode
10
Elastic Buffer
Serial to Parallel
Symbol Detection
K28.5 Detection
Recovered Bit Clock 2.5 GHz
Data Recovery Circuit
Clock Recovery PLL
RX Pads
RX_P
RX_N
Physical Layer Chip

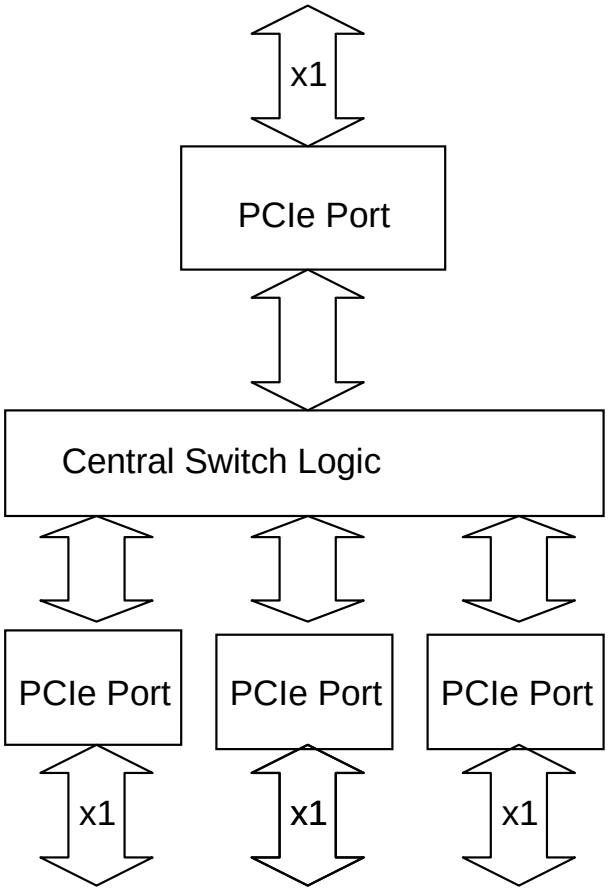

20 bits In & 20 bits Out
@ 250 MHz
X4 End Point Controller in FPGA
DK_PCIEx4 v2.0
www.plda.com
ALTERA
Stratix GX
EP1SGX40GF1020C5
L AAC950507A
KOREA

Need End Point Controller in silicon

Current Products

- Bridges:
  - PCIe <> PCI
  - PCIe <> Infiniband
  - PCIe <> Gig Ethernet
  - PCIe <> USB

- Need PCIe <> Generic Data IO Bus
  - Easy to use.
  - Just Supply Data and Configuration.
  - Chip does the rest.



4 Port x1 Switch

Switches

- 16 Lanes 4 Ports: Programmable x1,x2,x4,x8
- 32 Lanes 8 Ports: Programmable x1,x2,x4,x8,x16
- In Development; 8 Ports x8 Gen2 (5 GHz Signaling)

Bridges

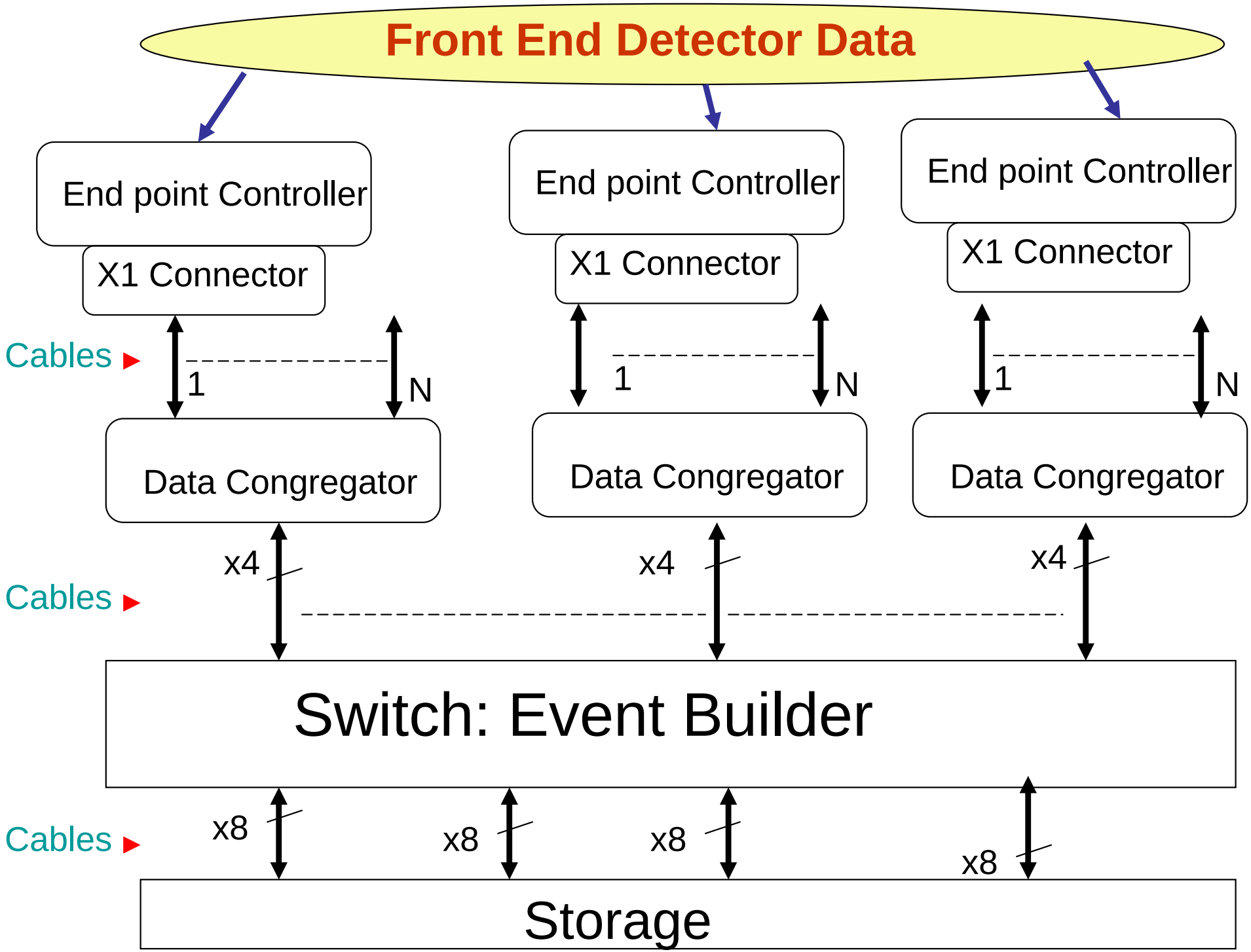
- PCI – PCIe
- PCX – PCIe
- USB – PCIe
- 1394A – PCIe 400 mb/s
- Infiniband – PCIe
- But no Generic DIo – PCIe Yet

Standard Connectors & Cables for PCIe

PCI Express Switching / Bridging

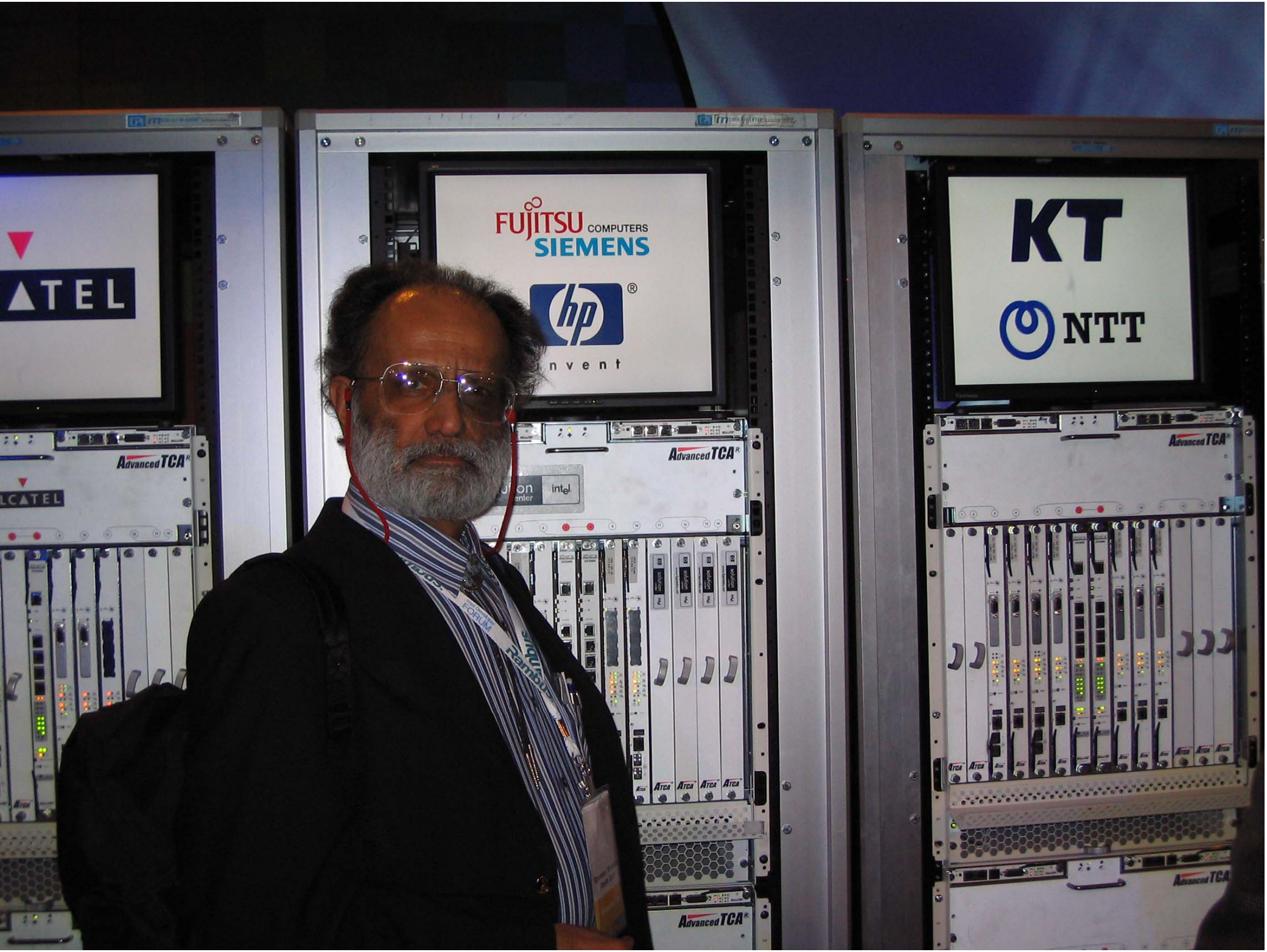
x4 Cable: Stackable Connectors

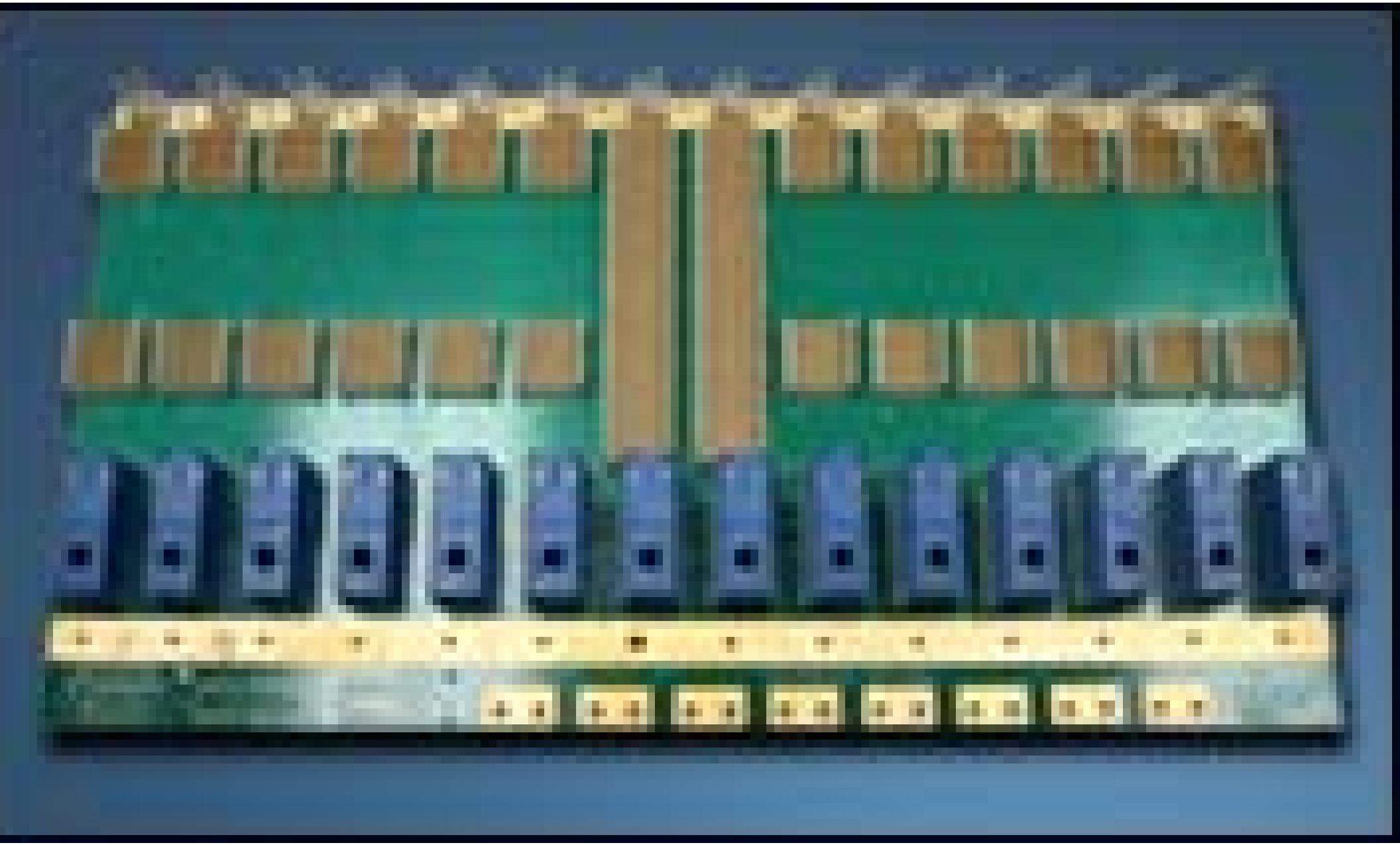
Length ~7 meters CAT6 Cable


Front End Detector Data
End point Controller
X1 Connector
End point Controller
X1 Connector
End point Controller
X1 Connector
Cables
1
N
1
N
1
N
Data Congregator
Data Congregator
Data Congregator
x4
x4
x4
Cables
Switch: Event Builder
Cables
x8
x8
x8
x8
Storage

ATCA

- Advance Telecom Computing Architecture
- Passive Back Plane
- 8 U Printed Circuit Board
- Dual Star or Full mesh
- - 48 Volt Power in
- x4 Traces
- Specifications for PCIe, Infiniband, GigE using
- the same Back plane
- AMC (Advanced mezzanine Card)
- mTCA based on ATC specifications - 4 U PCB
- Interconnections for Servers

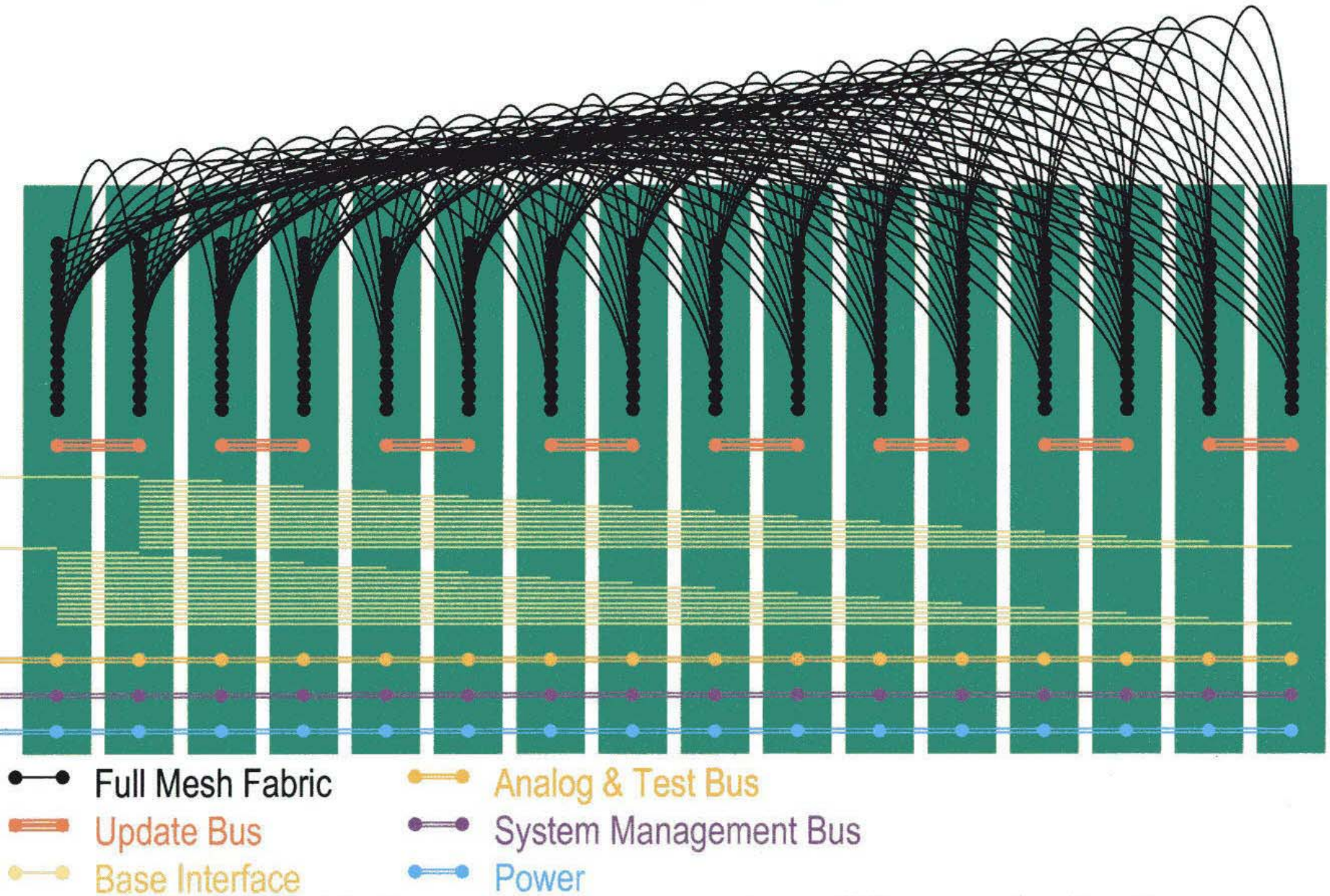

FUJITSU COMPUTERS
SIEMENS
hp
invent
KT
NTT
AdvancedTCA
intel



18 Layer / 36 Layer.

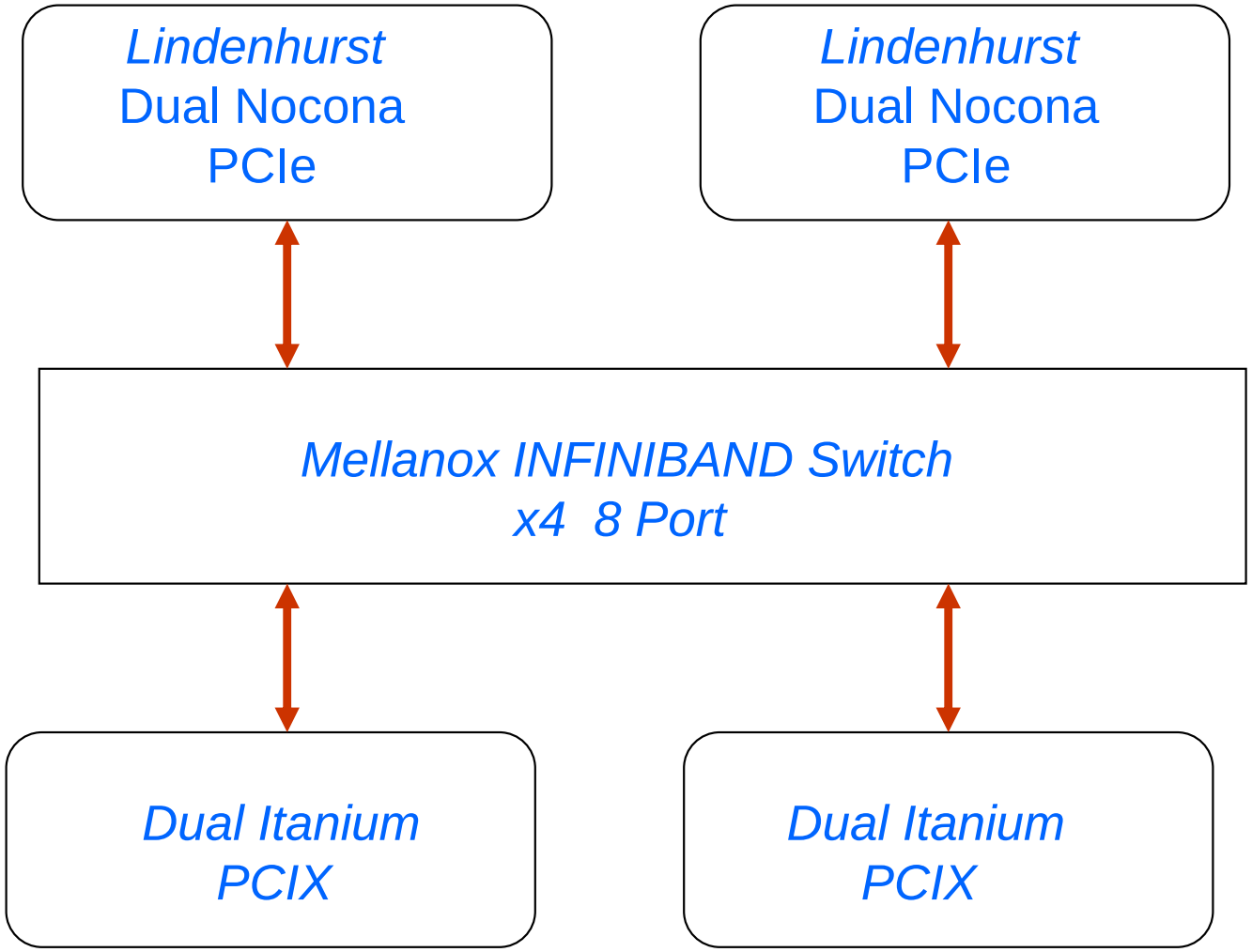
MAR 01 2004

PICMG 3 Backplane Topology



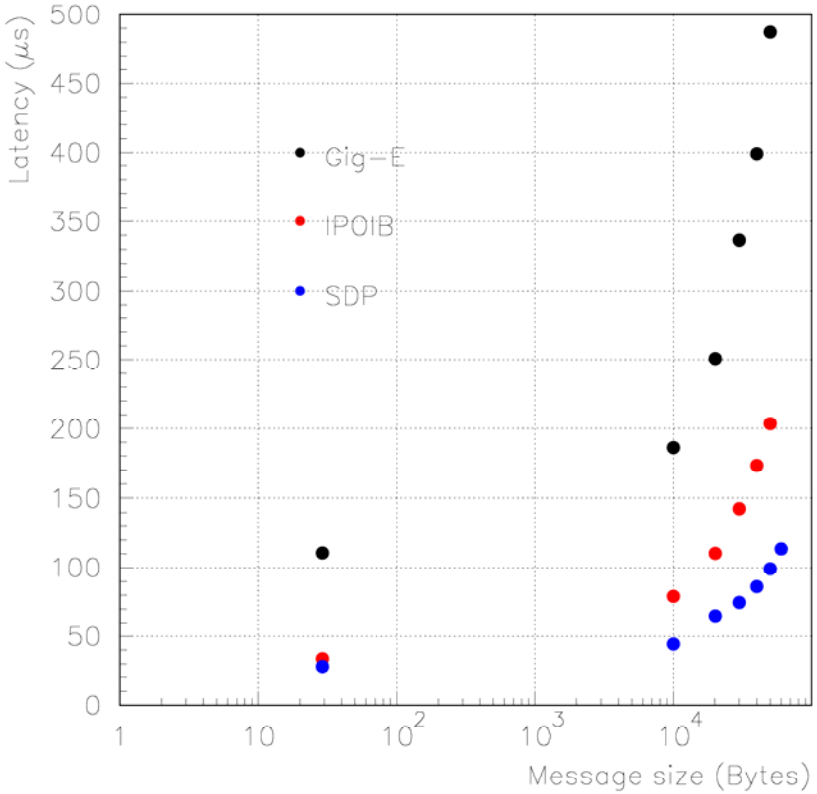
Need New Standard for Physics Instrumentation

- Based on Commodity components
- Signaling Standard implemented in Silicon
- Standard Hardware
- Serial interconnections
- Start a Standards Group –
  - NIM/CAMAC/FASTBUS/VITA
- Contact me at satish.dhwan@Yale.edu



Evaluation Hardware on Loan from Intel Corporation

Latency



igure 2: XDAQ RoundTrip benchmark. Latency versus message size for Gigabit thernet, IPOIB and SDP. Measured between Lindenhurst machines.

Transfer Rate

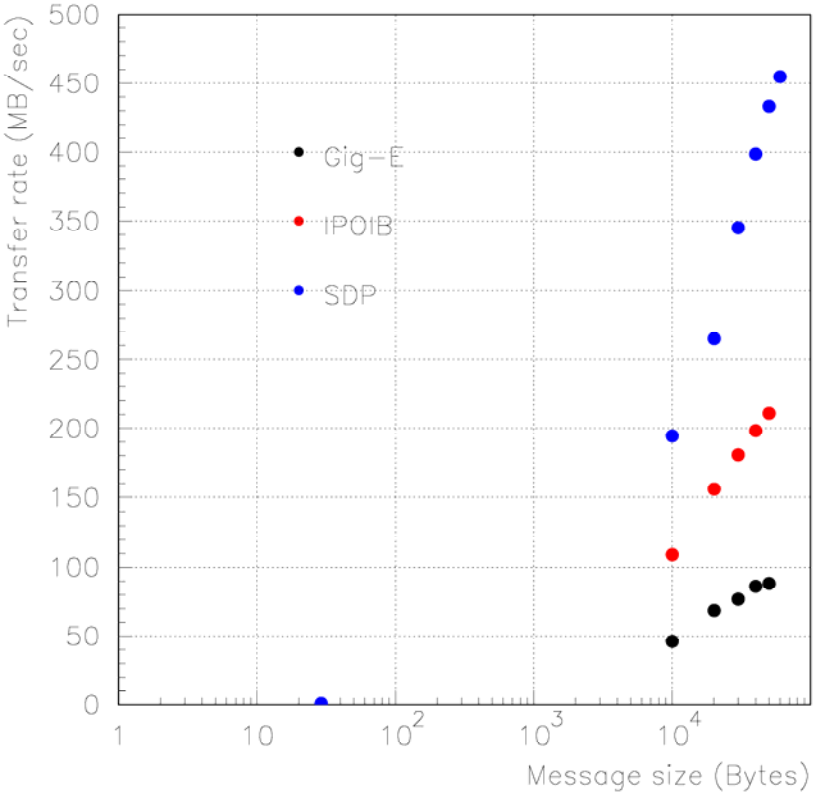
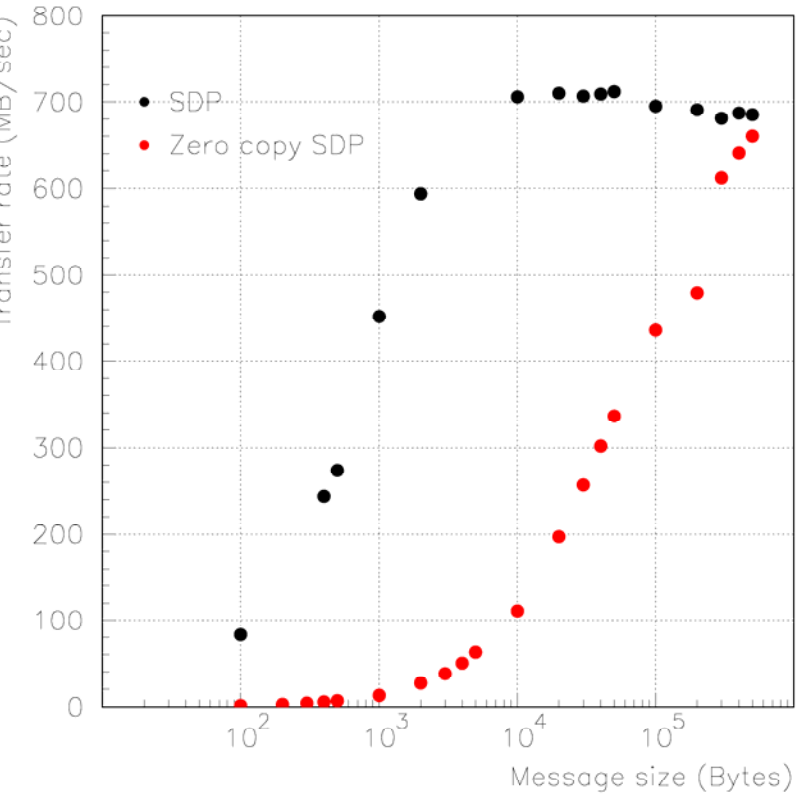


Figure 1: Test of XDAQ RoundTrip benchmark. Uni-directional data transfer rate between Lindenhurst machines in MByte/sec versus the message size. The cases of Gigabit ethernet, IPOIB and SDP are shown.

- Minimum Packet Size = 29 Bytes
- IPOIB: Internet Protocol over Infiniband
- SDP: Mellanox Drivers

Transfer Rate



ure 1: Uni-directional data transfer rate between Itanium machines in MByte/sec sus the message size. The cases of ordinary SDP and zero-copy SDP are shown.

CPU Utilization

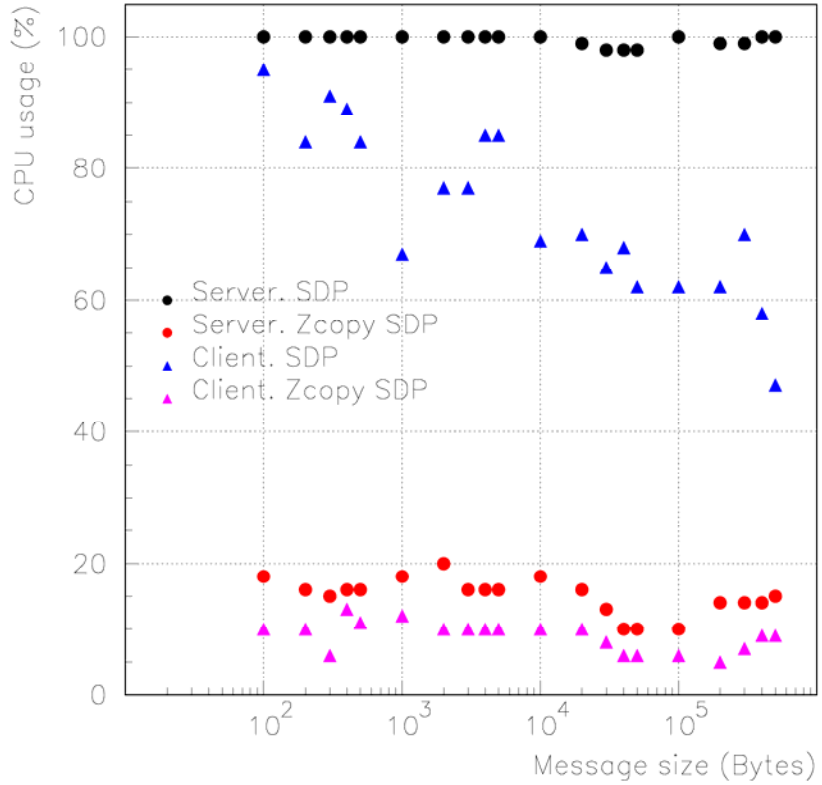


Figure 4: CPU usage on the server and client versus the message size for transfers between Lindenhurst machines. The cases of ordinary SDP and zero-copy SDP are shown.

- SDP: Mellanox Drivers
- Zcopy SDP