PCI SIG™

10TH YEAR ANNIVERSARY

Dave Dunning
Zale Schoenborn
Intel Corp.

2002 PCI-SIG Developers Conference Tour

Agenda

- **PCI Express Physical Layer (PHY) Basics**
  - ✓ Usages and requirements
  - ✓ Definitions
  - ✓ Signaling
- **PCI Express PHY Key Features**
  - ✓ Logical Sub-block
  - ✓ Electrical Sub-block
- **Summary**

PCI Express PHY Usages

Desktop
- Low Cost
- In-the-box Focus
- Chip-Chip
- Board-Board

Mobile
- Low Cost
- Low Power
- Small Form Factors
- Docking

Server
- Low Cost
- I/O expansion
- High Bandwidth
- Data Integrity

Comm
- Low Cost
- High Bandwidth
- I/O Connectivity
- Switching Capabilities

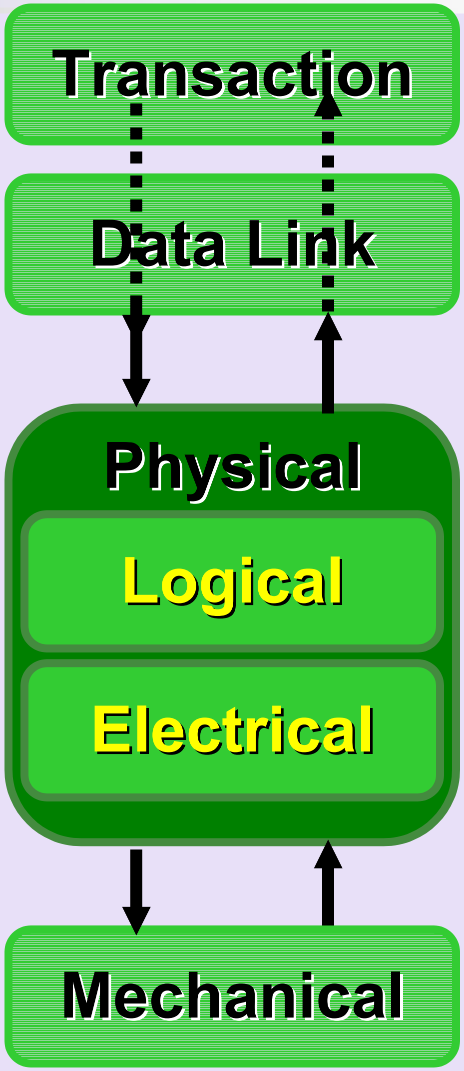
PHY Requirements

- Chip-to-chip data transfers through
  - ✓ 4 layer FR-4 PCBs
  - ✓ Silicon: small area, low power, low manufacturing costs
  - ✓ Low cost connectors
- Highest BW/$
  - ✓ Desktop, Mobile, Server, Comm segments
  - ✓ High Volume everything

Balance low cost with high performance

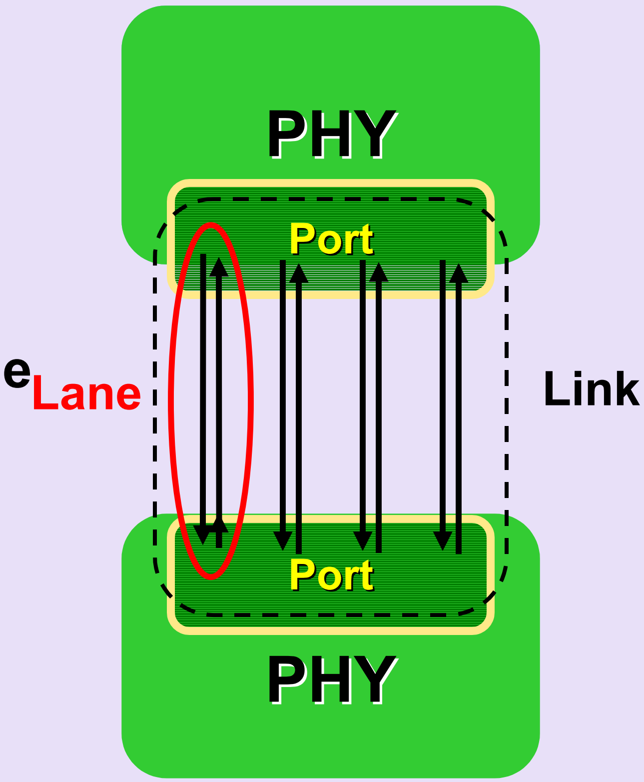
PHY Layer Partitioning

- Provides isolation from higher layers
  - ✓ PHY upgrades do not affect Data Link layer and above
- Logical Functions
  - ✓ Reset, initialization
  - ✓ Configuration: Rate, Link width, lane mapping and de-skew with status
  - ✓ Encoding/decoding/scrambling
- Electrical Functions
  - ✓ Packet Exchange
  - ✓ Power management mechanisms
  - ✓ Surprise attach/detach mechanisms



Key Definitions

- **Port**
  - ✓ A group of transmitters and receivers located on the same chip that define a link
- **Lane**
  - ✓ A set of differential signal pairs, one pair for transmission and one pair for reception
- **Link**
  - ✓ A dual-simplex communications path between two components
  - ✓ A xN link is composed of N lanes



Physical Layer Key Features

Link Configuration
Width Negotiation
Lane Reversal

Asynchronous Power Up and Power Down
Surprise Hot Swap

Power Management
3 modes that balance power vs. latency

Ease of Testing
Built in Compliance Pattern and Physical Layer Loopback

Silicon/Voltage Compatibility
AC Coupling and
Common mode Initialization

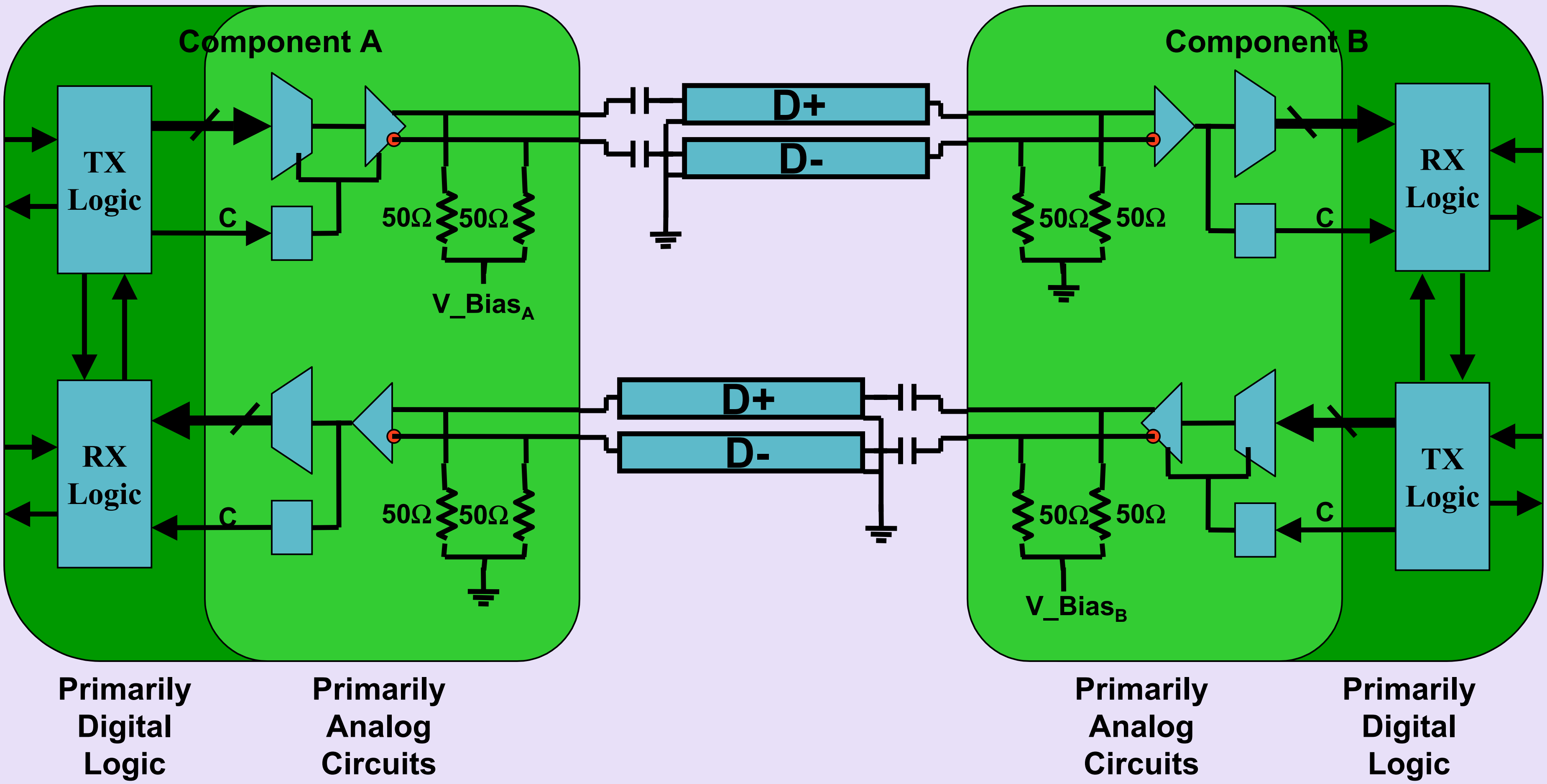
Embedded Clocking
Low Pin Count and
routing flexibility optimizes cost.
Frequency Scalability

Low EMI
Differential, low voltage swings
Randomized data patterns

No Sideband Signals
Lower Complexity
Low Pin Count optimizes cost

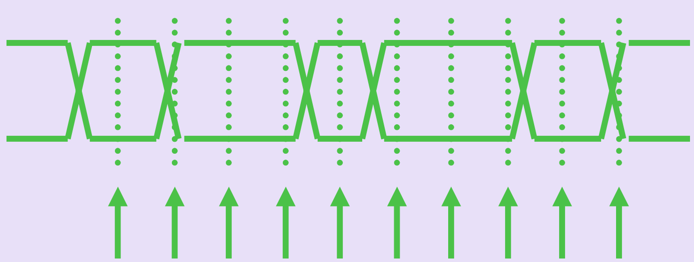
Advanced features beyond data transfer

PHY Pictorial; Example Only



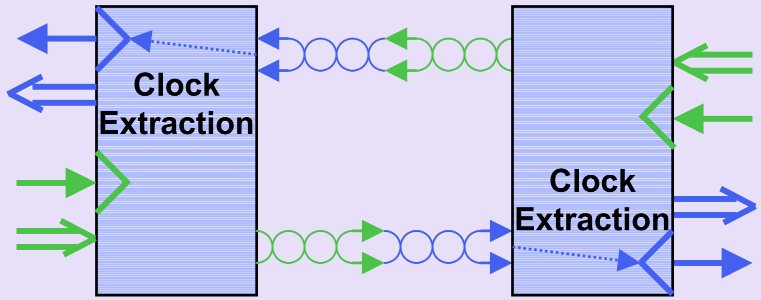
Embedded Clocking

- **Data transitions enable clock recovery at the receiver**
  - ✓ **Routing flexibility**
    - – **Removes skew relationships between clock and data signals**
  - ✓ **Higher performance**
    - – **High frequencies -> lossy transmission lines**
    - – **Skew/jitter between bits eliminated**



Clock Extraction

- 8b/10b encoding
  - ✓ Edge density facilitates clock extraction
  - ✓ DC balance allows AC Coupling
  - ✓ Frequency Scalability by eliminating the lowest 1/5 of transient bandwidth.
    - – Lowest 1/5 of bandwidth (<250Mhz) has approximately 40% of the loss seen between DC and 1.25Ghz.
- Allows for simple de-skew between lanes in a xN link (N ≥ 2)
  - ✓ All lanes use same source clock



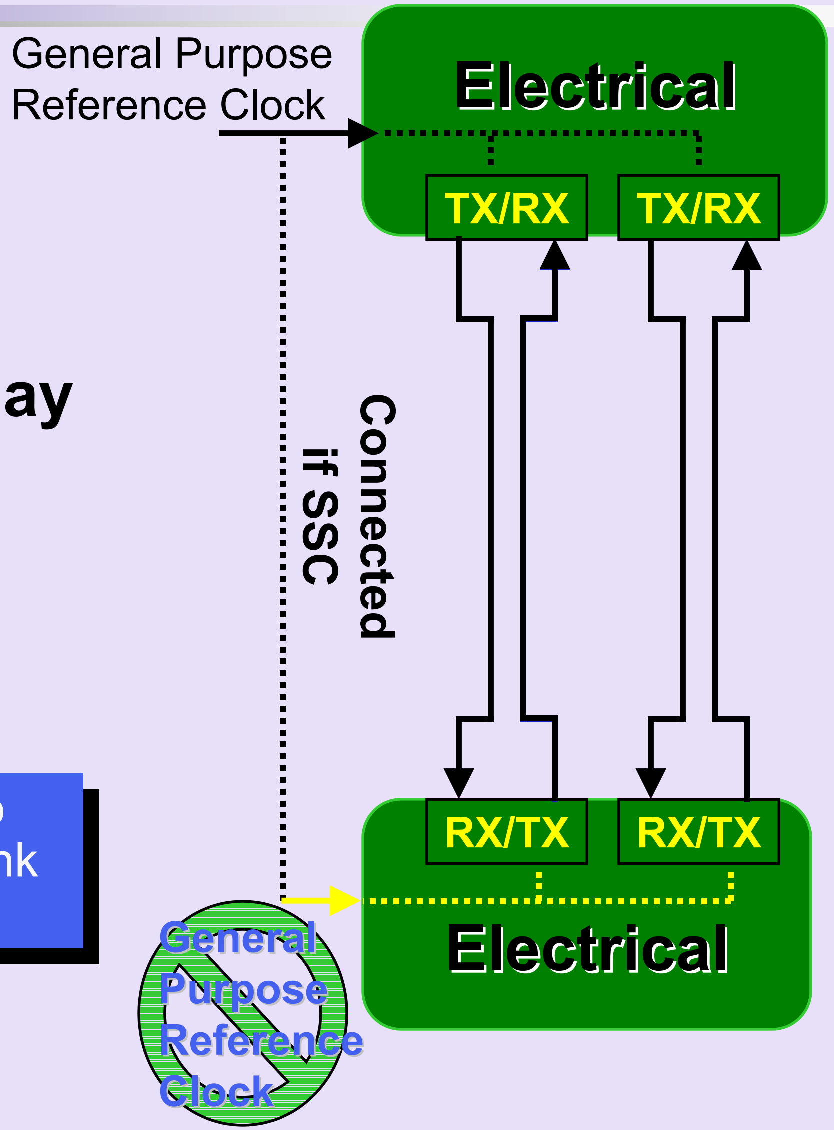
Clocking Options

- Distributed Frequency Reference
  - ✓Conducive to Spread Spectrum clocks
  - ✓Appealing for motherboard, add-in card systems
- Embedded/Derived from different sources
  - ✓Appealing where traces/pins are “expensive”
  - ✓Avoids any legacy clocks
- PCI Express allows for both
- Choice will be usage dependent

Clocking Options

- **All lanes within a port must transmit data using one frequency**
- **The two ports within a link may transmit data at different frequencies**
  - ✓ **Tolerance = <u>+/-300ppm</u> each**

<u>If</u> a spread spectrum clock (SSC) is used to modulate the data rate then both ports in a link must use the same modulation source



Lower EMI

- Low voltage reduces power and therefore potential for emissions
- Differential (balanced) signaling limits emitted energy
- Data scrambling eliminates repetitive data patterns on the wires
  - ✓ Minimizes strong harmonics and reduces EMI

Designed for Low EMI

Power Management

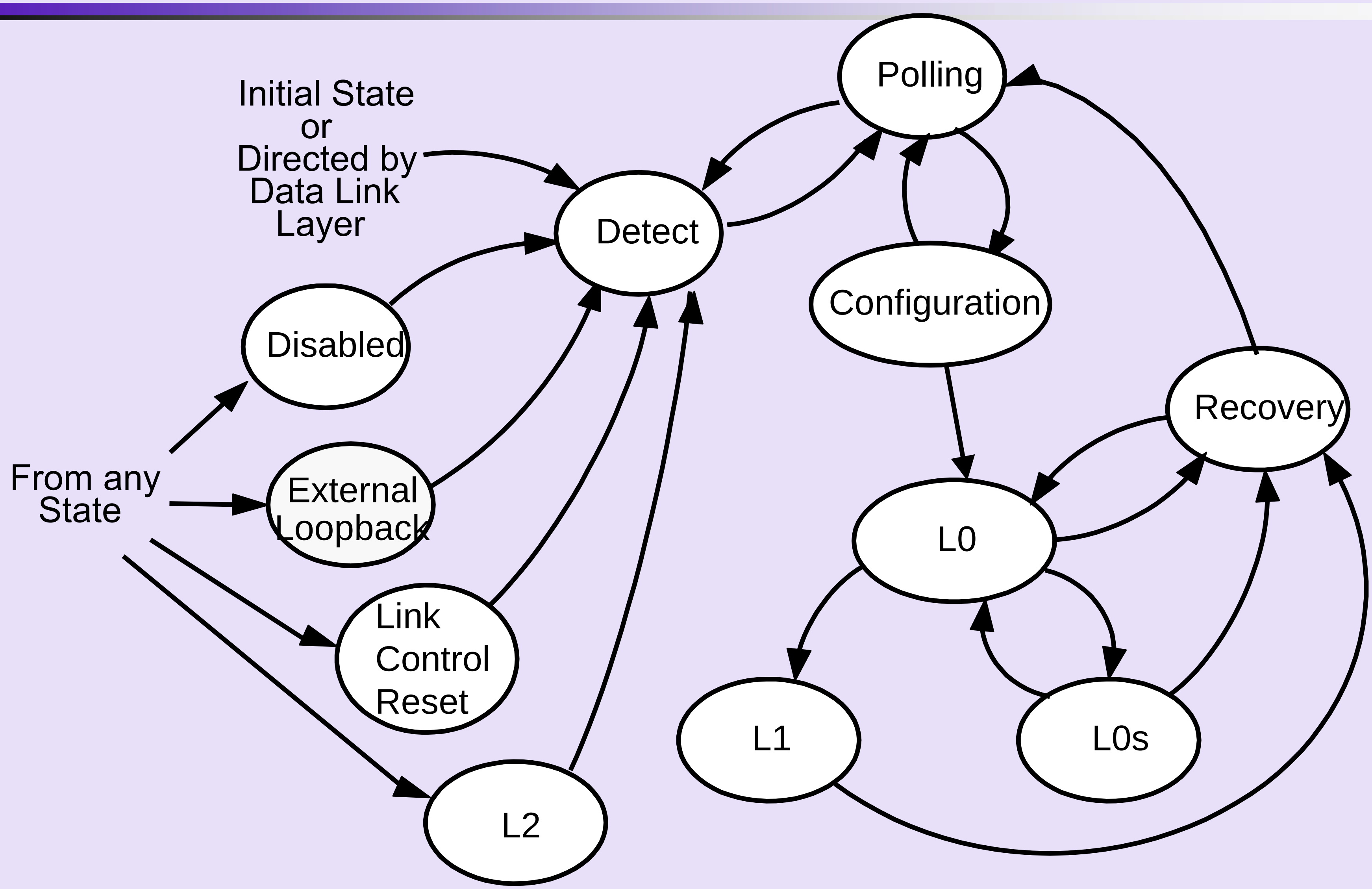
- 3 physical level states
  - ✓ Comprehends tradeoff between lowering power and quick response/wakeup time
  - ✓ Allows choices at system level
- Power: L0 > L0s > L1 > L2
  - ✓ L0 - normal operation
  - ✓ L0s – optional state
  - ✓ L1 – Clocks may be turned off, power remains
  - ✓ L2 – Main power removed, aux power only

Three levels for power/performance flexibility

PM PHY States

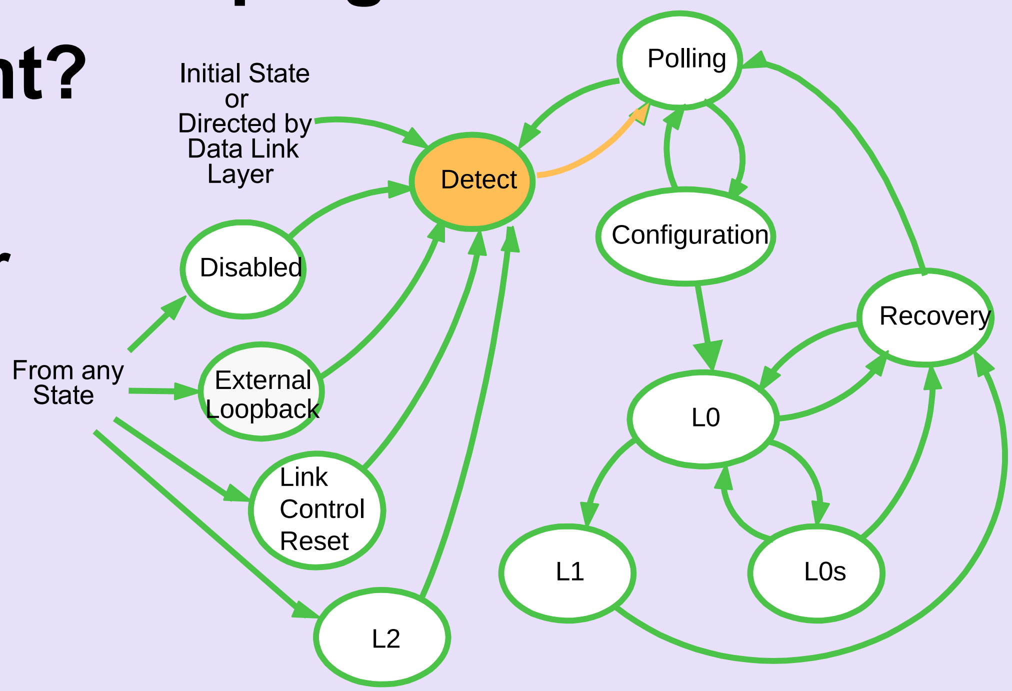
State	Description	Latency-->L0
L0	Channel active	Not Applicable
L0s	Channel pulled to common mode voltage On-chip clocks on Power on	16 nsec – 4 usec
L1	Channel pulled to common mode voltage On and off-chip clocks may be turned off Power on	1 – 10s of usec
L2	On-chip clocks not running Main power off, Aux power (if available) only	System Dependent (PLL spin-up time, Power supply Settling time)

Link Training & Status State Machine



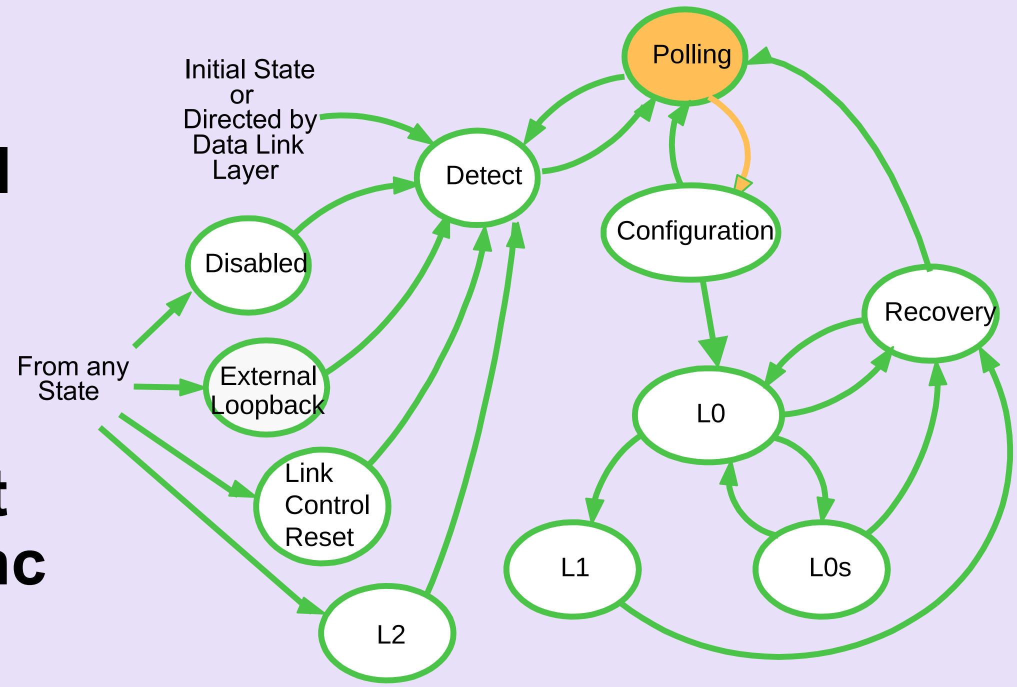
Detect

- **First state entered after power-on/reset or link detach**
- **Protects receiver against over-voltage during surprise attach and hot plug**
- **Is receiver present?**
  - ✓ **Yes: enter polling**
  - ✓ **No: try again later**



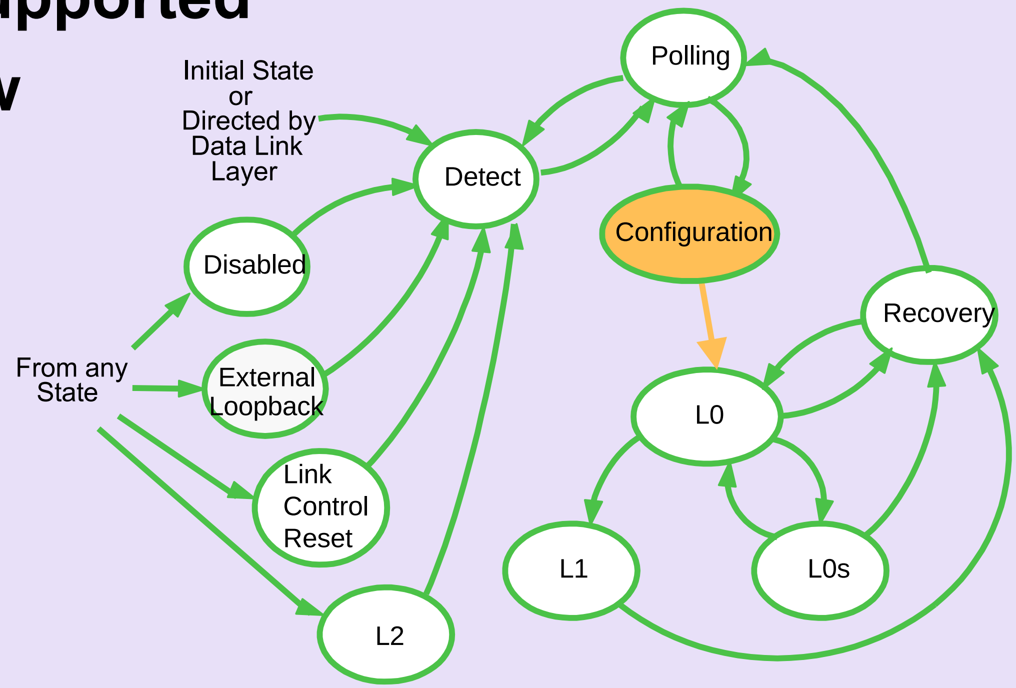
Polling

- Entered from Detect, Configuration or Recovery
- Establishes “health” of each lane
  - ✓ Bit, symbol synchronization
  - ✓ Polarity reversal if necessary
  - ✓ Lane data rate; future revisions
  - ✓ Compliance test pattern if no sync



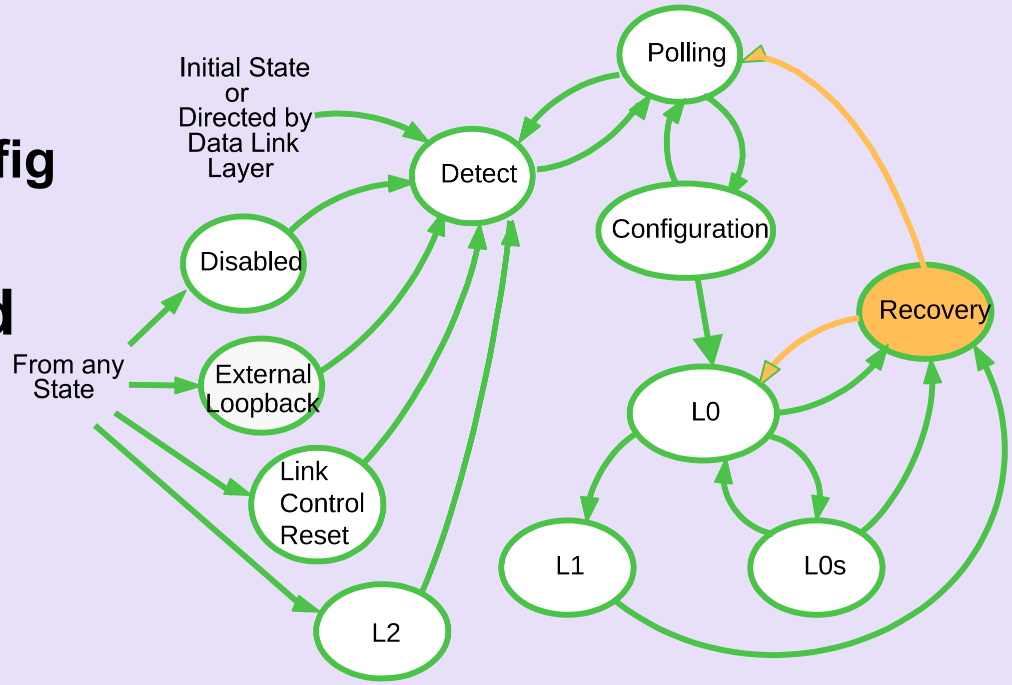
Configuration

- Entered from Polling
- Links formed from lanes
  - ✓ Width (# lanes in) link
  - ✓ Lane reversal if supported
  - ✓ Lane-lane de-skew within each link



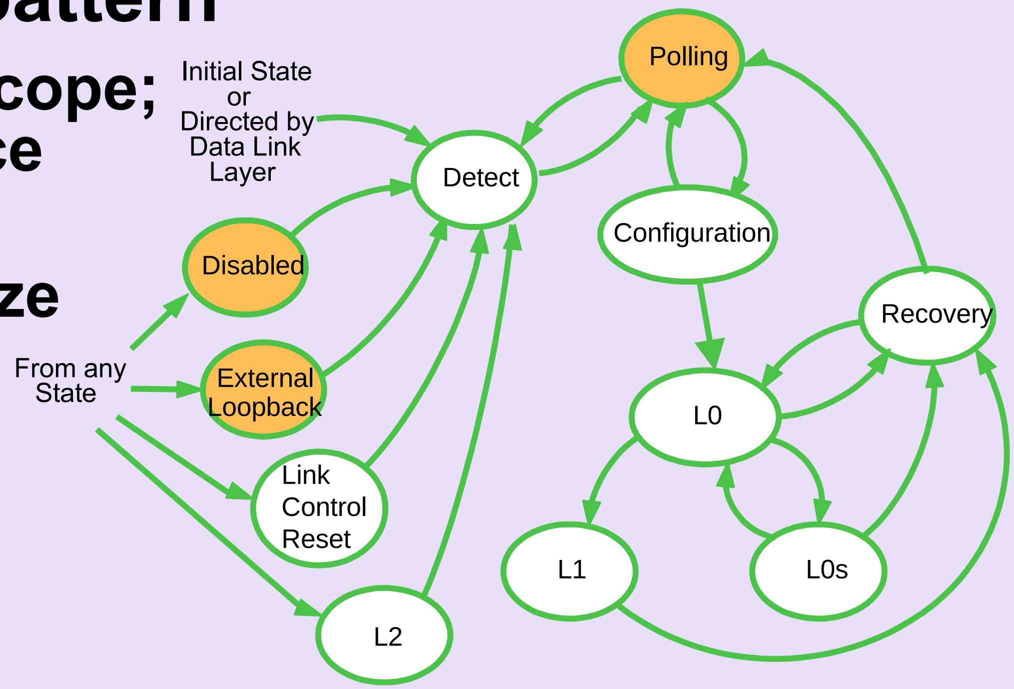
Recovery

- Entered from L0, L0s, L1
- Re-establish bit, symbol sync, polarity inversion if needed
  - ✓ Return to L0
    - – Retains previously negotiated Link config
  - ✓ Return to polling if power was removed



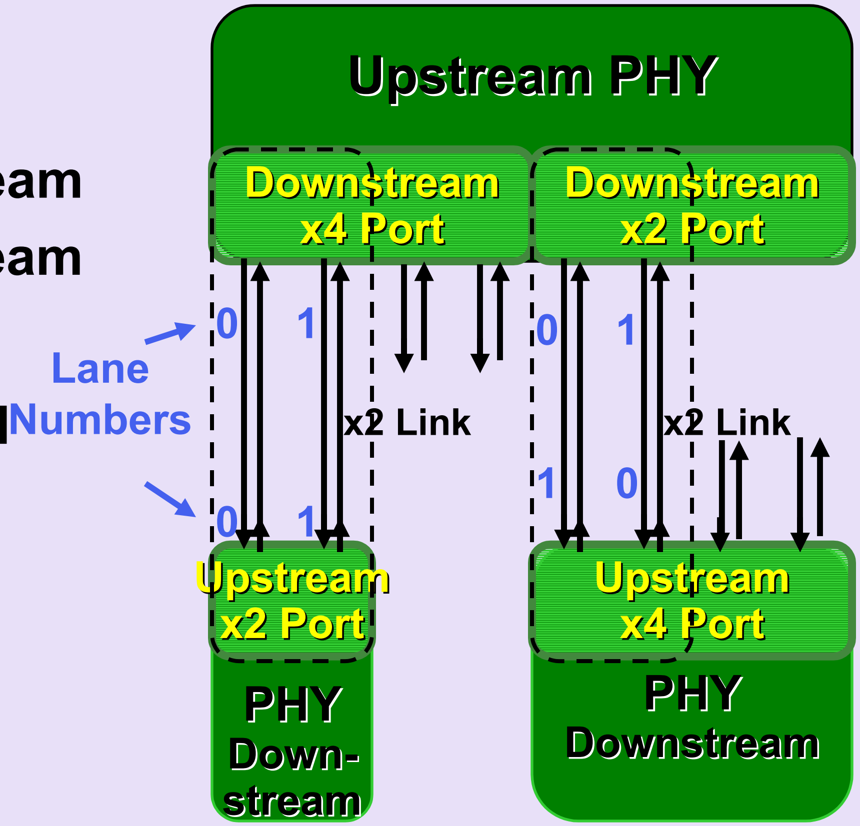
Test/Debug/System

- External Loopback Sub-state
  - ✓Loops data back; receive to transmit.
  - ✓Allows for system and component level BIST testing.
- Compliance test pattern
  - ✓Attach an oscilloscope; outputs compliance pattern
  - ✓Chosen to maximize ISI and crosstalk.
- Disabled
  - ✓"Turns off" ports



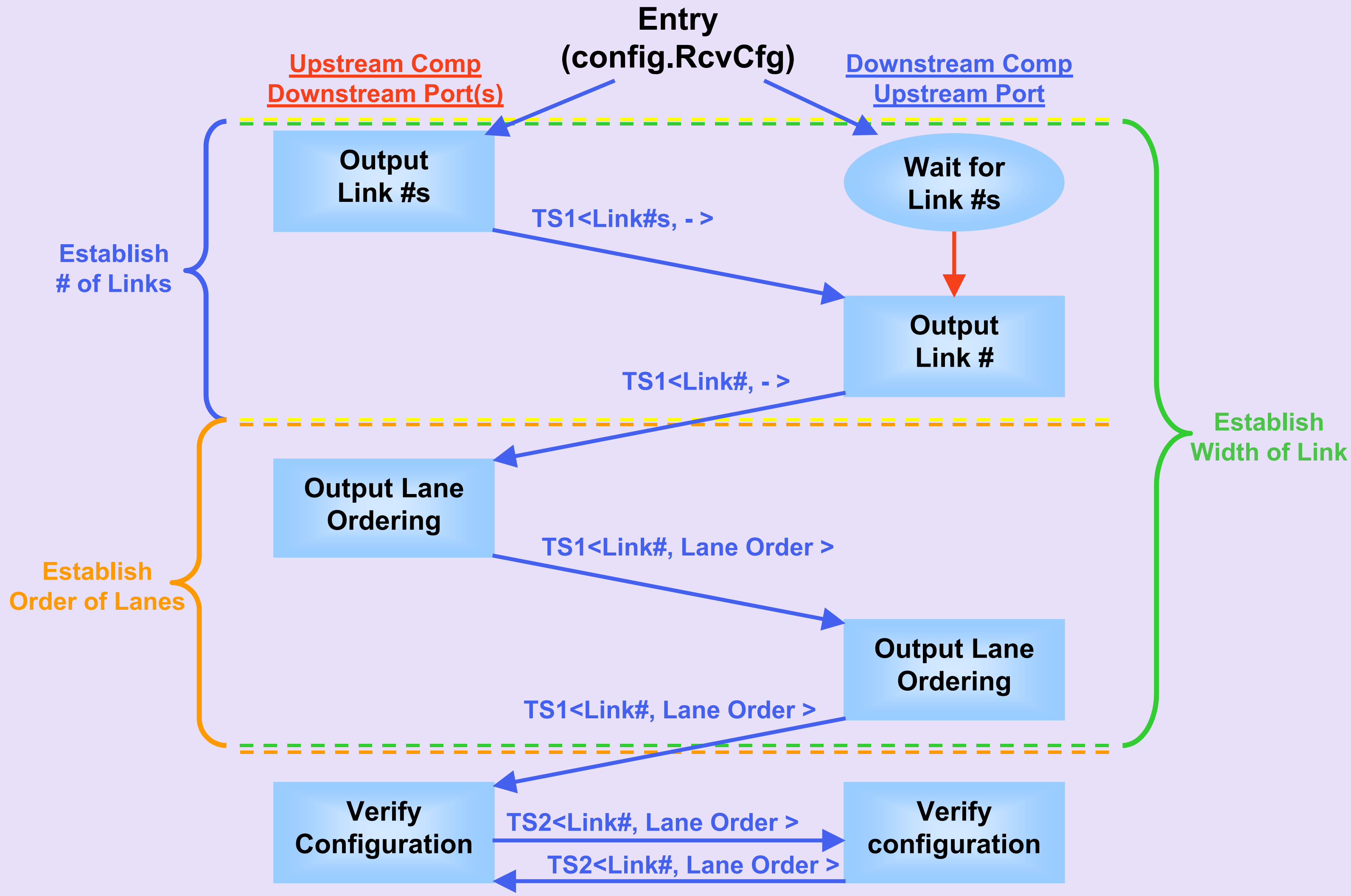
Link Configuration

- **Link Width**
  - ✓ Upstream wider than downstream
  - ✓ Downstream wider than upstream
- **Lane Ordering/Reversal**
  - ✓ Lane ordering can be swapped within a link

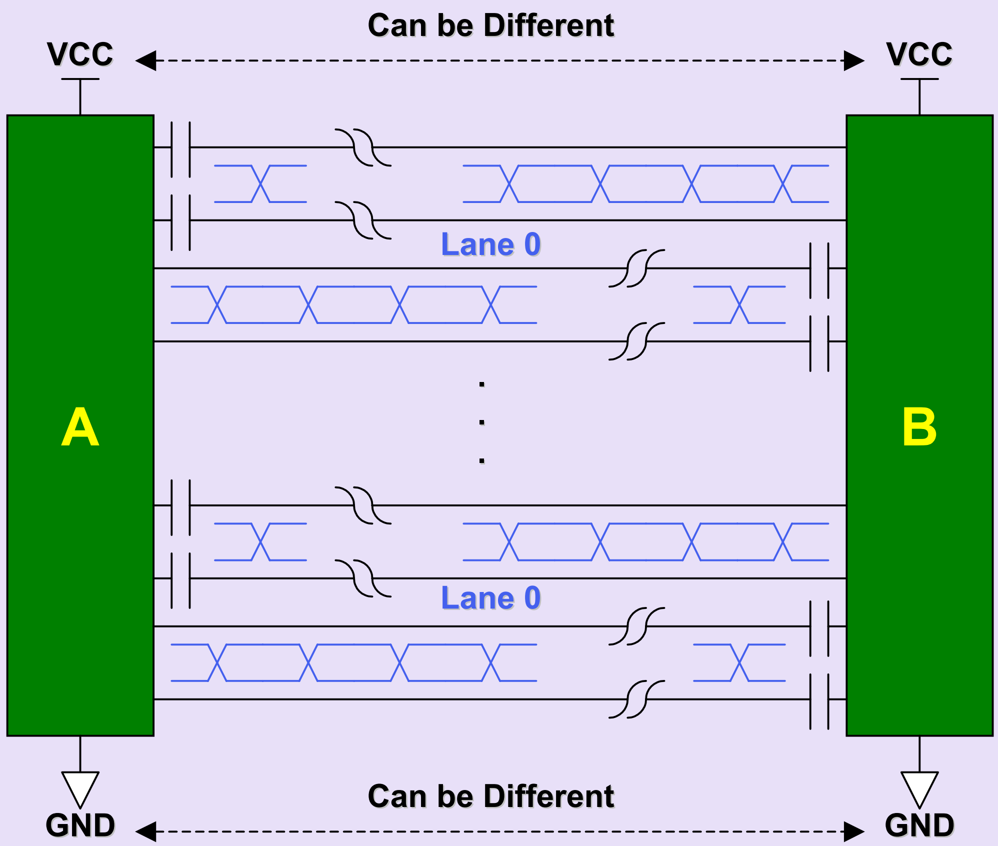


Simplifies PCB Routing

Link Width Negotiation

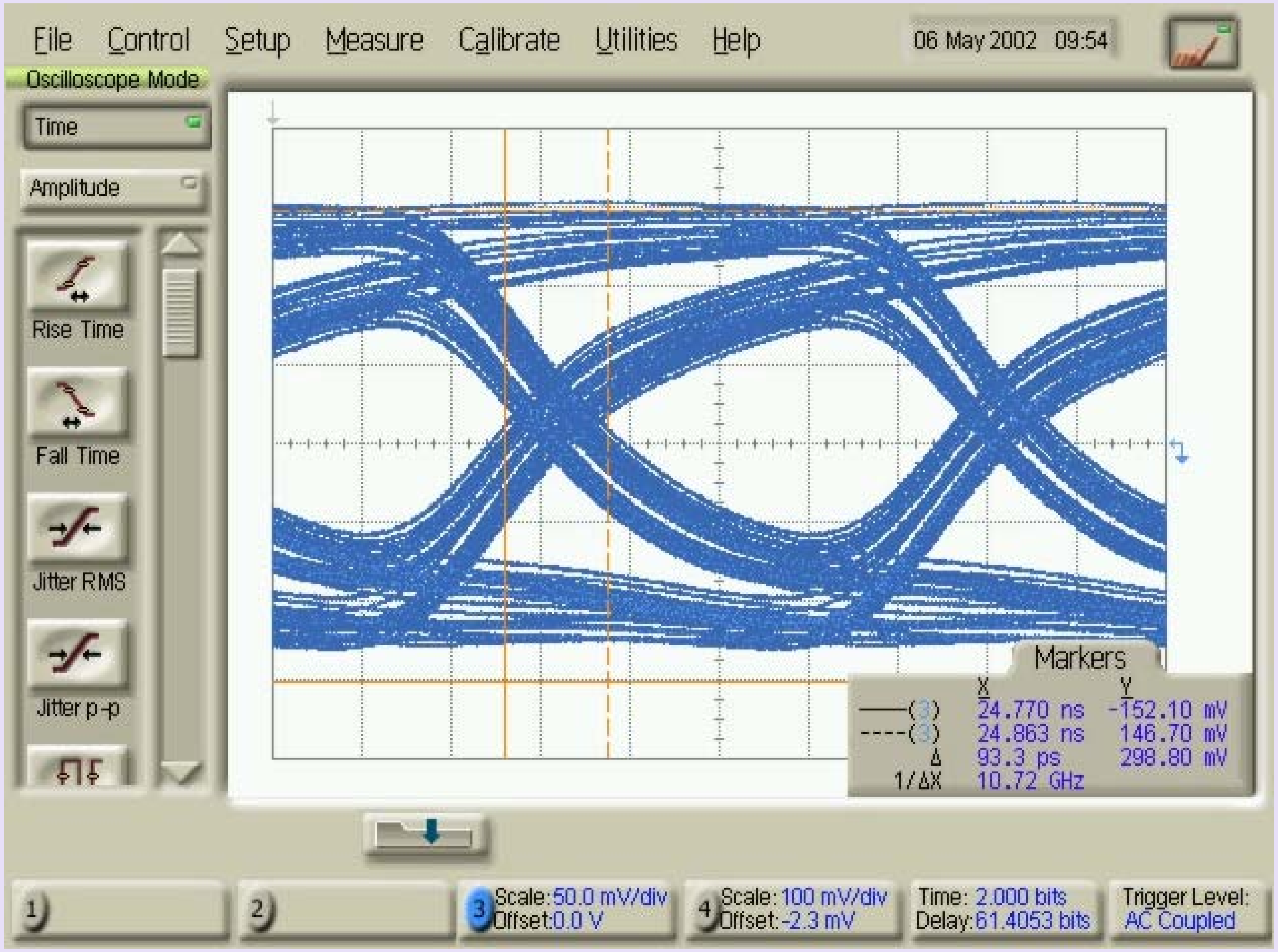


Signaling



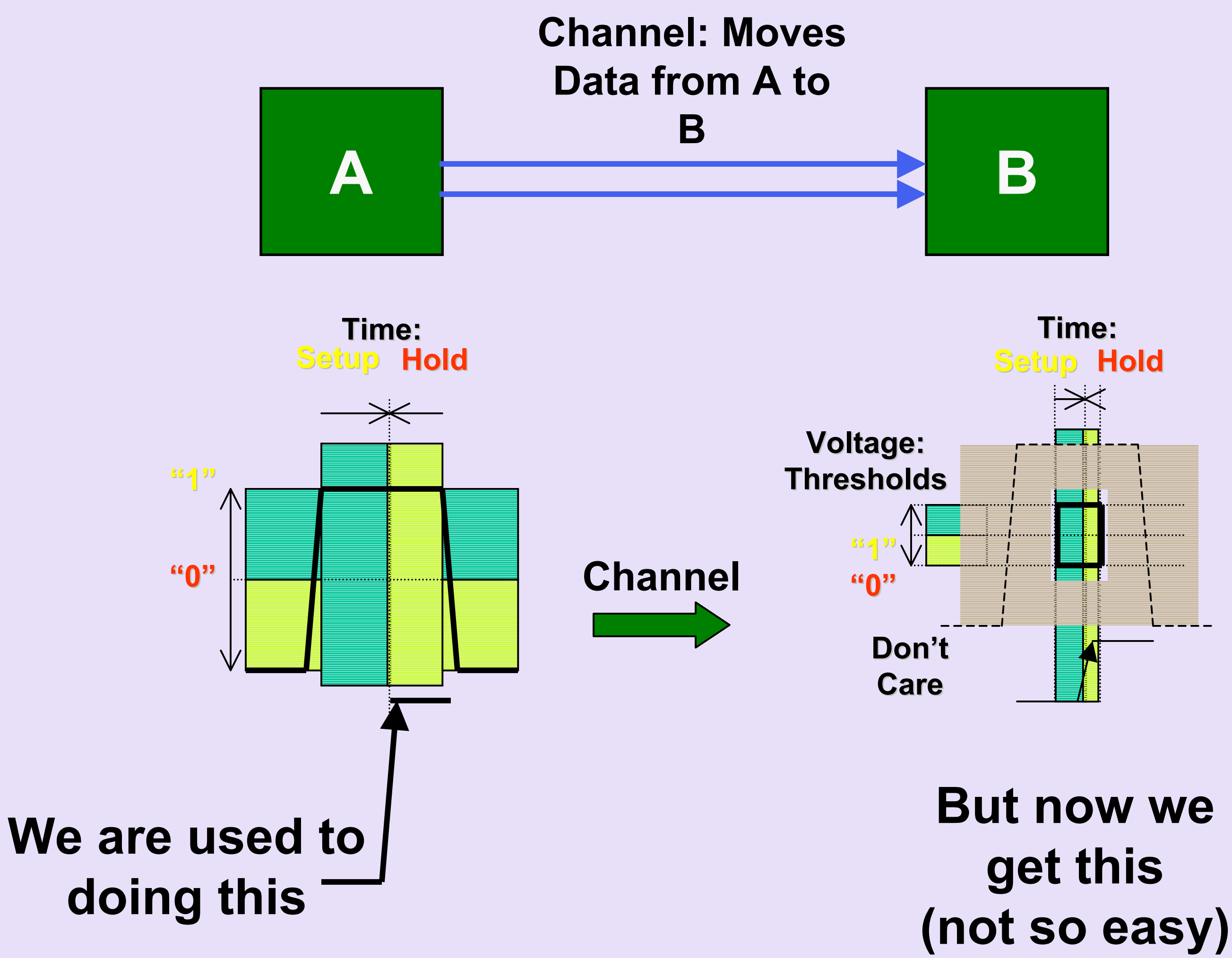
- **Differential**
  - ✓ Superior voltage margins to single-ended
  - ✓ Voltage independence; VCC can vary
  - ✓ Reduces EMI at a given frequency
- **Serial Lanes**
  - ✓ Higher Frequency; Removes skew requirements between parallel lanes
- **Relaxed routing rules within a link**

Example Eye Diagram

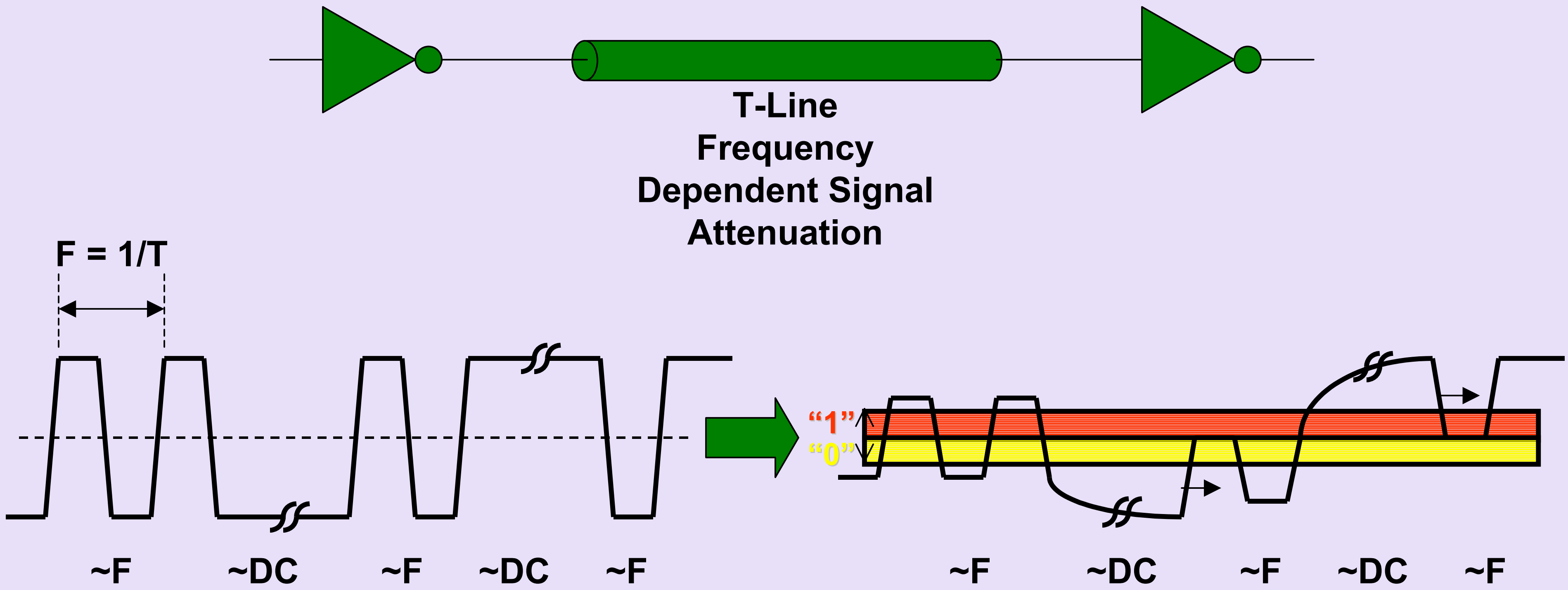


Source: A. Tripathi, Intel Corp

Electrical Specifications

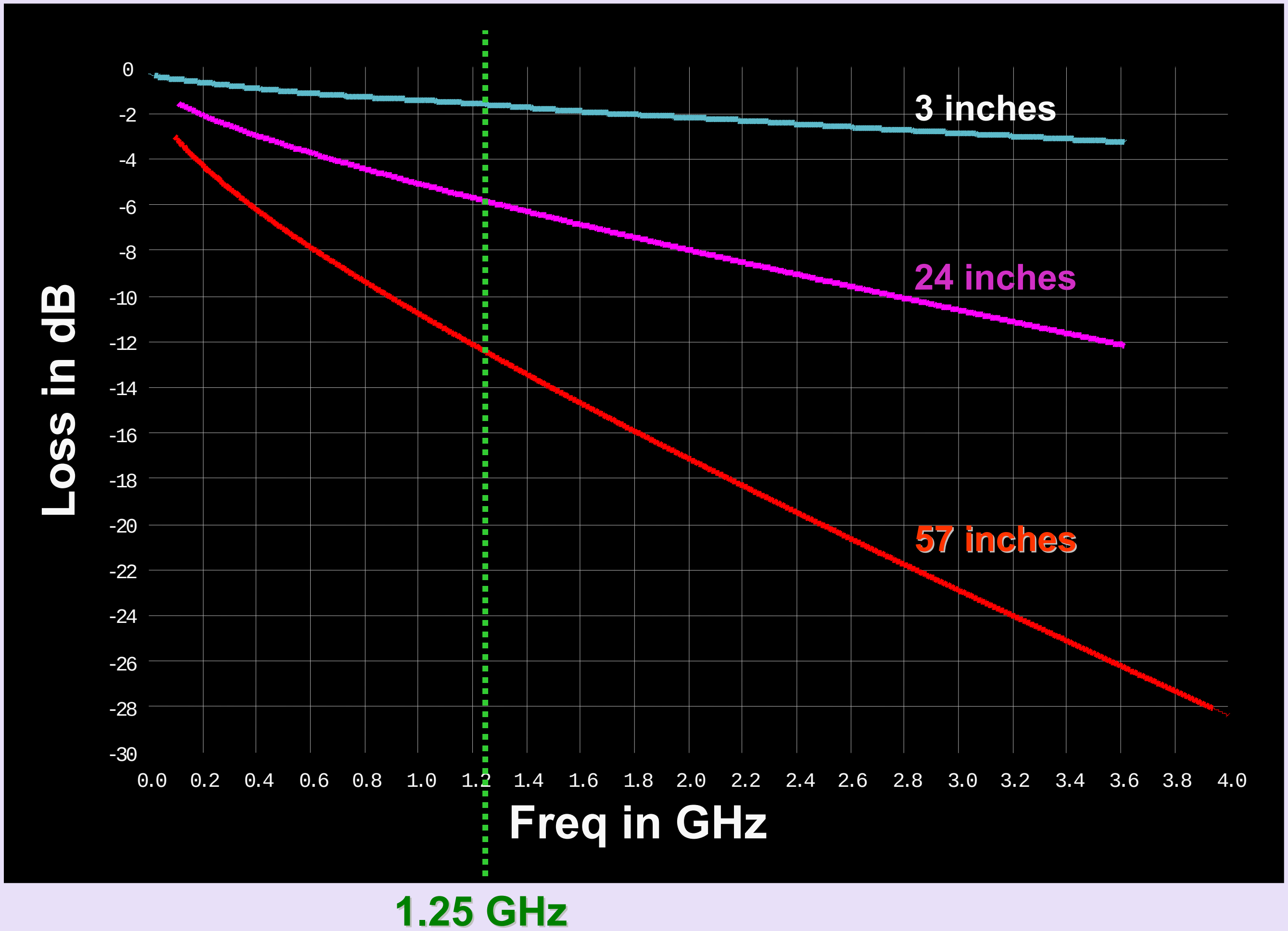


Frequency Dependent Loss



Lead to choice of Differential Signaling

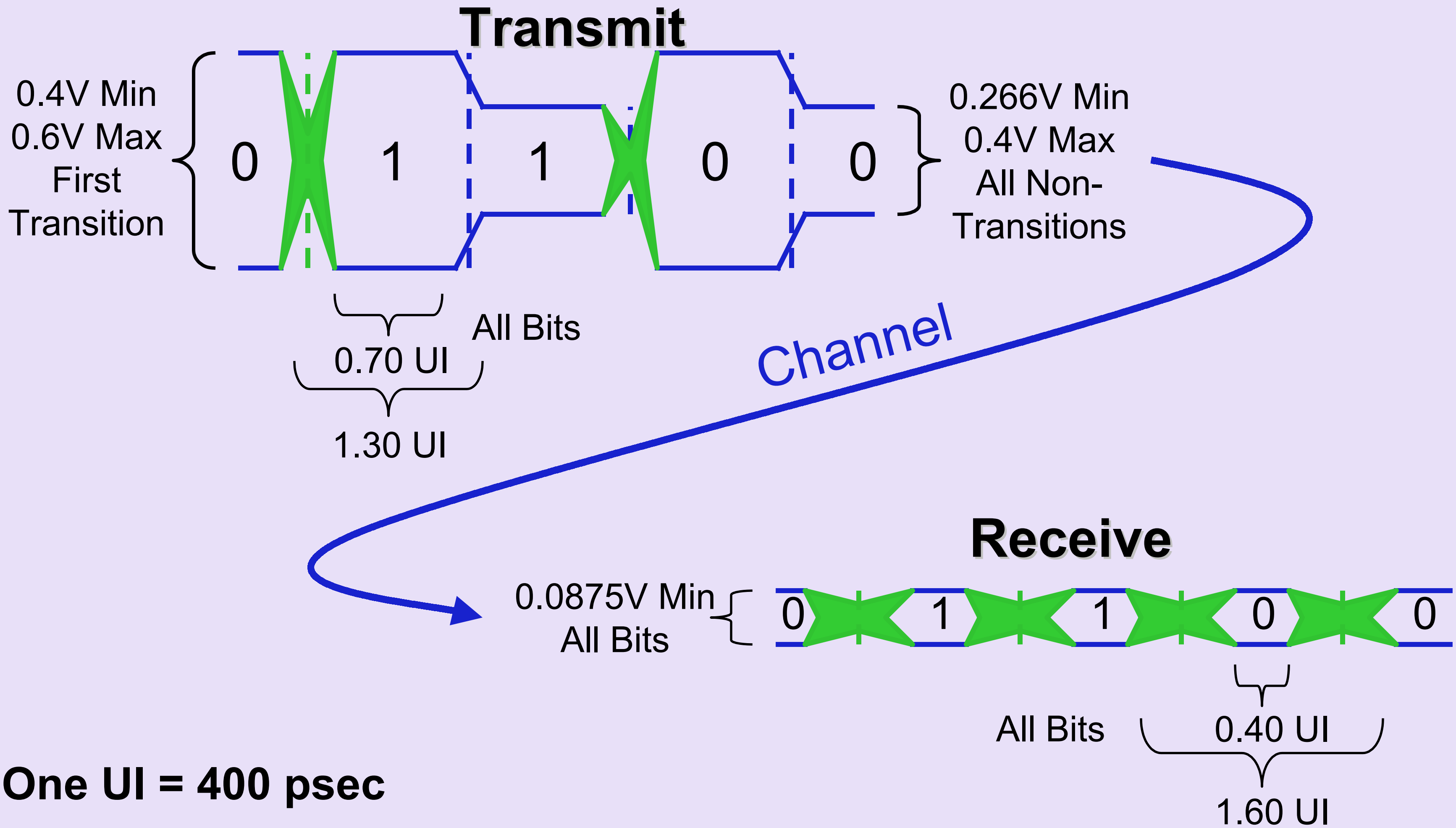
Channel Example; Loss/inch FR-4**



** Distances not expected to be representative of an actual channel

Used for example purposes only

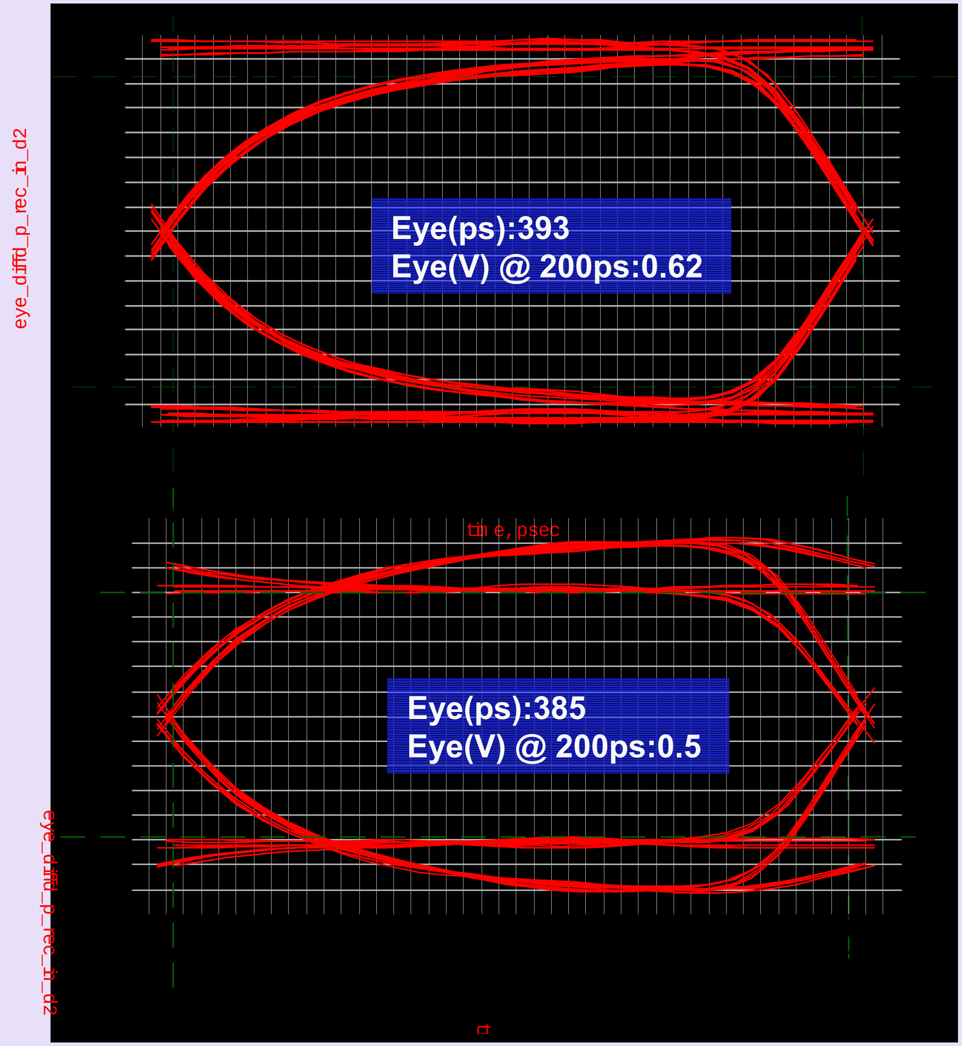
Electrical Specifications



One UI = 400 psec

Low Loss Channel

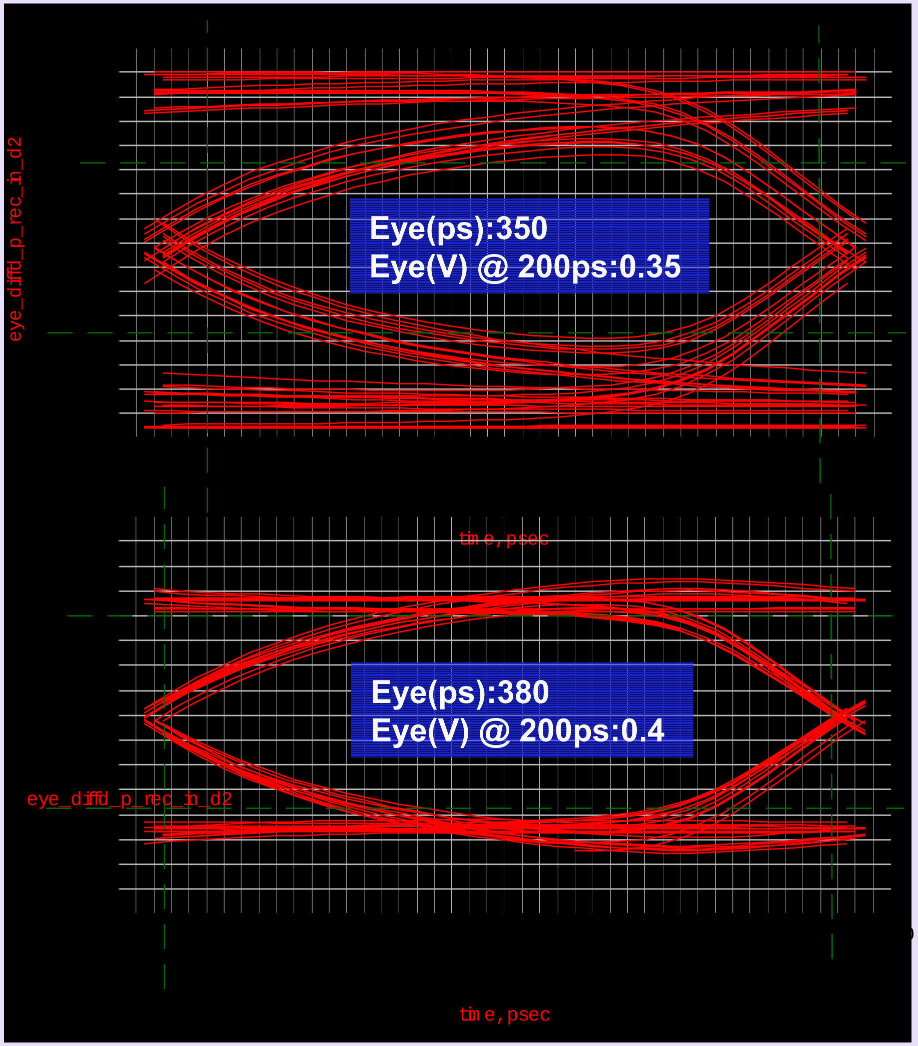
- **Channel; <2 dB of loss at 1.25 GHz**
  - ✓ **0 dB De-Emphasis**
  - ✓ **-3.5 dB De-emphasis**



Source: T. Secasiu, Z. Schoenborn, Intel Corp.

Medium Loss Channel

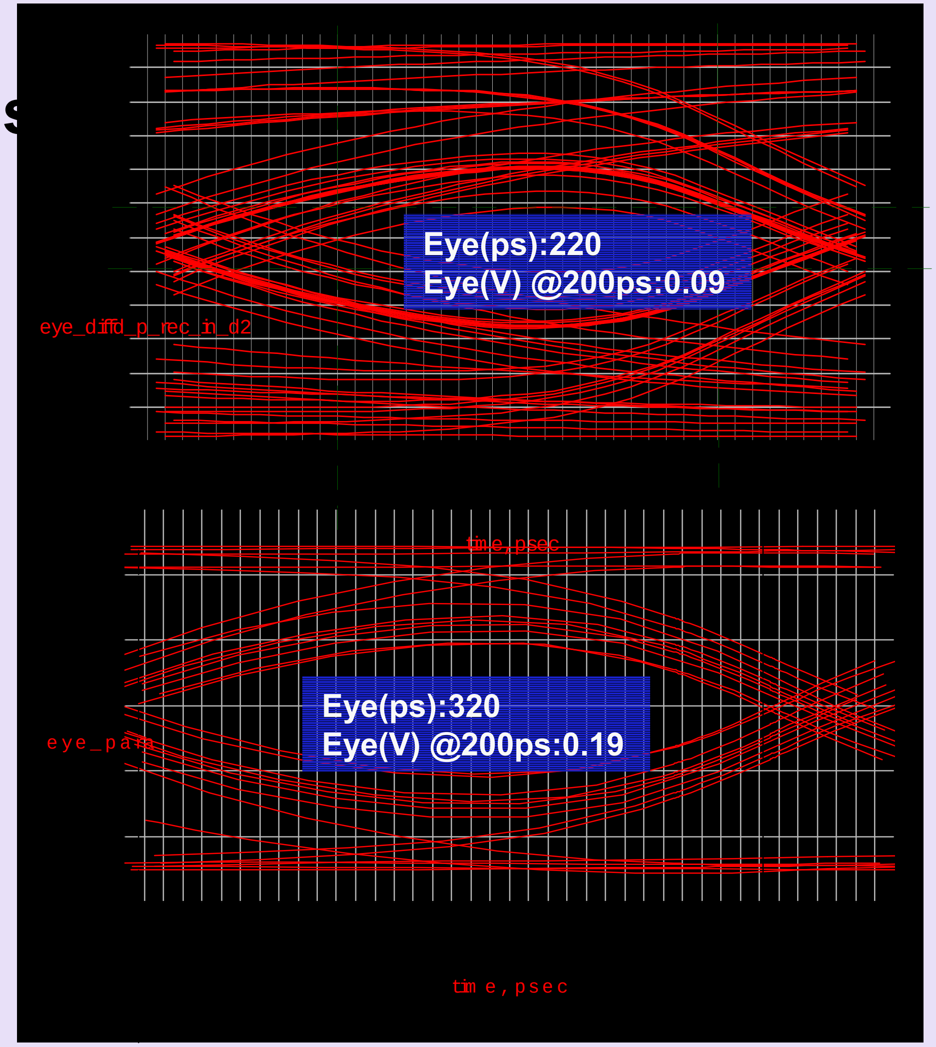
- Channel; ~6 dB of loss at 1.25 GHz
  - ✓ 0 dB De-emphasis
  - ✓ -3.5 dB De-emphasis



Source: T. Secasiu, Z. Schoenborn, Intel Corp.

High Loss Channel

- Channel; ~12.5 dB of loss at 1.25 GHz
  - ✓ 0 dB De-emphasis
  - ✓ -3.5 dB De-emphasis



Source: T. Secasiu, Z. Schoenborn, Intel Corp.

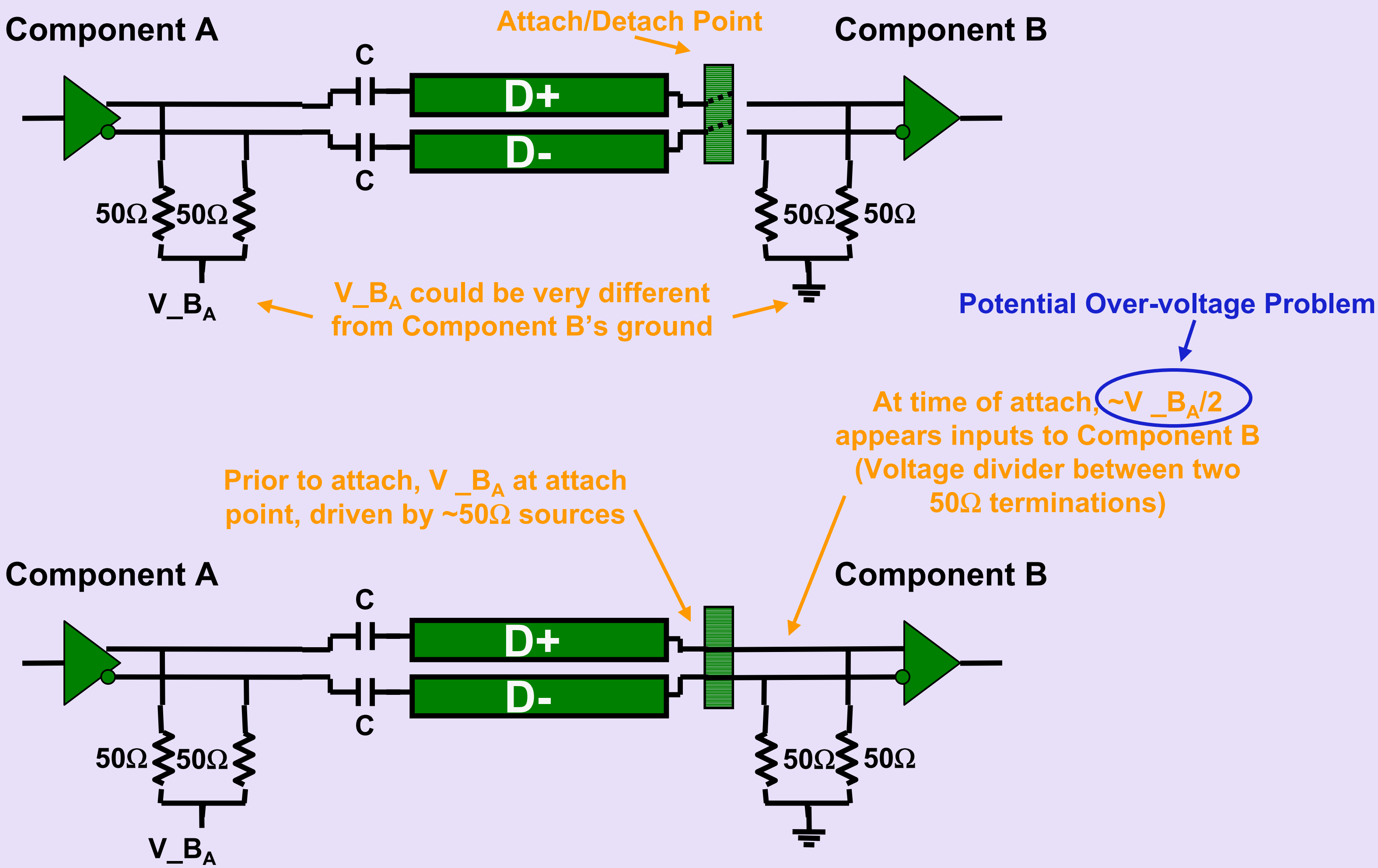
Termination

- **50 ohms single-ended**
  100 ohms differential
  - ✓+/- 10% tolerance
  - ✓Good match for test equipment
  - ✓Fair match for FR-4
- **Channel impedance not specified**
  - ✓AC and DC specs define channel requirements
    - e.g. Loss, jitter, AC common mode

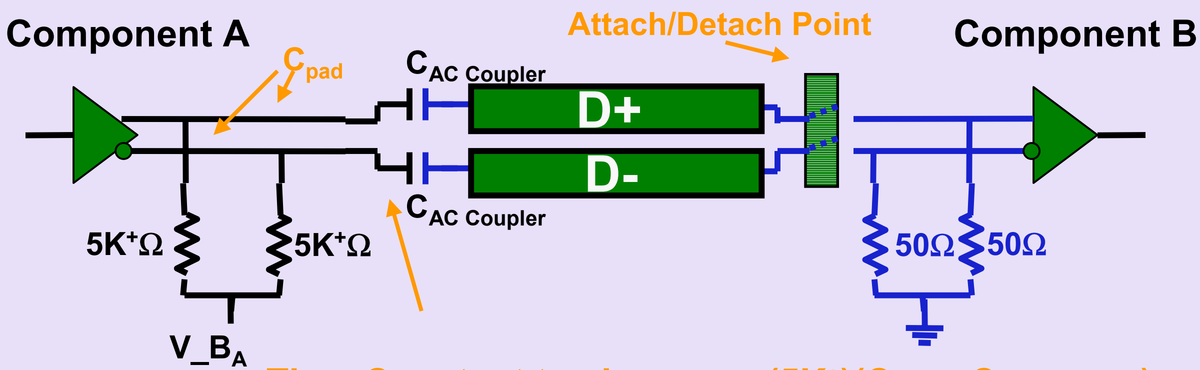
AC Coupling

- AC Coupling required on all channels
- Capacitance in range of 75 nF – 500 nF
- Located at transmitter end of channel
  - ✓Applicable to channels with connectors
- Transmitter, receiver common mod voltages do not need to match
  - ✓Allows for many different TX, RX designs
  - ✓Allows for simpler mixing of processes
  - ✓Allows for simpler mixing different supply voltages

Surprise Attach/Detach

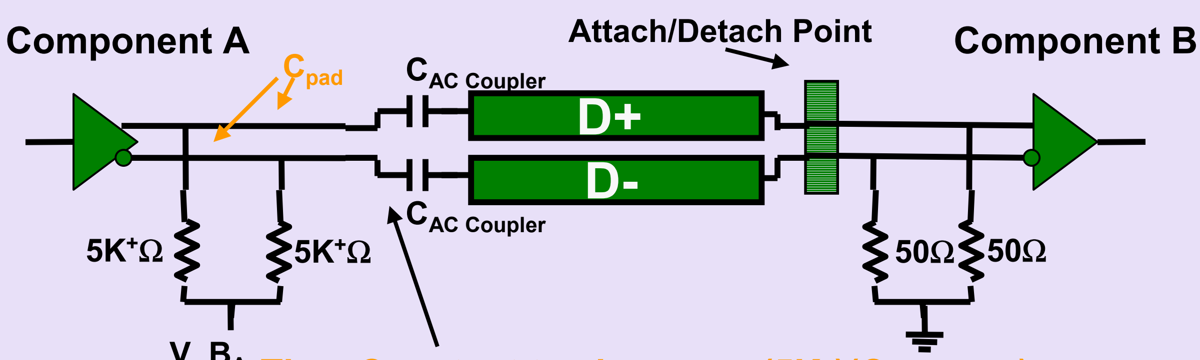


Surprise Attach/Detach



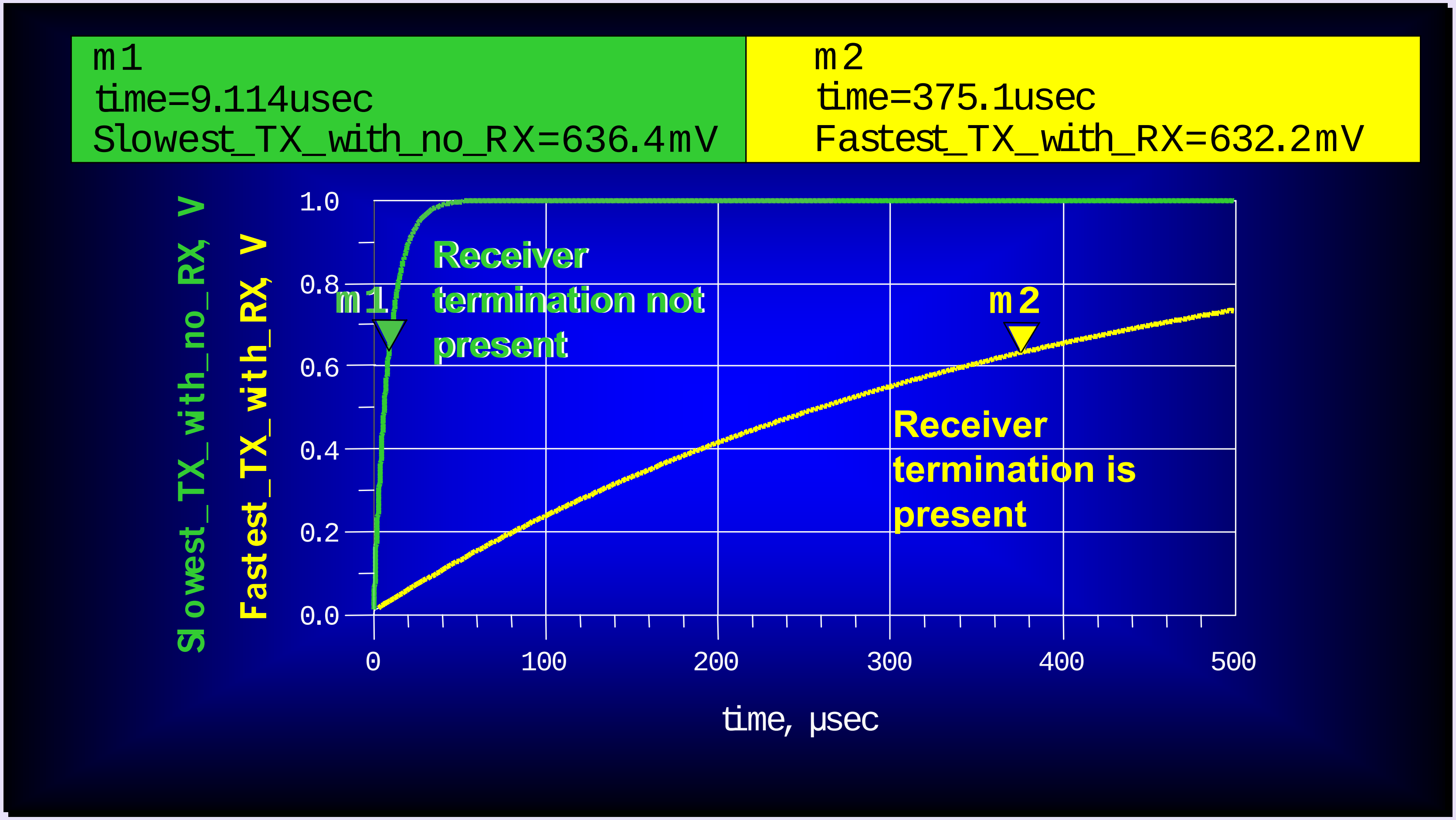
Time Constant to charge ~= $(5K^{+})(C_{pad} + C_{interconnect})$
= Small compared to below

$C_{pad} + C_{interconnect} << C_{AC\ Coupler}$



Time Constant to charge ~= $(5K^{+})(C_{AC\ Coupler})$
= Large compared to above

Time Constants Simulations



- **Time constants differ by a safe margin of ~40x**

Future Scaling

- **2.5 Gbps is the first PCI Express stop**
  - ✓ Benefit from many other differential interface learning curves
  - ✓ Roadmap for 2x, 4x planned
- **Multiple technology paths to investigate**
  - ✓ Evolution of signaling technologies
  - ✓ Channel characteristics and cost
  - ✓ Gain experience with noise sources

PCI Express PHY Summary

- The PCI Express PHY technology addresses interconnect bottlenecks
  - ✓ Balances low cost with high performance
- Investment in PCI Express PHY technology applies today
  - ✓ Friendly to future low voltage processes
- Advanced features beyond transfer speed
  - ✓ Auto-configuration, simple clock solutions, Low EMI, choice of power management mechanisms, Hot plug, Ease of Testing, and Process/VCC Interoperability.
- The technology is mature and silicon ready

Backup

Physical Layer Goals

- **Chip-chip data transfer through**
  - ✓ In-the-box "focus"; intra-board, board-board
- **10 year lifespan**
  - ✓ Tolerate ~0.8v ≤ VCC ≤ 2+v range
  - ✓ Technology friendly (higher frequencies)
- **Highest bandwidth/$ in computer and communication systems**
  - ✓ CMOS driver/receivers
  - ✓ 4 layer FR-4 boards; 20 – 30 inches, 3 connectors
  - ✓ High volume everything (connectors, clock sources…)

Electrical Specs

- **Differential**
  - ✓ 13 db Loss allowed across channel
  - ✓ Superior margins to single-ended signaling
  - ✓ Reduces EMI
- **Serial-Based Technology**
  - ✓ 0.30 UI budgeted to transmitter
  - ✓ 0.40 UI budgeted to receiver
  - ✓ Relaxed skew requirements between lanes
  - ✓ Conducive to higher frequencies, future scaling

Signaling Summary - Loss

- **PCI Express loss budget = ~13db**
  - ✓ Transmit voltage > ~400 mV SE** (~800 mV differential peak to peak)
    - – -3.5 dB de-emphasis non-changing bits
  - ✓ Receive sensitivity ≤ ~88 mV SE** (~175 mV differential peak to peak)
- **Generous silicon budgets**
  - ✓ Conducive to future speed increases without changing system interconnect

** SE = single ended

Signaling Summary - Jitter

- PCI Express Jitter budget = 60% UI**
  - ✓PLL [random] jitter 30% UI: not expected to increase as frequencies increases
  - ✓Interconnect jitter ~ 30% UI: random component relatively constant, deterministic component ~linear with frequency
- Balances silicon and interconnect
  - ✓Faster silicon processes allow for faster signaling
    - Silicon budget decreases, interconnect budget increases

** UI = Unit Interval = bit cell time

AC Coupling

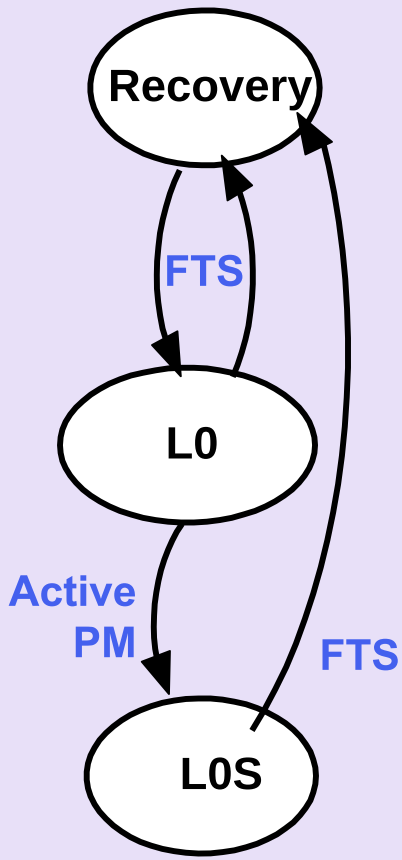
- **Adds flexibility to interface design**
  - ✓ **DC common mode voltage of transmitter can differ from DC common mode voltage of receiver**
    - **Transmitter and receiver can be built by different vendors with different processed**
    - **Simpler to design inter-operability between interfaces**
- **Allows for power saving when link is unused**
- **Allows two different ground references for communicating components**

Encoding/Decoding

- **8b/10b encode on transmit, decode on receiver**
  - ✓ **8 bit characters and control mapped to data (D) and control (K) symbols**
  - ✓ **Maximum run length = 5 (bits)**
  - ✓ **Running disparity to maintain DC balance**
    - – **Above two specs allows for AC coupling**
  - ✓ **Ordered sets of K and D symbols use to construct training and configuration sequences**

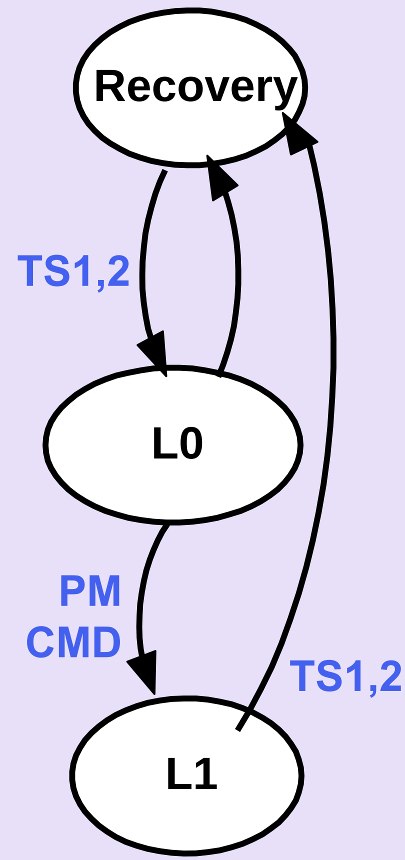
PHY Power States – L0s

- Channel pulled to common-mode voltage, on-chip clocks running, power is on
- Used by active power management
- One-way Fast Training Sequence (FTS) to exit
  - ✓ 16 nsec – 4 usec L0s –> L0 latency
- Optional if distributed reference clock
  - ✓ Not available with local reference clock
- Default = disabled



PHY Power States – L1

- **Channel pulled to common-mode voltage, on-chip clocks may or may not be running, power is on**
- **TX, RX entered via higher layer command**
  - ✓ **Simple handshake between two sides**
  - ✓ **Avoids race condition**
- **TS1/TS2 handshake to exit**
  - ✓ **≤ 64 usec L2 –> L0 latency**
  - ✓ **Implementation dependent**



PHY Power States – L2

- **On-chip, off-chip clocks may be turned off, power may be off**
- **TX, RX enter via higher layer command**
  - ✓ **Simple handshake between two sides**
  - ✓ **Avoids race condition**
- **TS1/TS2 to exit (after power, clocks restored)**
  - ✓ **L2 –> L0 latency not specified (currently)**
  - ✓ **Heavily dependent on PLL spin up, settling time and power supply settling time**
- **Must be able to detect Beacon**
  - ✓ **30 KHz – 500 MHz signal**