June 26, 2007

Hardware Design Considerations for *PCI Express® and SGMII*

AN307

Tiffany Tran-Chandler

Senior Hardware Design Engineer

Power™

freescale™
semiconductor

PCI Express® – Hardware Interconnect Design

- Abstract
  - Reviews the importance of high-speed serial interfaces to the next generation of system interconnect design. Provides the groundwork for successful PCI Express® and Serial Gigabit Media Independent Interface (SGMII) system design, including a focus on the careful attention to PCB design and interconnect that these systems demand.

freescale
™
semiconductor

PCI Express® – Hardware Interconnect Design

► Agenda

- PCI Express® Electrical Architecture Overview
- PCI Express Electrical Fundamentals
    - Loss
    - Jitter
- Interconnect Design
    - Practical PCB Design Recommendations
    - Loss and Jitter Mitigation
    - Simulation
- System Interconnect Example
- Brief introduction to PCI Express 2.0
- SGMII fundamentals and comparison

freescale ™
semiconductor

PCI Express® – Electrical Architecture Overview

► PCI SIG - Several Specs define requirements

- PCI Express® Base Specification 1.1 and 2.0

- Companion Specifications

  - PCI Express Card Electromechanical Specification 1.1 and 2.0 (still unreleased)
    - Addresses Add-in cards for ATX-based desktop applications

  - PCI Express Mini Card Electromechanical Specification Revision 1.1
    - Addresses small card form factor for mobile and communications systems

  - PCI Express ExpressModule™ Electromechanical Specification Revision 1.0
    - Latest addition defines removable modular I/O adapters for closed chassis servers and workstations

  - PCI Express External Cabling Specification 1.0 (Feb 07)

    www.pcisig.com/specifications/pciexpress

freescale
semiconductor

PCI Express® – Electrical Architecture Overview

► PCI SIG – Several Specs define requirements

- PCI Express® Base Specification Revision 1.1
- Companion Specifications
  - PCI Express® Card Electromechanical Specification Revision 1.1
    - Addresses Add-in cards for ATX-based desktop applications
  - PCI Express® Mini Card Electromechanical Specification Revision 1.1
    - Addresses small card form factor for mobile and communications systems
  - PCI Express® ExpressModule™ Electromechanical Specification Revision 1.0
    - Latest addition defines removable modular I/O adapters for closed chassis servers and workstations (previously referred to as "Server I/O Module" or "SIOM" spec)

  http://www.pcisig.com/specifications/pciexpress

freescale™
semiconductor

PCI Express® – Electrical Architecture Overview

► *PCI Express® Electrical Interconnect Design Resources*

- "Checklists" aid electrical interconnect design
  - Motherboard Compliance Checklist
  - Expansion Card Compliance Checklist

► *PCI Express® System Architecture, Ravi Budruk, et.al.*

► *High-Speed Digital Design* and *High Speed Signal Propagation, Howard Johnson, et.al.*

► Xilinx / Howard Johnson Signal Integrity DVD at

www.xilinx.com/products/design_resources/signal_integrity/resource/hojo_dvd.htm

freescale
semiconductor ™

PCI Express® – Interconnect Topologies

► Two devices on a single system board

► One device on system board, second device on add-in card with a connector in between
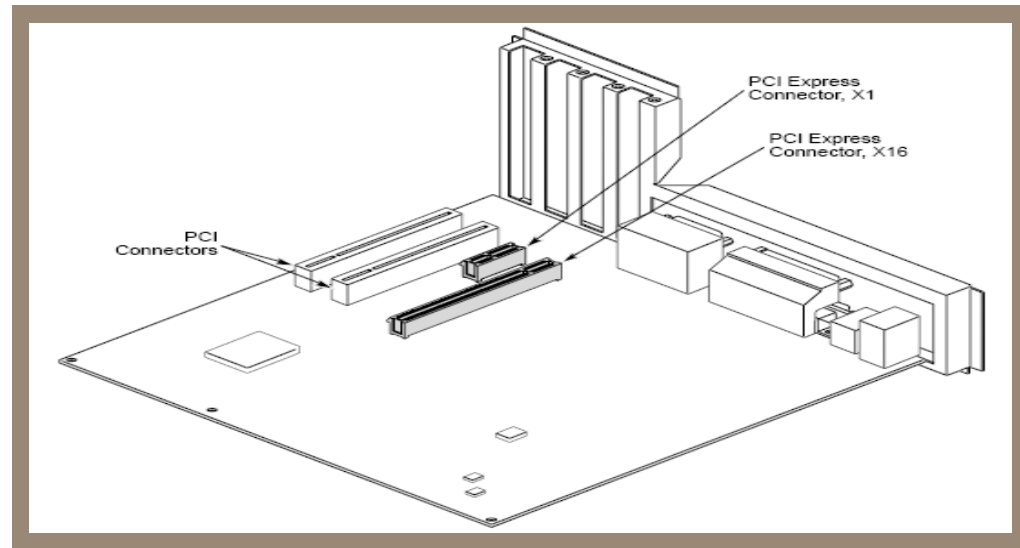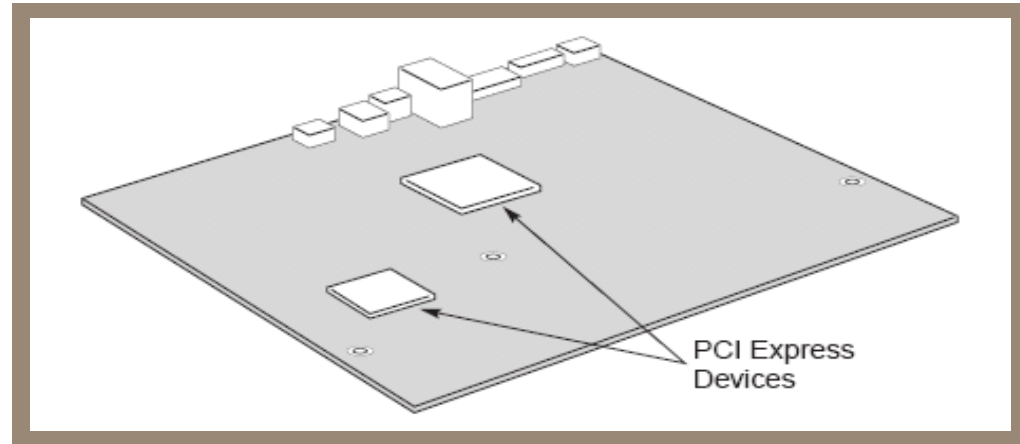
PCI Express® – Interconnect Topologies

▶ One device on system board, second device on add-in card with a riser card and two connectors in between

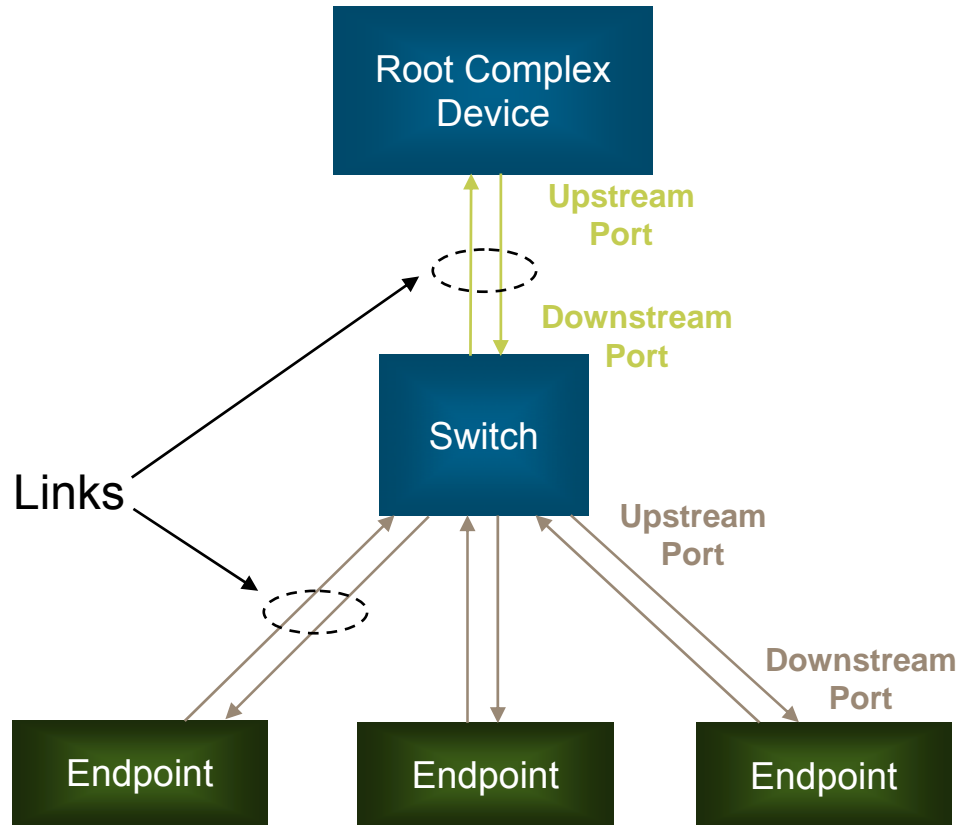▶ Two ExpressModules – each with a device. The ExpressModules™ plug into a small, compact chassis backplane

freescale
semiconductor
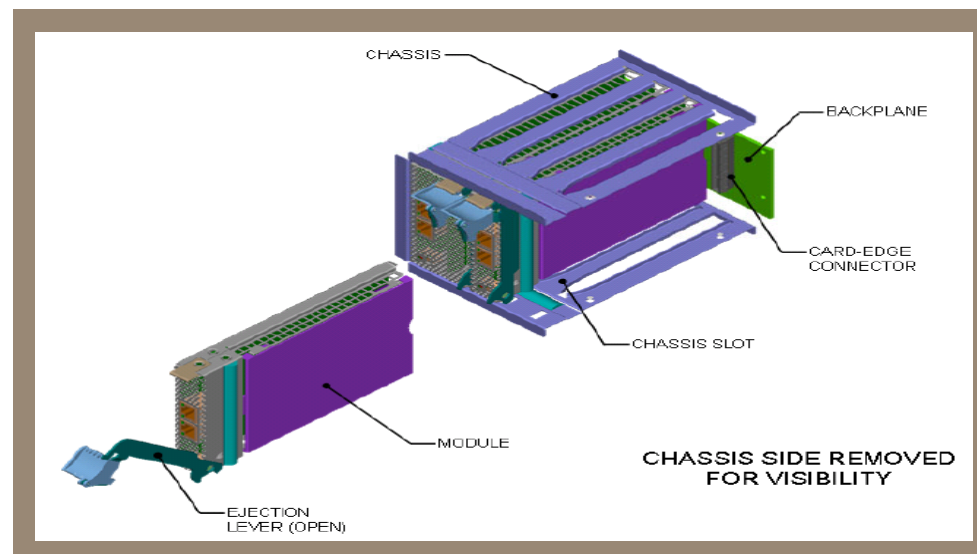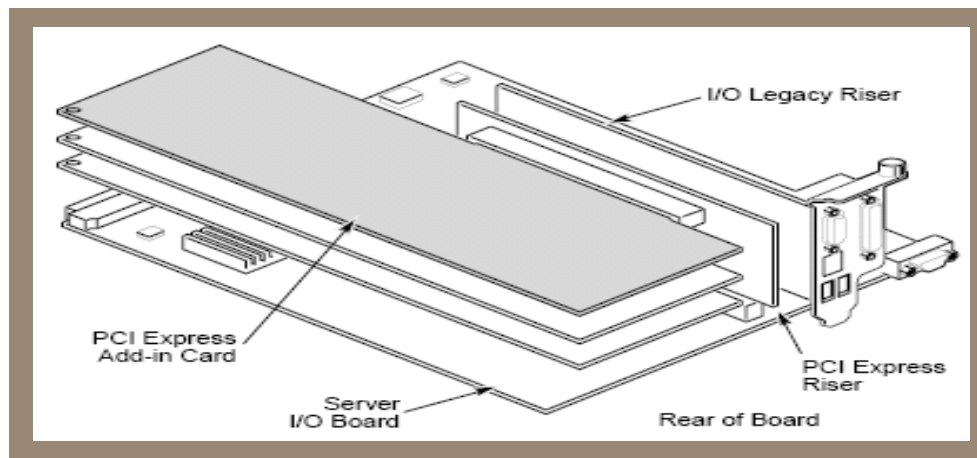
PCI Express® – Electrical Architecture Overview

► *PCI Express*® Key Electrical Features

- High speed signaling extension to PCI and PCI-X
  - 1.x: 2.5 Gb/s raw bit rate per lane (diff pair) / per direction
  - 2.0: 5.0 Gb/s raw bit rate per lane (diff pair) / per direction
- Serial Interface on a dual simplex bus
- Point-to-point connections
- Differential (*LVDS*), AC coupled signaling
- Terminations built into devices
- Embedded clock in data stream (8b/10b Encoding)
- In-band sidebands (interrupts, resets …)

freescale ™
semiconductor

PCI Express® – System Topology

Root Complex Device

Upstream Port

Downstream Port

Switch

Upstream Port

Downstream Port

Links

Endpoint

Endpoint

Endpoint

freescale ™
semiconductor

PCI Express® – Link Topology

► Number of Lanes in a Link can be x1 x2 x4 x8 x12 x16 or x32

► Each Lane consists of an upstream and downstream channel

► Each Channel consists of one differential pair of signals

freescale™
semiconductor
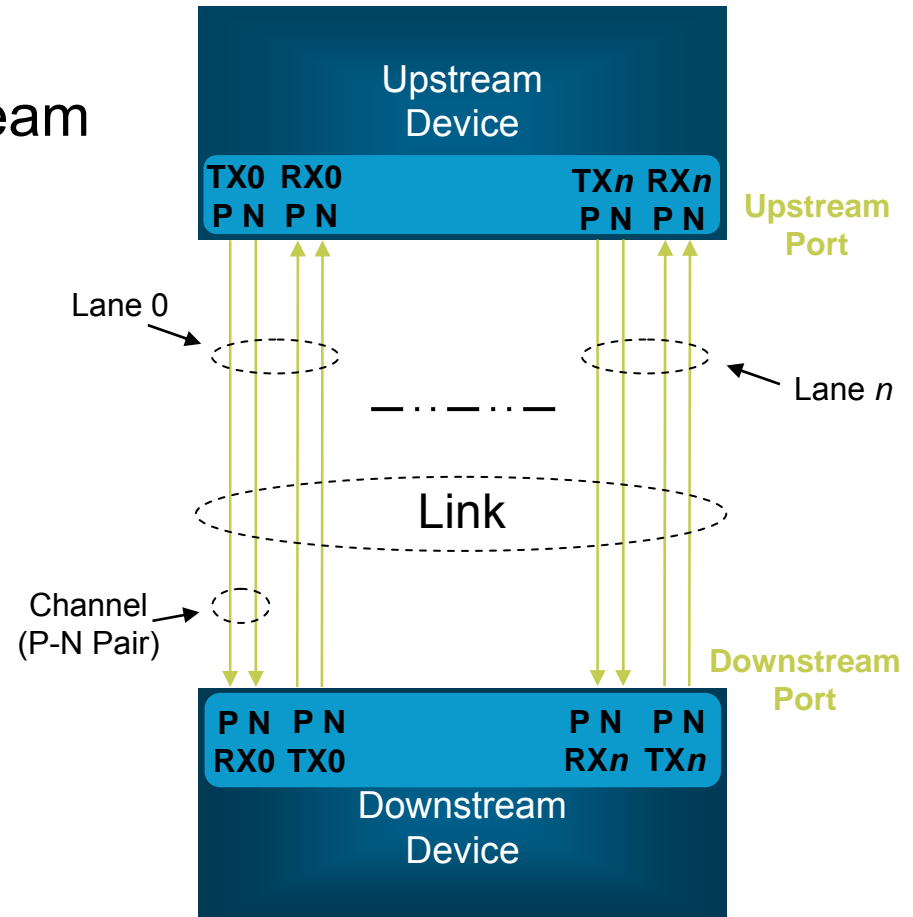
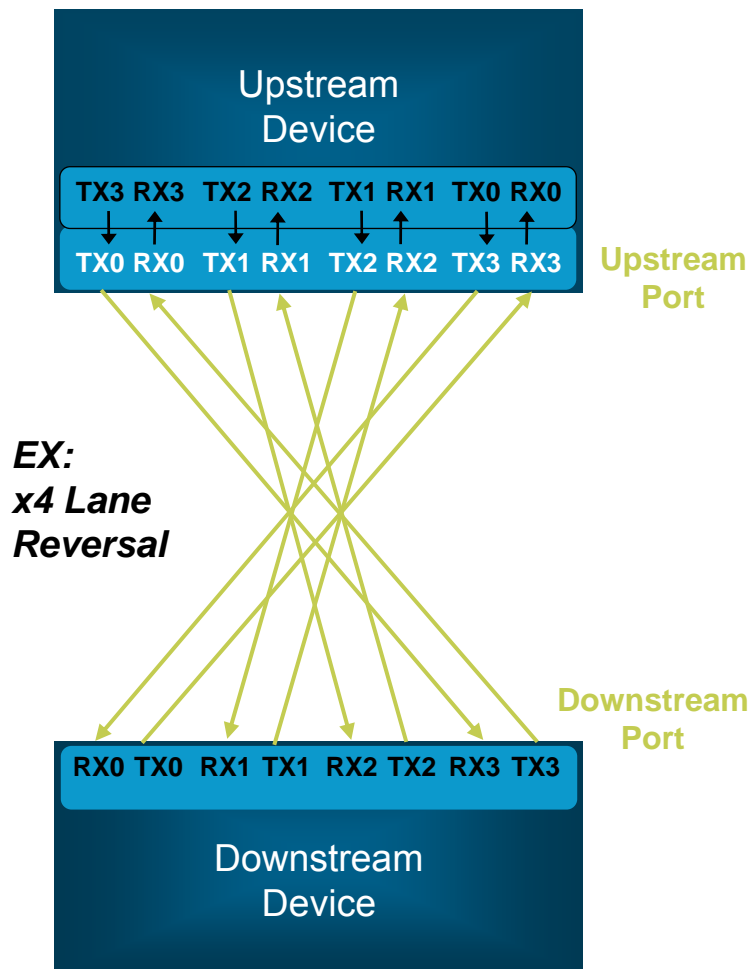# PCI Express® – Link Connectivity Variations

► Polarity Inversion

- Any channel can be connected with the positive and negative signals reversed
- Training sequence detects polarity inversion condition
- RX port is **required** to invert its signal polarity
- Eases routing if differential signals are normally bow-tied
- Reduces vias

Example: Upstream channel 0 polarity reversed
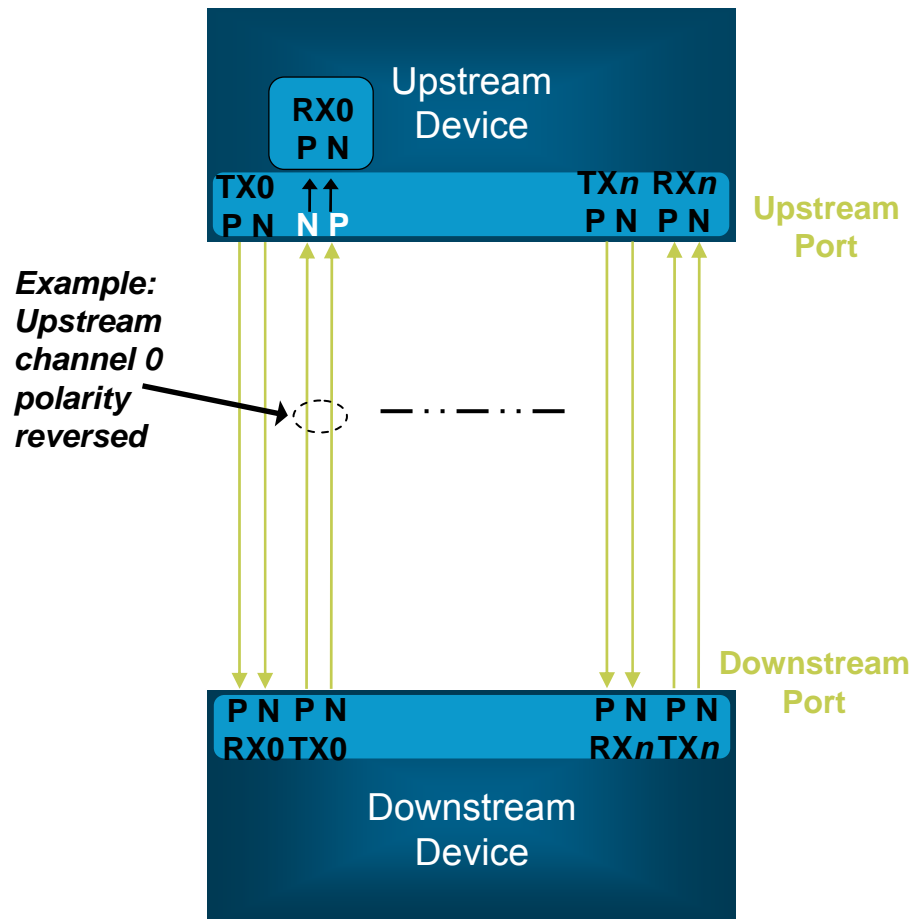
PCI Express® – Link Connectivity Variations

► Lane Reversal

- Either <u>all</u> lanes or <u>no</u> lanes in a link are reversed – no subset of reordering allowed
- Training sequence detects lane reversal condition
- Either device may change the lane ordering
- Eases routing, reduces vias, if lanes are naturally bow-tied
- **Optional feature** (may not be supported in some devices!)

Upstream Device

TX3 RX3 TX2 RX2 TX1 RX1 TX0 RX0

TX0 RX0 TX1 RX1 TX2 RX2 TX3 RX3   **Upstream Port**

EX: x4 Lane Reversal

Downstream Port

RX0 TX0 RX1 TX1 RX2 TX2 RX3 TX3

Downstream Device

freescale ™
semiconductor

▶ Differential Signaling
- Lower Crosstalk
- Improved EMI
- Matched signal return path
- Z_{diff} = 100 ohms +/- 20%, Z_0 = 60 ohms +/- 15% recommended

▶ AC Coupled Differential Pairs
- Isolates DC voltage component from TX device to RX device
  - Allows different reference planes at each device on a link
- Used to employ link receiver detection during link initialization
- Hot insertion protection
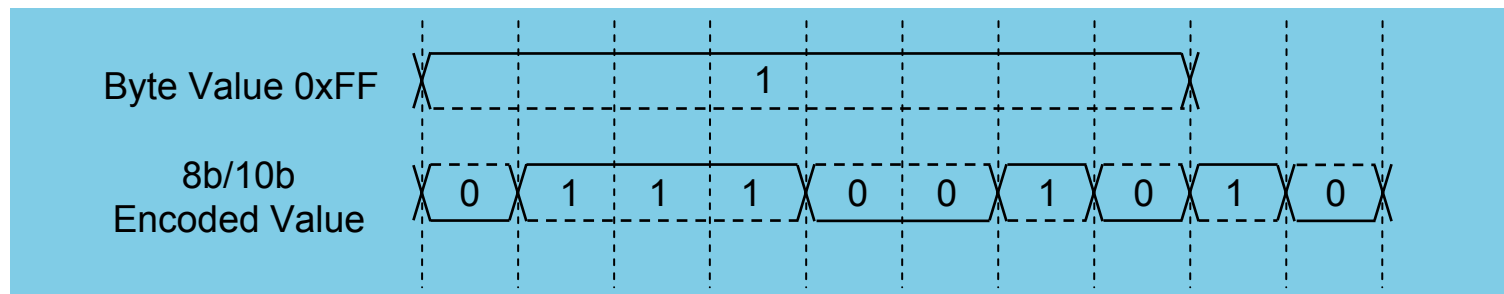- 75nF – 200nF, placed near device transmitter outputs

▶ Terminations on-die to 100 ohms differential

freescale ™
semiconductor

▶ Data scrambling (optional)
- Limits repeated occurrences of specific data patterns
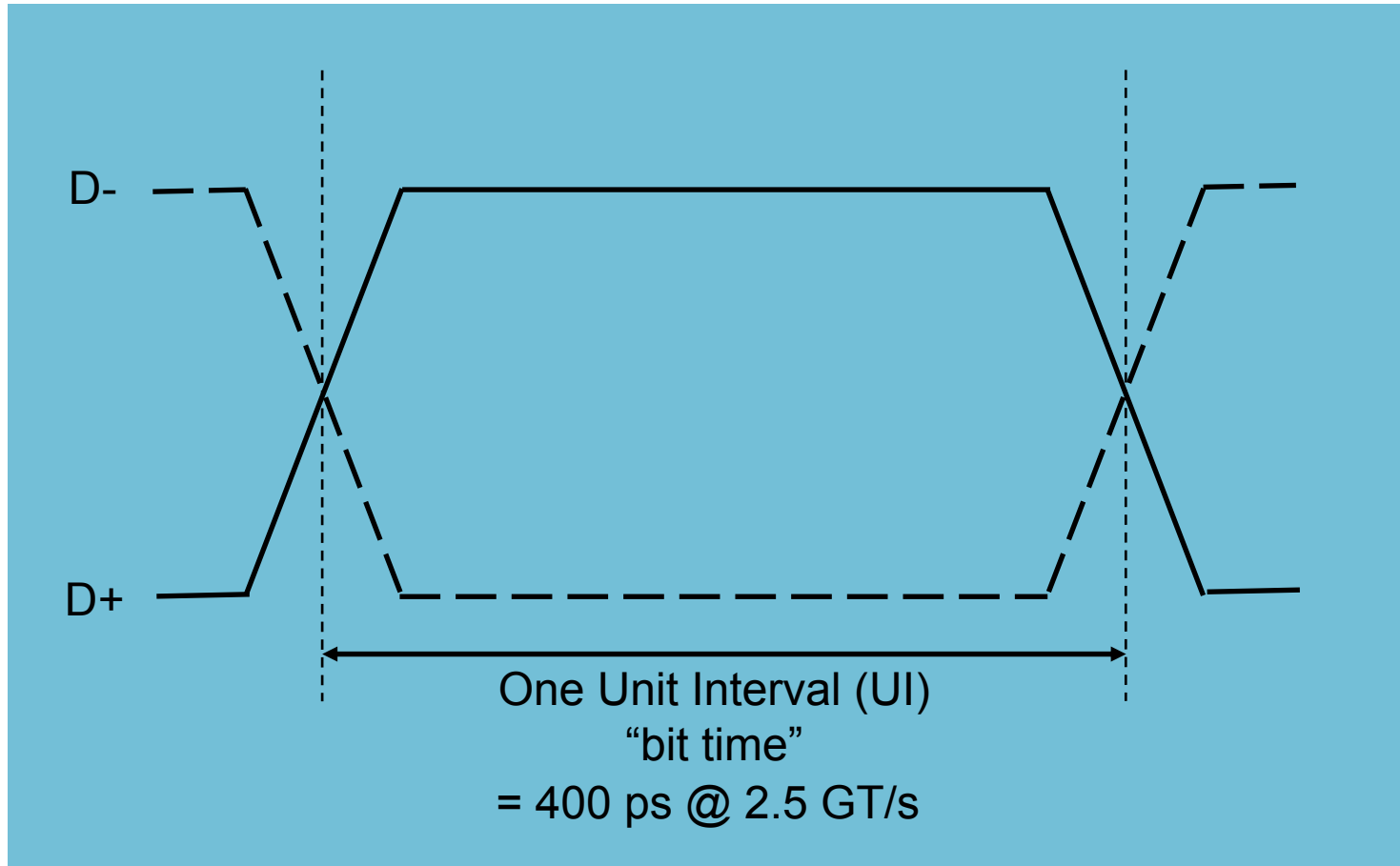- Significantly improves EMI radiation

▶ 8b/10b data encoding (required)
- No more than five bits in a row are transmitted at the same state – limits "dead time" and allows AC coupling scheme
- Frequency range is bounded from 250 MHz to 1.25 GHz
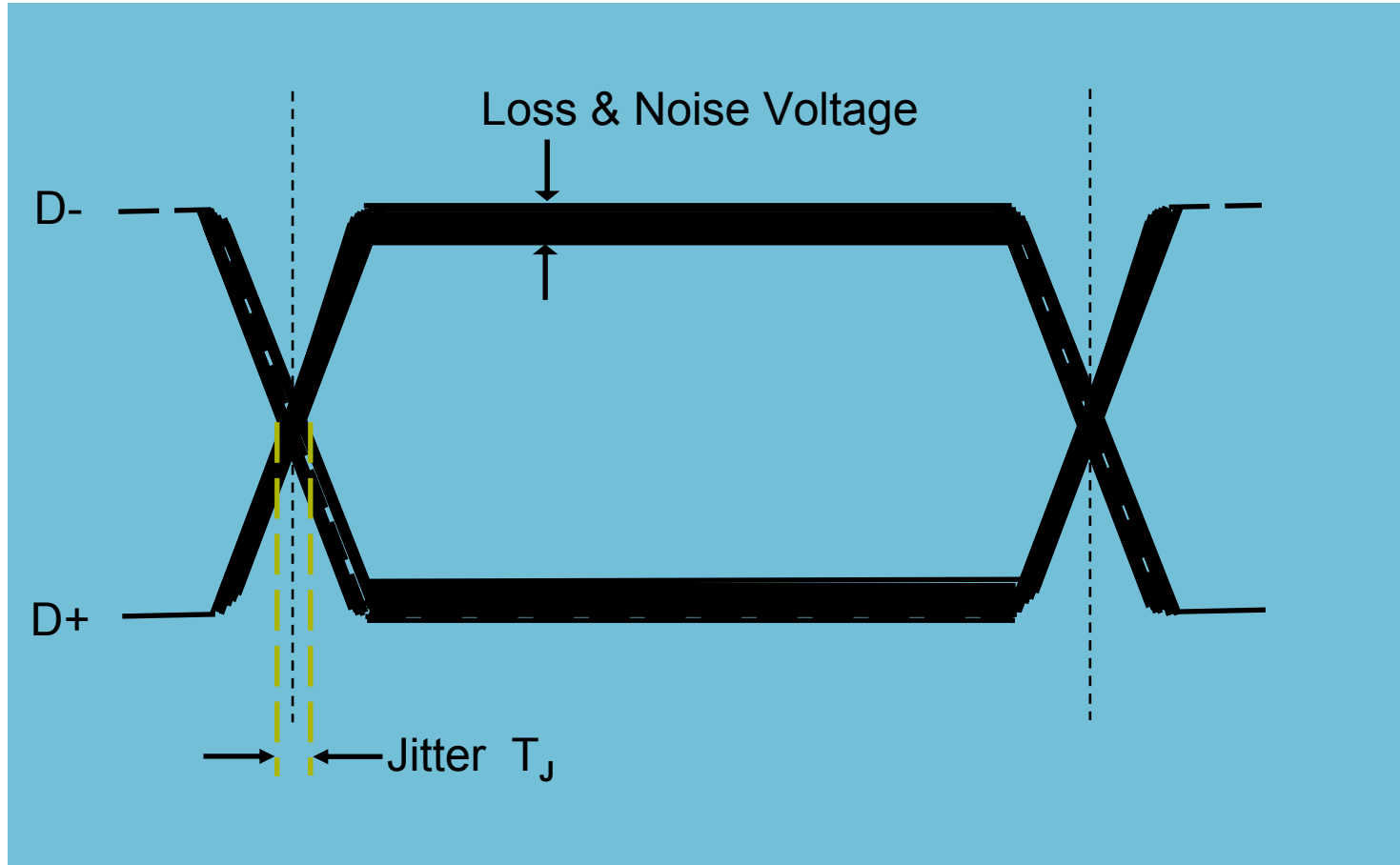- Guarantees clock recovery from transitioning data

Byte Value 0xFF	1
8b/10b Encoded Value	0 1 1 1 0 0 1 0 1 0

freescale ™
semiconductor

PCI Express® – Clocking

► TX/RX Clock is embedded in data
- Clock is reconstructed from data signal
- Removes skew relationships between a separate clock and data path

► Reference clock
- Provided by system to each device on a link
- HCSL(High Speed Current Sensing Logic) (0mV to 700mV level)
  - Static Reference Clock
    – 100 MHz +/- 300 ppm
  - Spread Spectrum Clocking (SSC)
    – 100 MHz +/- 300 ppm with modulation frequency of 30 to 33 kHz, down spread 0% to –0.5%
    – Both ports must adhere to +/- 300 ppm requirement, so SSC usually requires the clock for both devices on a link come from same SSC clock source

freescale™
semiconductor

PCI Express® – SERDES Data Eye

D-

D+

One Unit Interval (UI)
"bit time"
= 400 ps @ 2.5 GT/s

freescale™
semiconductor

PCI Express® – SERDES Data Eye

Loss & Noise Voltage

D-

D+

Jitter T_J

freescale ™
semiconductor
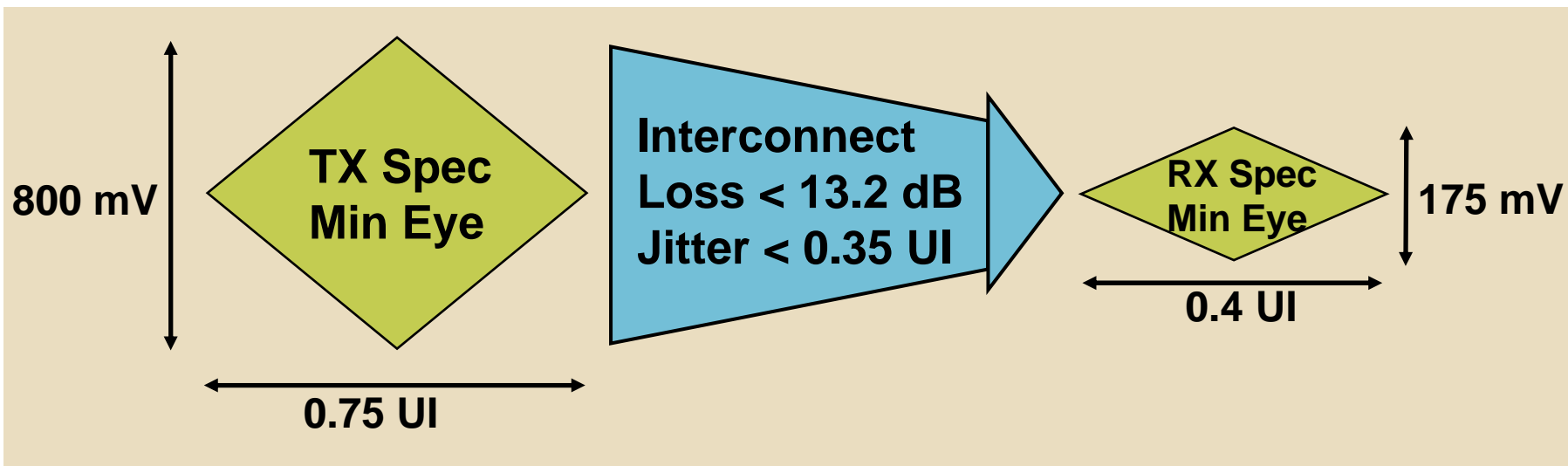
PCI Express® – Key Electrical Parameters

►Loss
- Insertion Loss
- Return Loss

►Jitter
- Random Jitter (Rj)
- Deterministic Jitter (Dj)

800 mV

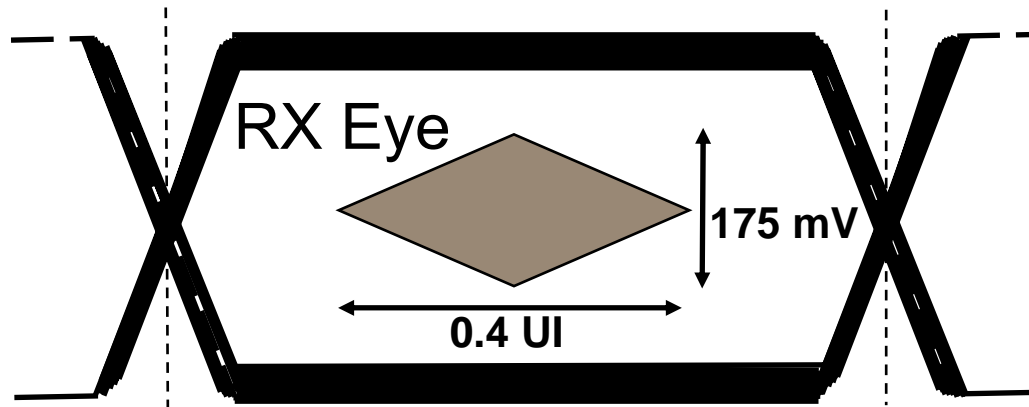**TX Spec Min Eye**

0.75 UI

Interconnect Loss < 13.2 dB Jitter < 0.35 UI

RX Spec Min Eye

175 mV

0.4 UI

- 1 UI = 400 ps
- TX Eye shown without de-emphasis

freescale ™
semiconductor

TX Eye

800 mV

0.75 UI

Eye Diagram at the Transmitter

freescale ™
semiconductor

PCI Express® – SERDES Data Eye

RX Eye

175 mV

0.4 UI
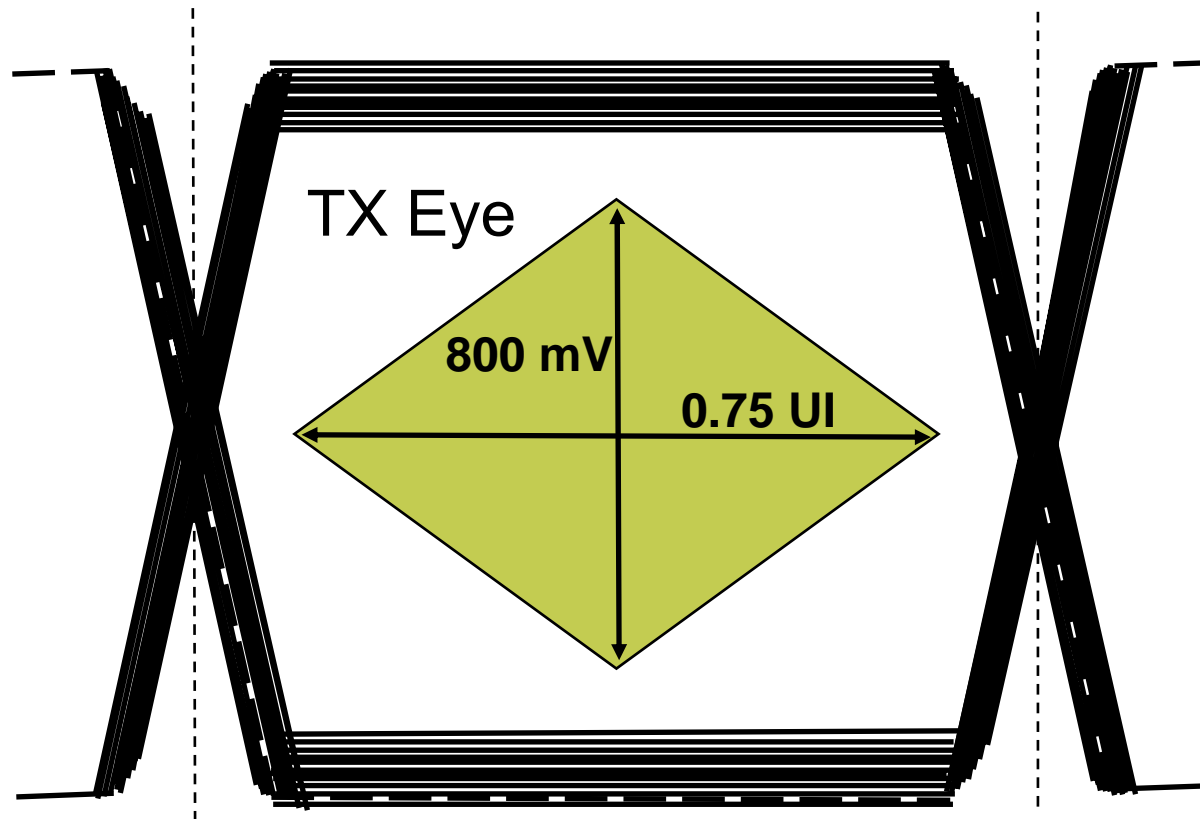
Eye Diagram at the Receiver

freescale ™
semiconductor

System Issues vs. Frequency
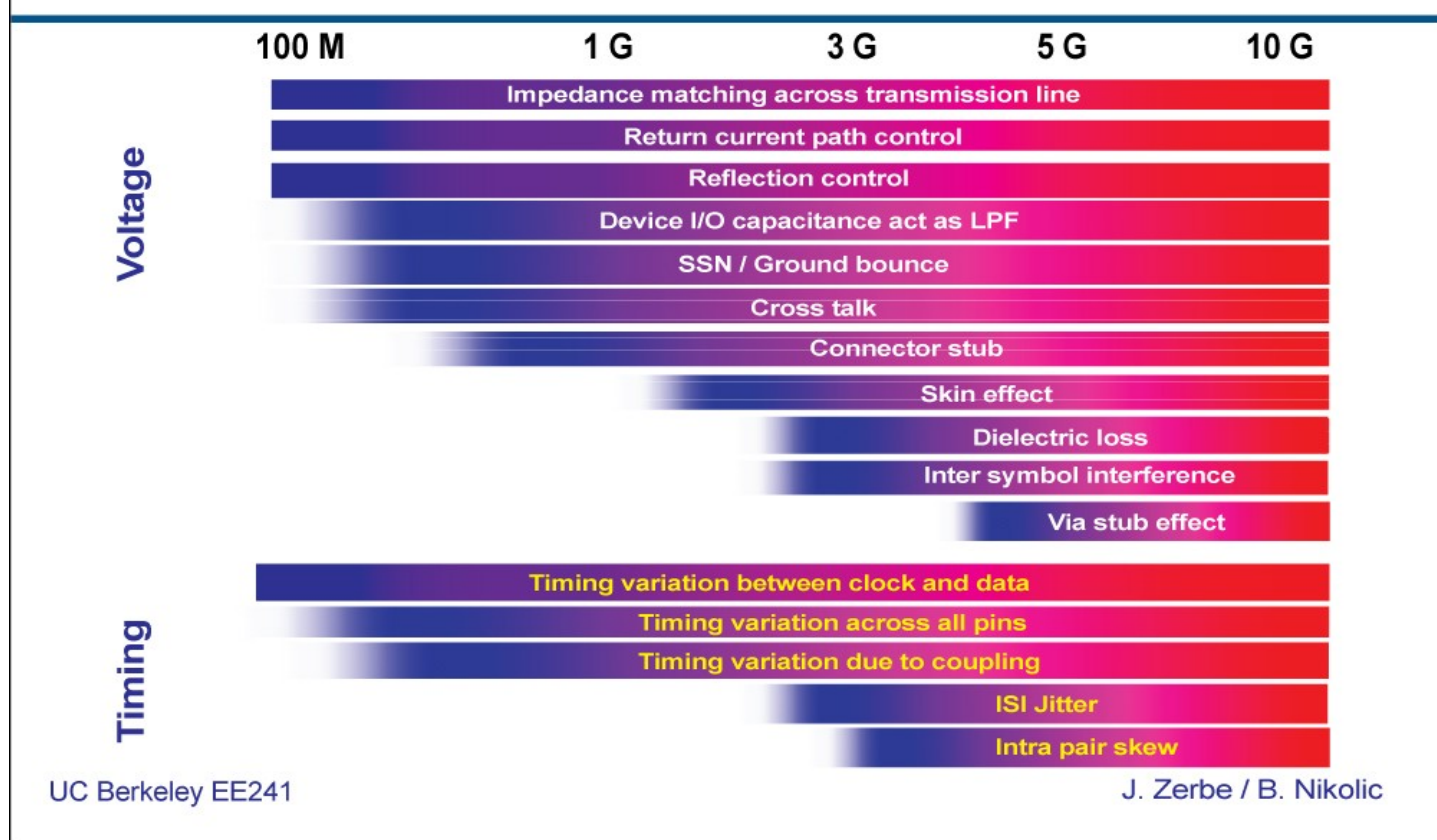
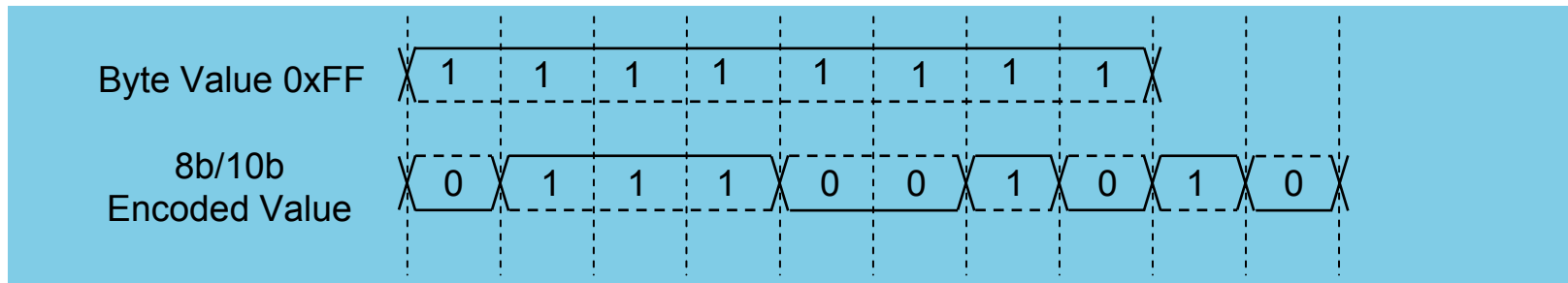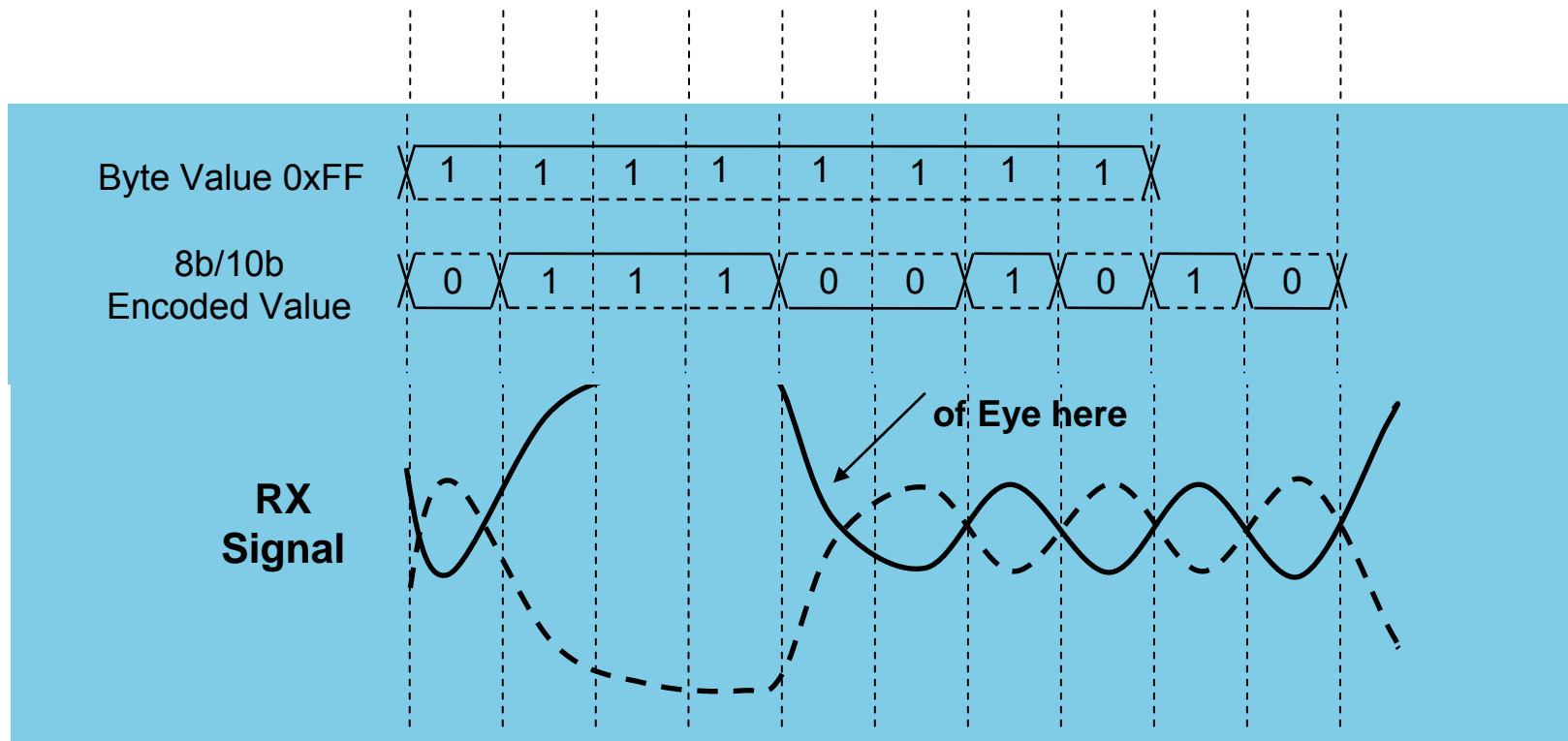► **ISI (Inter-Symbol Interference)**

- The signal degradation effect on a signal state, called a symbol, from *previous* data signal transitions

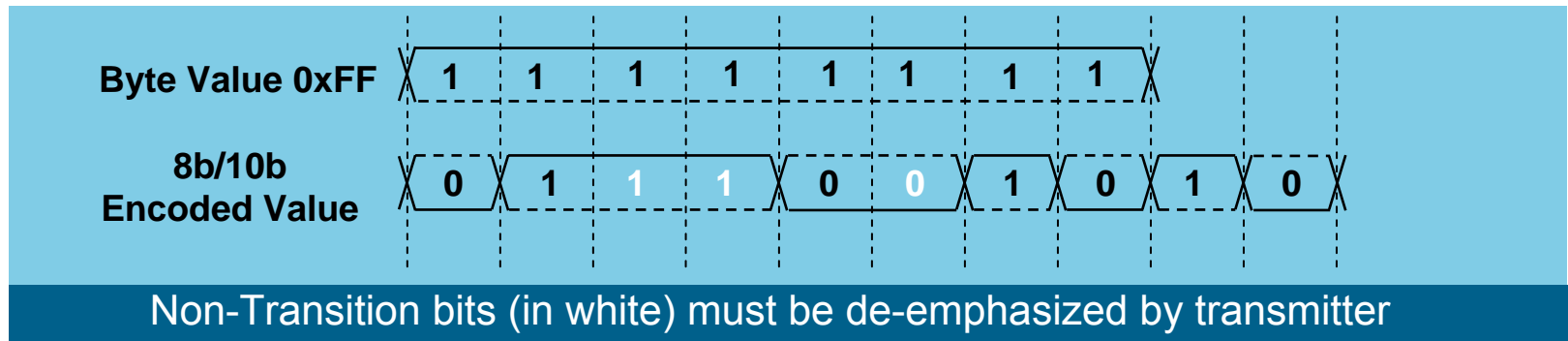Byte Value 0xFF: 1 1 1 1 1 1 1 1

8b/10b Encoded Value: 0 1 1 1 0 0 1 0 1 0

freescale ™
semiconductor

PCI Express® – Loss

► Signal transitions start at different voltage levels depending on their previous signal state
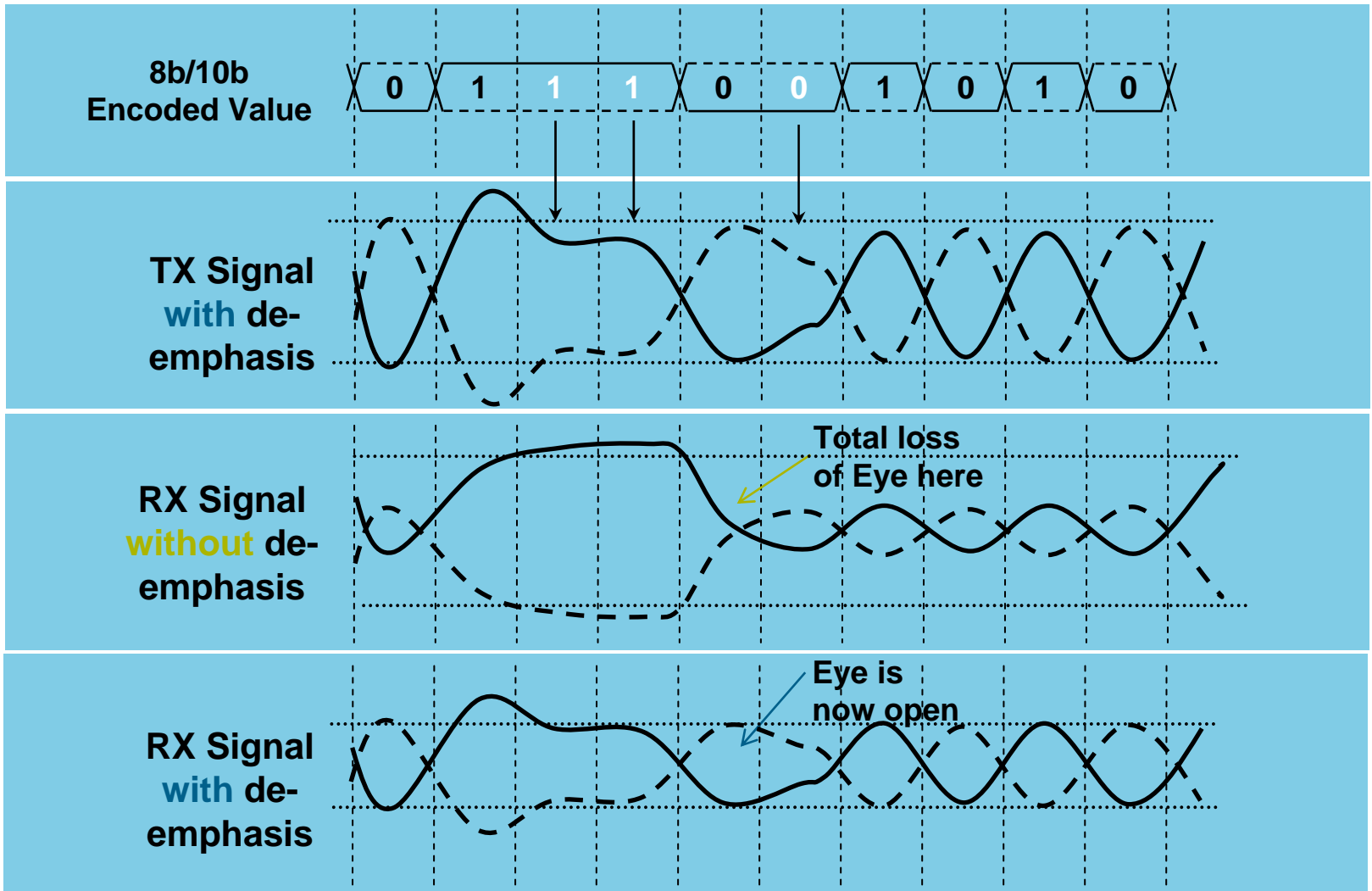
freescale ™
semiconductor

► De-emphasis

- An attenuation of the voltage level at the transmitter of all consecutive bits of the same signal state except the first
- Non-transition bits are de-emphasized by the transmitter (required by spec)
- The cure for ISI

Byte Value 0xFF	1	1	1	1	1	1	1	1

8b/10b Encoded Value	0	1	1	1	0	0	1	0	1	0

Non-Transition bits (in white) must be de-emphasized by transmitter

freescale™
semiconductor

PCI Express® – SERDES Data Eye
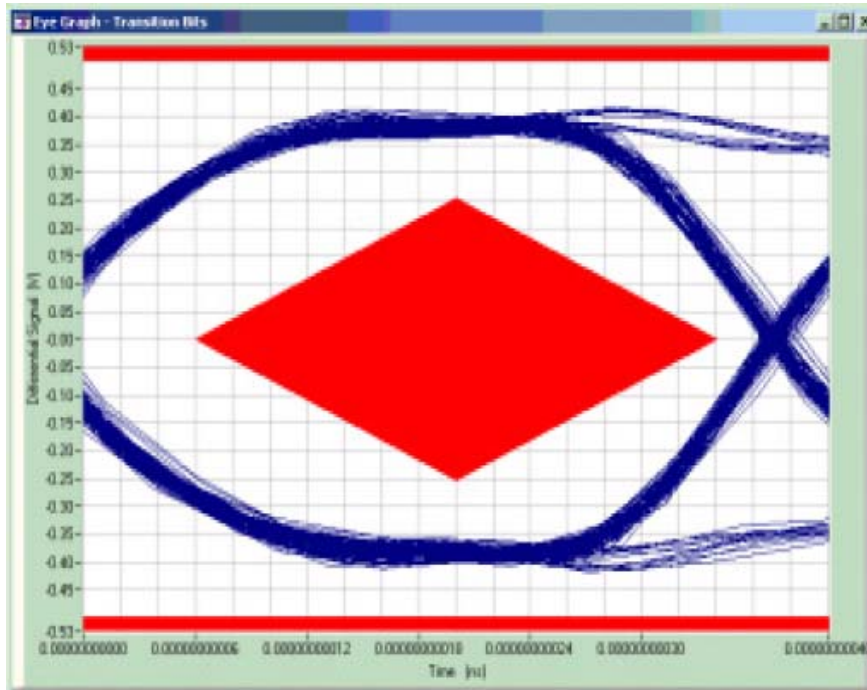
Transition Bit Eye Diagram at the Transmitter

Non-Transition Bit Eye Diagram at the Transmitter (with De-Emphasis)

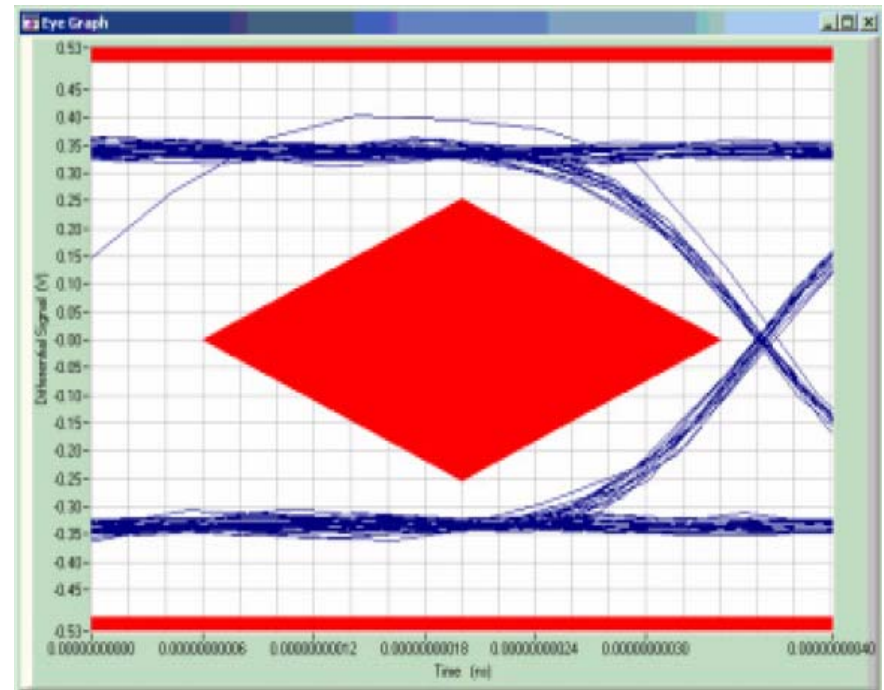PCI Express® Design and Interoperability Considerations
Ajay Bhatt 2003

PCI Express® – Key Electrical Parameters

800 mV

Transition Bit TX Eye (1.25 GHz)

0.75 UI

Interconnect Loss < 13.2 dB Jitter < 0.35 UI

RX Spec Min Eye

175 mV

0.4 UI

De-emphasis

505 mV

Non-Transition Bit TX Eye (625 MHz)

0.75 UI

Interconnect Loss < 9.2 dB Jitter < 0.35 UI

RX Spec Min Eye

175 mV

0.4 UI

freescale ™
semiconductor
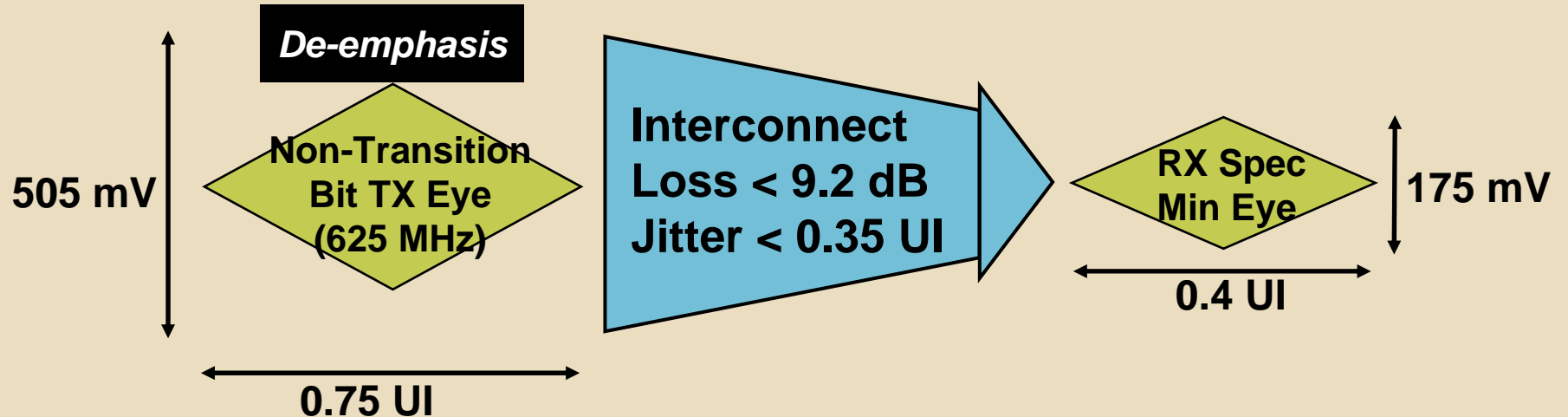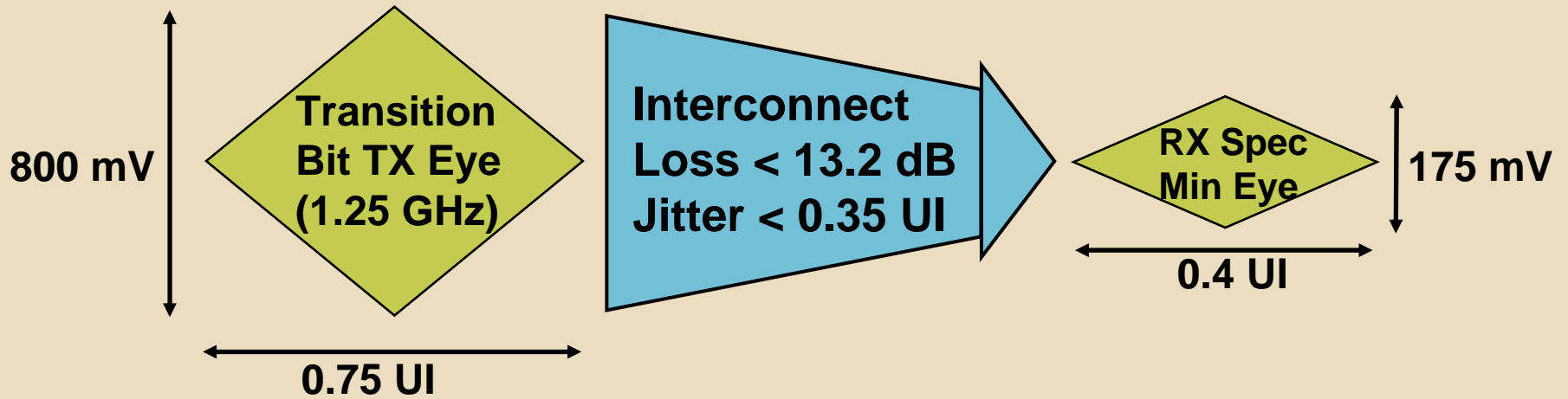
► Insertion Loss

- Signal amplitude attenuation in the forward signal direction (from TX to RX)

- Contributors to loss

  - KEY ➔ consistent L_0/C_0 along the path
  - Package
  - PCB traces and vias
  - AC coupling caps (parasitics)
  - Connectors
  - PCB dielectric variations (FR4 fabric weave)
  - Data Pattern (Intersymbol Interference)

$$Z_0 = \sqrt{\frac{L_0}{C_0}}$$

Ideal (lossless) transmission line

freescale™
semiconductor

PCI Express® – Loss
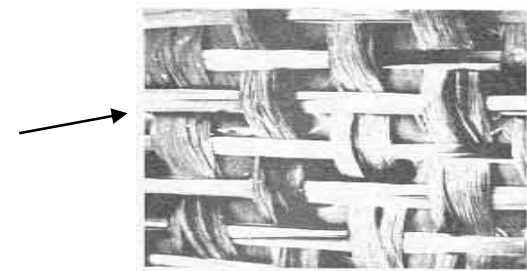
► Insertion Loss

- Conservative estimates (standard FR4)
  - PCB traces ➔ 0.25 — 0.35 db/inch at 1.25 GHz
  - Standard FR4 Vias (T < 150 mils) Vias ➔ 0.25 db each
  - Backplane Vias (T > 150 mils) ➔ 0.5 db each
  - Connectors (CEM Spec only allows 1db loss up to 1.25 GHz
- Two types of Loss to consider
  - Skin Effect loss
    – Resistive loss due to the tendency of current to flow near the surface of a conductor at high frequencies
  - Dielectric Loss
    – Loss due to heating effects in the surrounding dielectric materials
  - $Loss_{Total} = Loss_{Skin}\ Effect + Loss_{Dielectic}$

Both are frequency dependent

freescale ™
semiconductor

▶ **Skin effect loss**
- grows in proportion to the square root of frequency
- grows in proportion to decreasing trace width

$$\alpha \propto \sqrt{f} / t_w$$

t_w = **trace width**

▶ **Dielectric loss**
- grows in proportion to the frequency
- grows in proportion to loss tangent of material
- grows in proportion to the square root of dielectric constant

$$\alpha = 2.3 \, (f) \, \tan(\theta) \sqrt{e_r}$$

$\tan(\theta)$ = **loss tangent**
e_r = **dielectric constant**

$$Loss_{Total} = Loss_{Skin\ Effect} + Loss_{Dielectric}$$

freescale ™
semiconductor

►Return Loss

- Reflected signal amplitude in the reverse direction (in path from RX to TX)
- Matching Impedance along the path minimizes return loss

System Card   **Add-in Card**

AC Caps   Connector

TX   RX

freescale ™
semiconductor
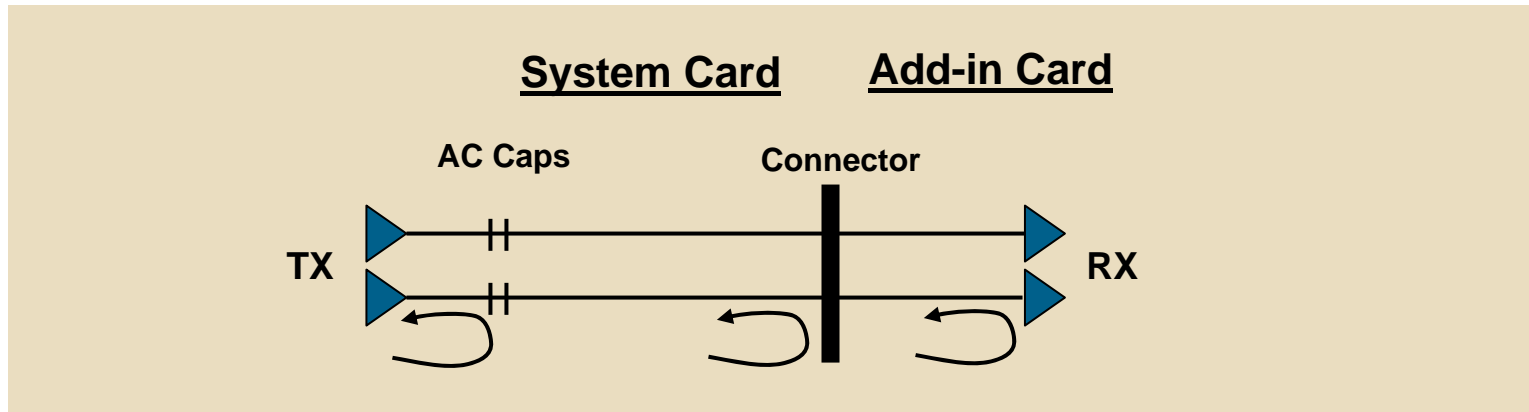
► Jitter

- Random Jitter (Rj)
  - Random noise and thermal effects in silicon
  - Gaussian distribution (unbounded)
- Deterministic Jitter (Dj)
  - Data dependent (ISI)
  - System and transmission line effects
    - ISI
    - Crosstalk
    - Impedance discontinuities
    - Power supply noise
    - Non-Gaussian distribution (bounded)

freescale
semiconductor ™

►Jitter

- Total Jitter (Tj) measured directly from Eye Diagram

- Spec allows Tj of 0.35 UI on the interconnect ➔ 140 ps

- Recently released CEM spec breaks out phase jitter budget

- Refer to *PCI Express® Jitter and BER* and *PCI Express Jitter Modeling* at PCI-SIG website for additional details on jitter definition and budgeting

freescale
semiconductor ™

PCI Express® – PCB Design Recommendations

► Use standard FR4

► Z_{diff} = 100 ohms +/- 20%, Z_o = 60 ohms +/- 15%

► Routing Length

- System Board ➔ up to 12 inches
- Add-in Card ➔ up to 3.5 inches
- System Board (chip-to-chip) ➔ up to 15.5 inches
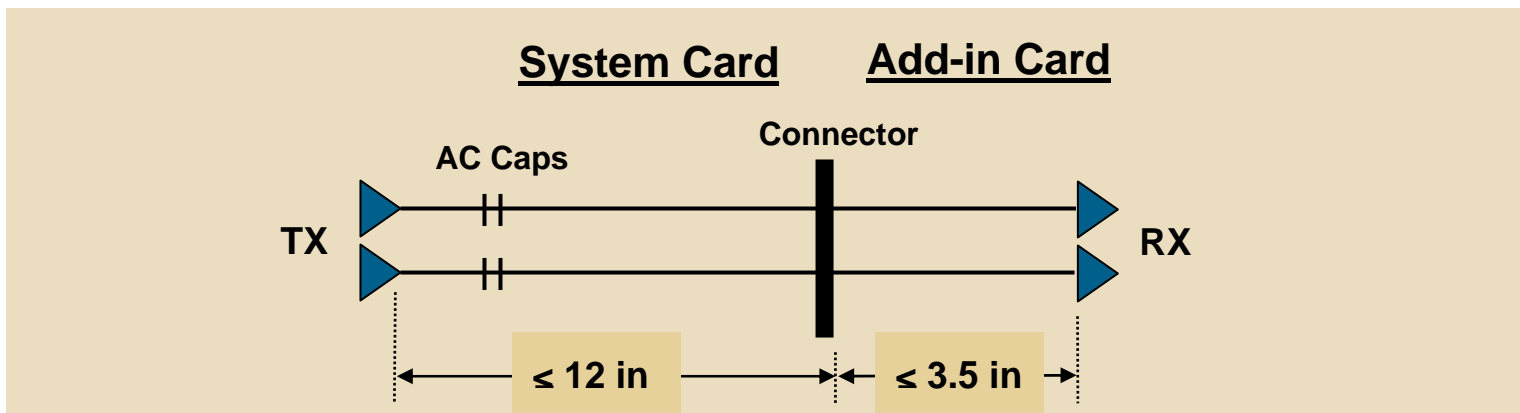
PCI Express® – PCB Design Recommendations

► PCB Trace routing

- Trace thickness
  - ➔ Microstrip (outside signal layers) ½ oz. Copper plated - **OR** -
  - ➔ Stripline (inner signal layers) 1 oz. Copper

- Trace width = 5 mils

- Solid reference plane GND (strongly recommended)

- Close differential intra-pair air gap spacing
  - ➔ Microstrip = 7 mils
  - ➔ Stripline = 5 mils

freescale ™
semiconductor
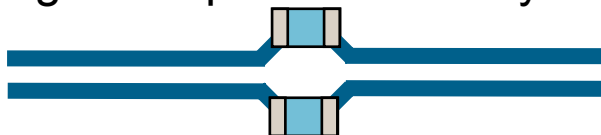
PCI Express®– PCB Design Recommendations (cont.)

- Differential pair-to-pair spacing should be > 20 mils ➔ minimizes crosstalk
- Length Matching between pairs not required
    - ➔ Recommend matching to within 3 nsec to reduce latency
        - ➔ ~20 inches on a microstrip
        - ➔ ~16 inches on a stripline
- Match diff pair (P to N) to within 5 mils

freescale™
semiconductor

PCI Express®– PCB Design Recommendations (cont.)

► PCB Trace routing

- Maintain symmetry of diff pair routing
  - Placing AC capacitors side-by-side symmetrically will help this

- Side-by-side breakout from package pins

- Serpentine or loop-end route where non-symmetrical breakouts patterns cannot be avoided

freescale™
semiconductor

PCI Express®– PCB Design Recommendations (cont.)

▶ PCB Trace routing
- Bends
  - Avoid tight bends in trace: 45° or less, no 90° bends
  - Match number of left and right bends when possible
- No Stubs!
- Do not route diff pairs over voids in reference plane – includes antipads
- Maximum of 6 vias in entire path
- Stitching vias (to GND) at all diff pair via sites
- Vertical and horizontal routing increases dielectric loss

freescale ™
semiconductor
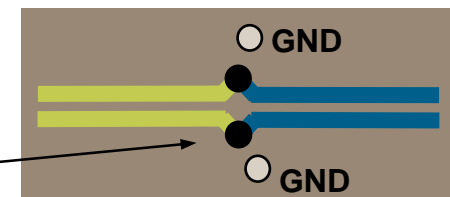
PCI Express® – PCB Design Recommendations (cont.)

▶ Reference Clock

- Net lengths do not have to be equal
- May need to filter Reference Clock VDD
- Reference differential pair to solid plane
- Recommend using common clock source if SSC is needed
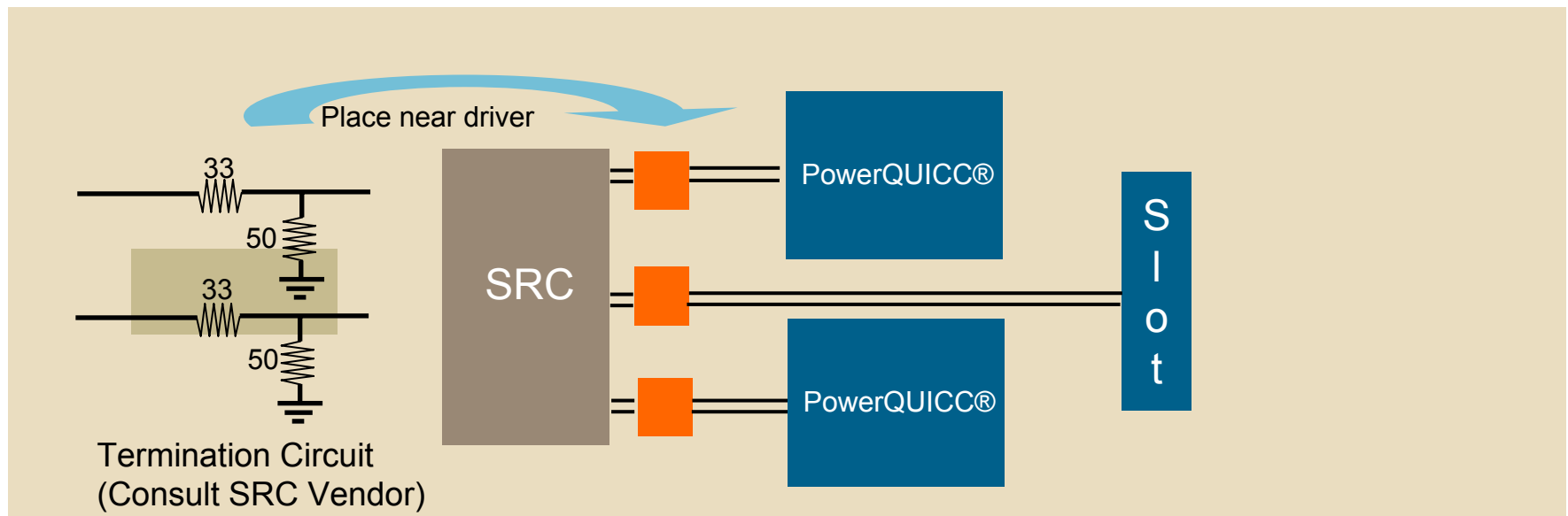
PCI Express® – HPCN EVAL Platform Routing

PCIe Bus Routing for HPCN (MPC8641)
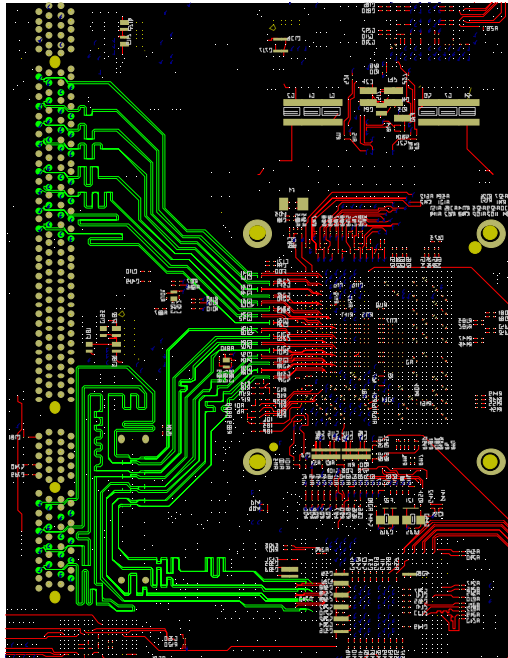
PCIe
Routed on
Bottom Layer

PCIe
Routed on
Top Layer Layer

PCI Express® – HPCN EVAL Platform routing

PCIe Bus Routing for HPCN (MPC8641)

PCIe Routed on Top Layer

Logic Analyzer Mid-bus Probe

freescale™
semiconductor
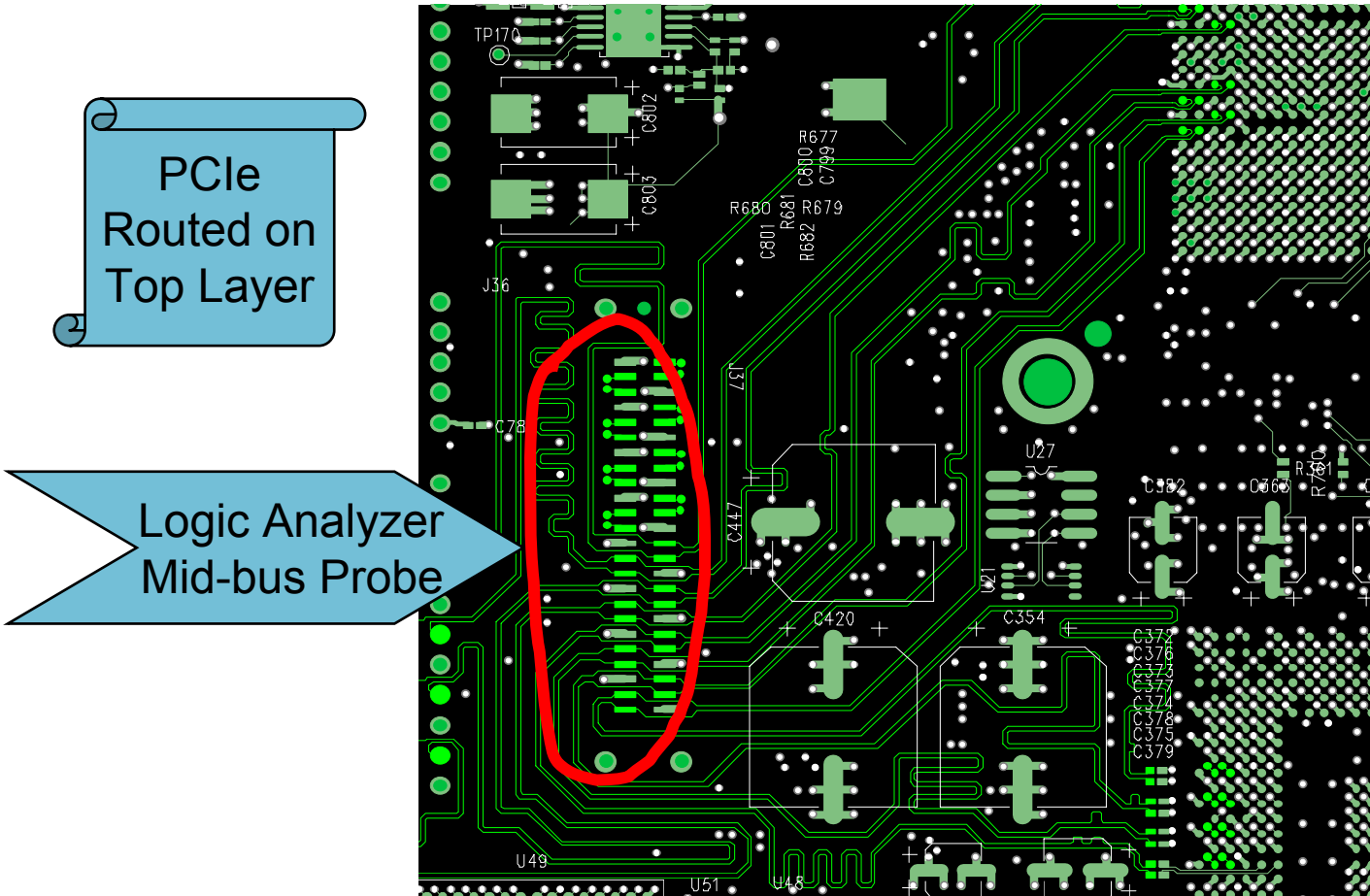
PCI Express® – Ways to Mitigate Loss and Jitter

► **Trace width**

- Use wider traces
  - (+) Improves skin effect loss
  - (+) No increase in material cost
  - (-) Uses more routing area
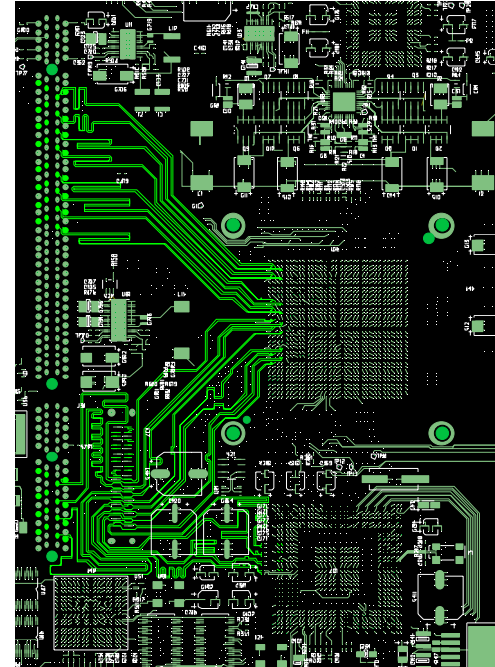 - (-) Increases PCB thickness to maintain impedance targets
  - Ex: ½ oz. Copper stripline, Zdiff = 100 ohms
    - 5 mil width / 6 mil air gap yields 1 db loss per 10.8 inches
    - 8 mil width / 14 mil air gap yields 1 db loss per 15.0 inches

► **Vias**

- Use blind/buried vias
- Use via-in-pad
  - (+) Smaller via geometry ➔ lower parasitics
- Backdrill vias with stubs ➔ eliminates signal reflections at via stub

size scales linearly with frequency (2x smaller ==> 2x frequency)

freescale ™
semiconductor

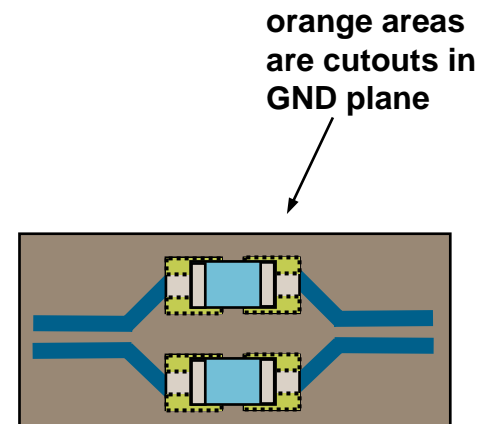# PCI Express® – Ways to Mitigate Loss and Jitter

►PCB Materials

- Use "High-speed" FR4
  - (+) Lower e_r ➔ lowers dielectric loss
  - (+) Lower loss tangent ➔ lowers dielectric loss
    - Loss tangent can be cut in half with modified FR4 materials
  - (-) can cost 2x to 5x of standard FR4
- Use "smooth" copper
  - (+) Lower dielectric loss
  - (-) Caution! Peel strength is reduced

►AC Coupling Caps

- Use smaller body style (0402 recommended for most designs)
  - (+) 0201 pads reduce parasitic capacitance by 70% over 0402 pads
  - (-) 0201 assembly and rework more challenging
- Cutout Reference planes underneath capacitor pads

orange areas are cutouts in GND plane

freescale™ semiconductor

► Simulate the design!

- Preliminary simulation

  - As soon as you know the topology structure (trace length estimate, number of connectors, etc)

  - Goal: Will this topology work?

- Second pass simulation

  - Done when details of the elements are known (using trace lengths and width, actual connector model, AC capacitor geometry, number of vias, etc.)

  - Improves confidence of success

freescale ™
semiconductor

PCI Express® – Simulation Example

System Card

Riser Card

Add-in Card

TX

AC Caps
0.1 uF
0201

Connector

PCI Express
midbus
probe

SMT
Connector

0.1"
via
stubs

RX

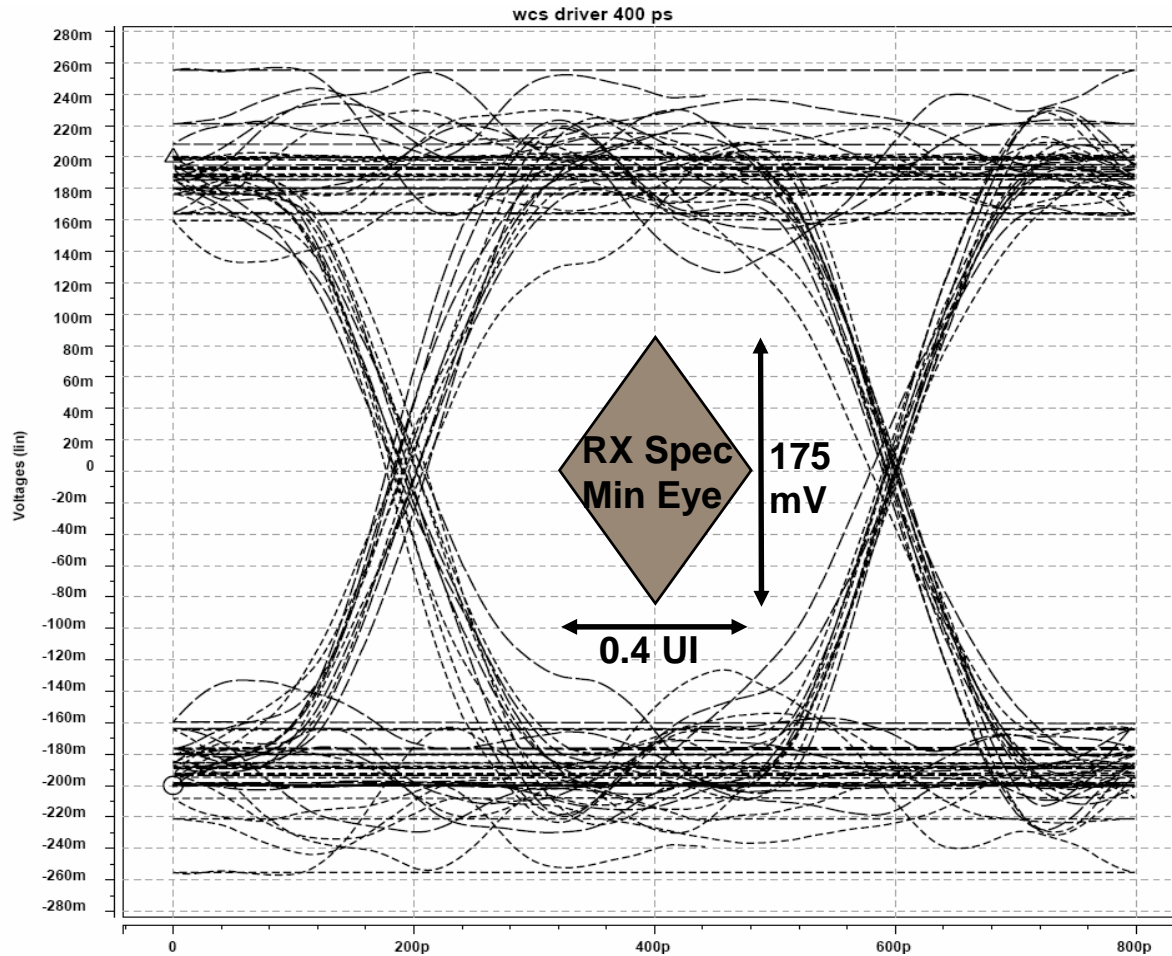| Length → | 0.7" | 6.1" | 0.6" | 5.2" | 0.8" | 3.1" | 0.6" | 6.7" | 0.2" |
Type →	stripline	stripline		*u*strip	*u*strip	*u*strip		stripline	stripline
T width →	4mil	6mil		8mil	8mil	8mil		6mil	4mil
Zdiff →	100 ohms	100 ohms		100 ohms	100 ohms	100 ohms		100 ohms	100 ohms
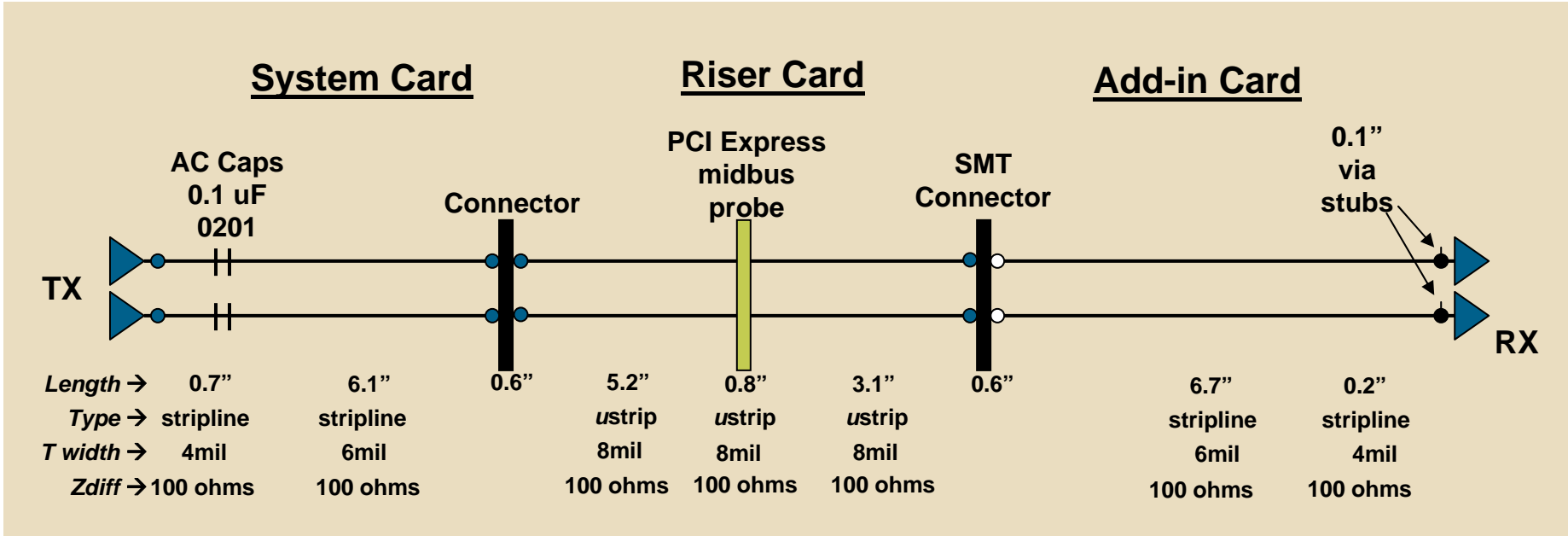
Total Length = 24 in. !!
Will stubs at RX be ok?

- ● Thru via (no stub)
- ● Thru via (0.1 stub)
- ○ Backdrilled via (no stub)

freescale ™
semiconductor

PCI Express® – Simulation Example

Jon Burnett
Signal Integrity Engineer
Freescale Semiconductor
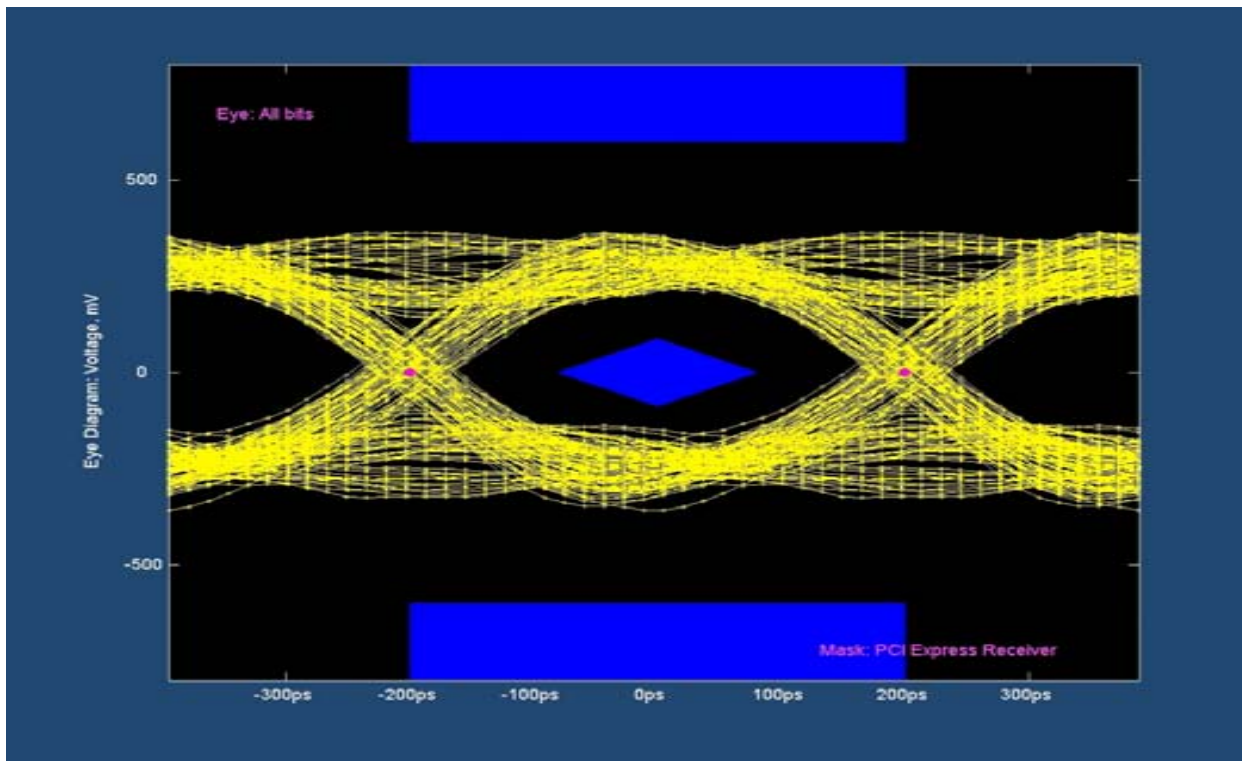
RX Data Eye: 24 inch path through two connectors

Next Generation PCI Express® 2.0

▶ PCI Express® 2.0

- Double the Data Rate of PCIe 1.x – 5 GT/s
- AC parameters and measurements redefined
  - Jitter budget
  - Insertion loss
- Backwards compatible to PCIe 1.x
- "short reach" <= 20" achievable with 1.x routing guidelines and materials
- "long reach" >20" difficult
  - Requires one or more of the following:
    - Improved PCB materials (not FR4)
    - Backdrilling of via stubs
    - Improved (higher speed) connectors
  - PCI SIG struggling to find inexpensive solution to 31" ATCA backplane solution

freescale ™
semiconductor

PCI Express® 1.1 vs 2.0

	PCIe (Gen1)	PCIe2 (Gen2)
Data rate	2.5 GT/s	5.0 GT/s
Unit Interval	400 ps ± 300ppm	200 ps ± 300ppm
TX min voltage	800 mV p-p	800 mV p-p
RX min voltage	175 mV p-p	120 mV p-p
Jitter: Tj @ Receiver	0.6 UI (240 ps)	0.6 UI (120ps)
REFCLK jitter	150 ps (cyc-cyc) 108 ps $J_{phase,p-p}$ @ BER=10^{-12}	150 ps (cyc-cyc) 3.1 ps rms J_{rnd}

freescale™
semiconductor

SGMII Overview

► Transmit and receive data paths leverage the 1000BASE-SX PCS defined in the IEEE 802.3z specification (clause 36).

► Cisco ownership of *Serial-GMII Specification*

► Half the data rate of PCI Express$^®$ 1.x –> 1.25 GT/s

► 8-wire, 6-wire and 4-wire interconnect

  • Cisco SGMII spec rev 1.7 only calls out 6-wire and 8-wire interface– the 4-wire interface is an industry de-facto standard

  • Freescale PowerQUICC$^®$ devices to date only implement 4-wire interface

  • 4-wire interface relies on Clock Data Recovery to reconstruct clock from transmitted data.

► 1 Lane only

► No polarity inversion capability in the differential pair

► LVDS Signaling

► No receiver detection circuit hardware required (AC coupling caps).

  • AC caps are optional, but are required if the transmitter common mode (DC) signaling is not compatible with receiving devices input voltage DC window.

freescale ™
semiconductor

SGMII vs PCI Express® (System Design)

	PCIe (Gen1)	SGMII
Data rate	2.5 GT/s	1.25 GT/s
Unit Interval	400 ps ± 300ppm	800 ps ± 100ppm*
TX min voltage	800 mV p-p	300 mV p-p
RX min voltage	175 mV p-p	200 mV p-p
Jitter: Tj @ Receiver	0.6 UI (240 ps)	0.375 UI (300ps)
REFCLK jitter	150 ps (cyc-cyc) 108 ps $J_{phase,p-p}$ @ BER=10^{-12}	100 ps (cyc-cyc)

freescale™
semiconductor

SGMII System Design Guidelines

► PCI Express® 1.x and SGMII interconnect are very similar

► SGMII is less strict on nearly all spec parameters

- Watch out for stricter frequency tolerance and jitter specs on REFCLK.

► 4-wire SGMII is a defacto standard and has no published interconnect guidelines, therefore…

► Use PCI Express 1.x guidelines for interconnect and routing

- Can route even further with SGMII if necessary due to its reduced data rate → simulate first!!!

- Calculate insertion loss budget using device's minimum TX output and minimum RX input voltages.

- Backdrilling and advanced PCB materials likely not needed.

- Eliminate AC coupling caps if I/O voltage signaling levels are compatible between devices.

freescale ™
semiconductor

PCI Express® – Hardware Interconnect Design

►Summary

- *PCI Express*® offers flexibility in a broad range of architectural topologies

- *PCI Express* provides significant performance advantages over previous generation busses

- Electrical Interconnect design focuses on the two key performance parameters of loss and jitter

- Well-defined interconnect design guidelines facilitate successful first-time designs for *PCI Express* as well as SGMII

freescale
semiconductor