PCIe® 2.0 PHY Architecture

Debendra Das Sharma

Member, EWG

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements
- Enhancements for Robustness
- Summary & Call to Action

Architectural Extensions: Overview

- Changes limited to physical layer only
- Backwards compatible with PCIe 1.1 spec
- Considerations
  - ✓RAS
  - ✓Power efficiency
  - ✓Robustness in design (HVM considerations)
  - ✓Ease of design and validation

Architectural Extensions: Overview

Extensions	Explanation	Benefits
Speed Negotiation	1. Capability to upgrade or downgrade link speed 2. Programmable and inband compliance testing	RAS (improved link uptime), power savings (25%+) Flexibility in compliance testing
Electrical Idle Entry and Exit	Protocol changes to facilitate circuit design	Enhanced robustness, yield, power savings, ease of design (TTM)
Link width upconfigure	Capability to increase the link width up to the initial trained link width	Power savings
Testability enhancements	Receiver Compliance Transmitter Margining Loopback enhancements	Receiver margining Ease of loopback testing at higher speeds
Compliance Entry/Exit	Device Configuration despite link failures	High Availability, enhanced robustness.

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements
- Enhancements for Robustness
- Summary & Call to Action

Speed Negotiation: Philosophy

- Ability to change link speed multiple times without taking the link down
  - ✓ Reliability:
    - Link goes to 2.5G speed if it fails to operate at a higher data rate after being operational for a while
    - Revert back to prior speed if a speed change did not succeed
  - ✓ Power improvements: can change to a lower data rate when bandwidth requirements are low to conserve power and upgrade later when one needs the higher bandwidth
- Ability to notify software on bandwidth change

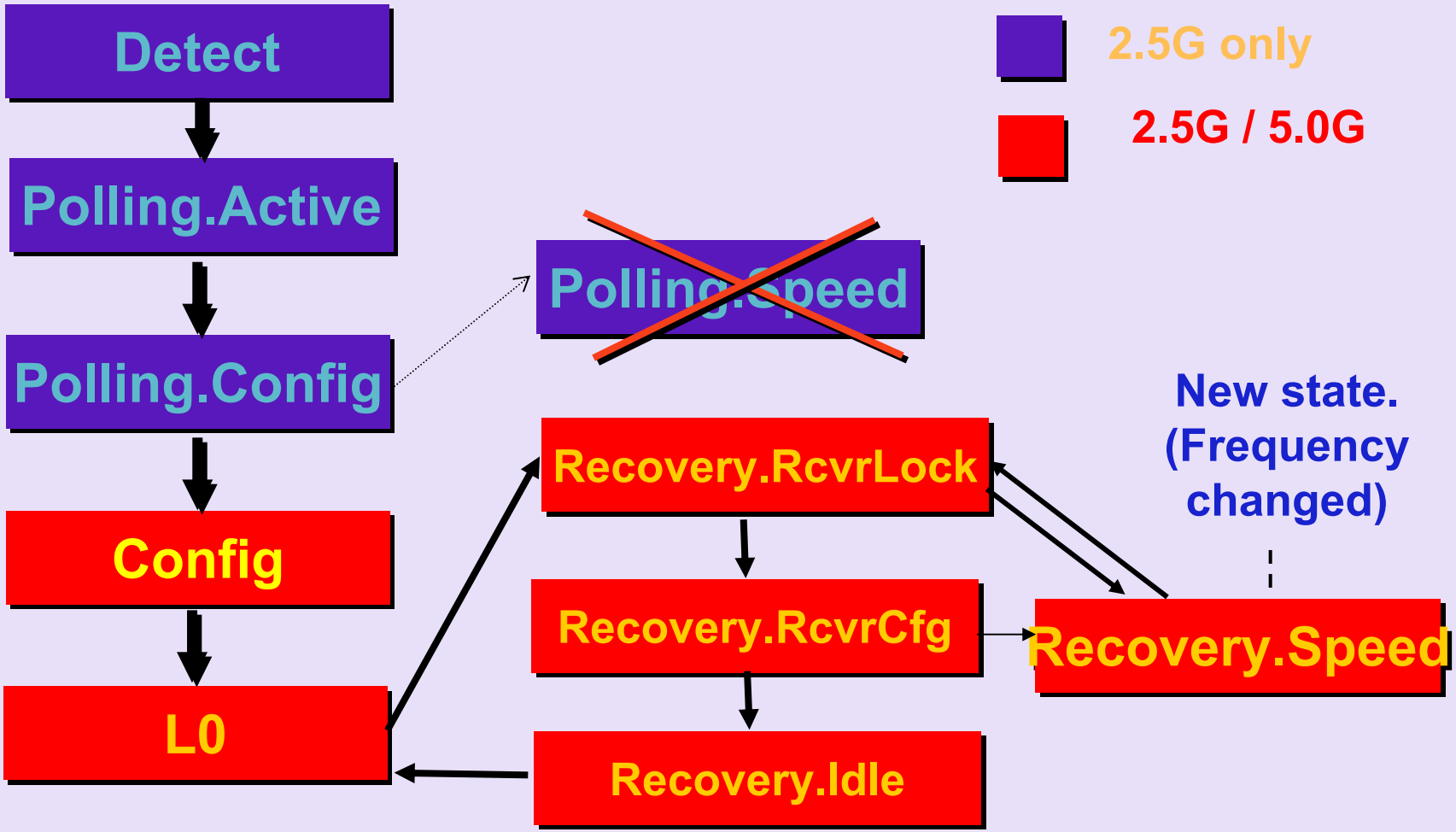
Speed Negotiation

- Initially link trains to L0 in 2.5G data rate
- Supported speeds advertised in all TS ordered sets
  - ✓Supported speeds by the other component noted in Config.Complete and Recovery.RcvrCfg
- Speed changed through handshake in Recovery state
  - ✓New substate: Recovery.Speed
  - ✓Speed changed in Recovery.Speed
- Polling.Speed state obsolete

Speed Negotiation through Recovery

- LTSSM enters Recovery from L0 if a speed change is desired
  - ✓ Can be initiated by hardware or software
  - ✓ Both sides exchange speed information through TS ordered sets in Recovery.RcvrLock and Recovery.RcvrCfg
    - Includes supported speeds as well as intent to change speed
  - ✓ Speed changed in Recovery.Speed

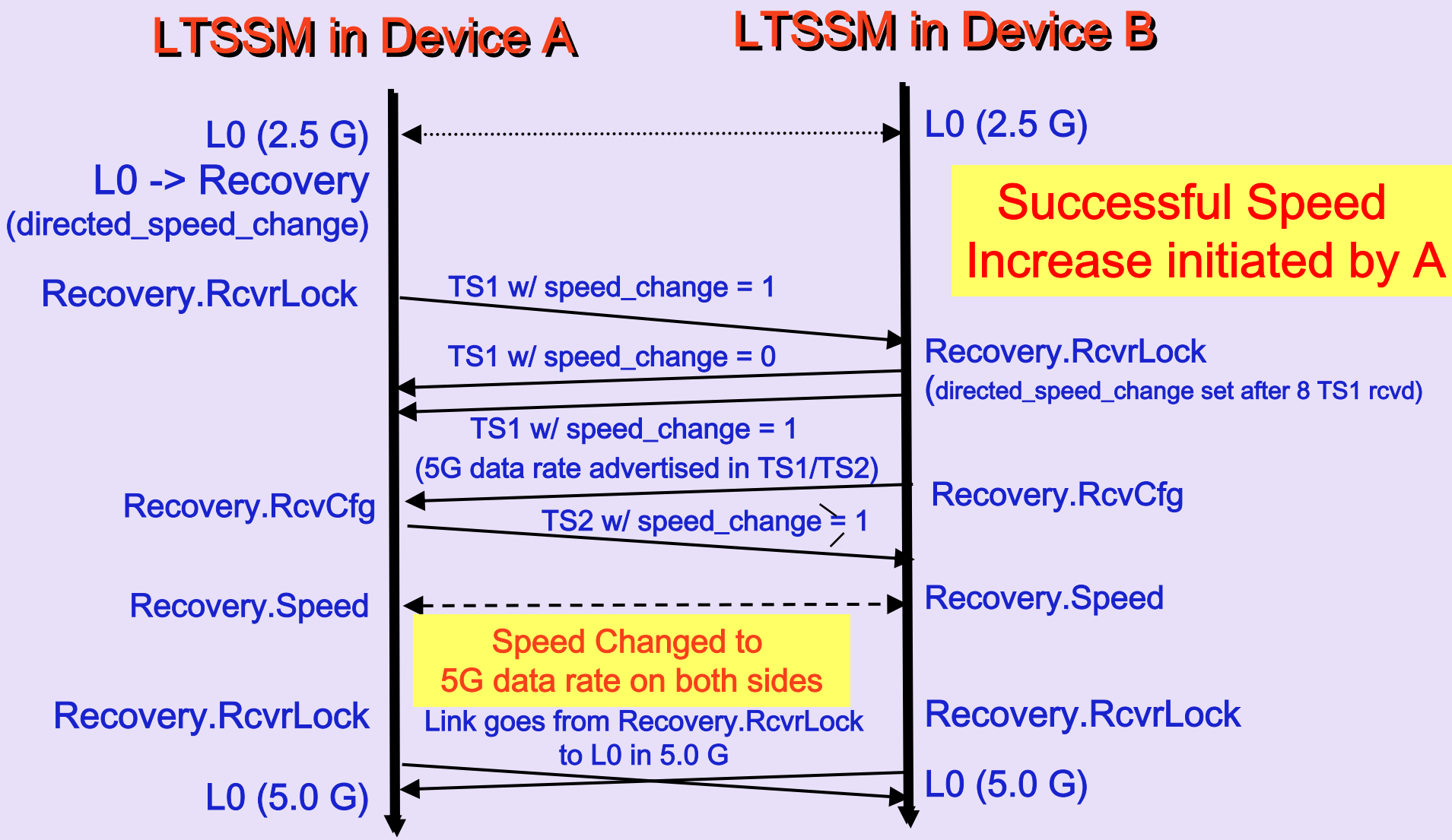
Relevant LTSSM States



Training Sequence (TS1/TS2 changes)

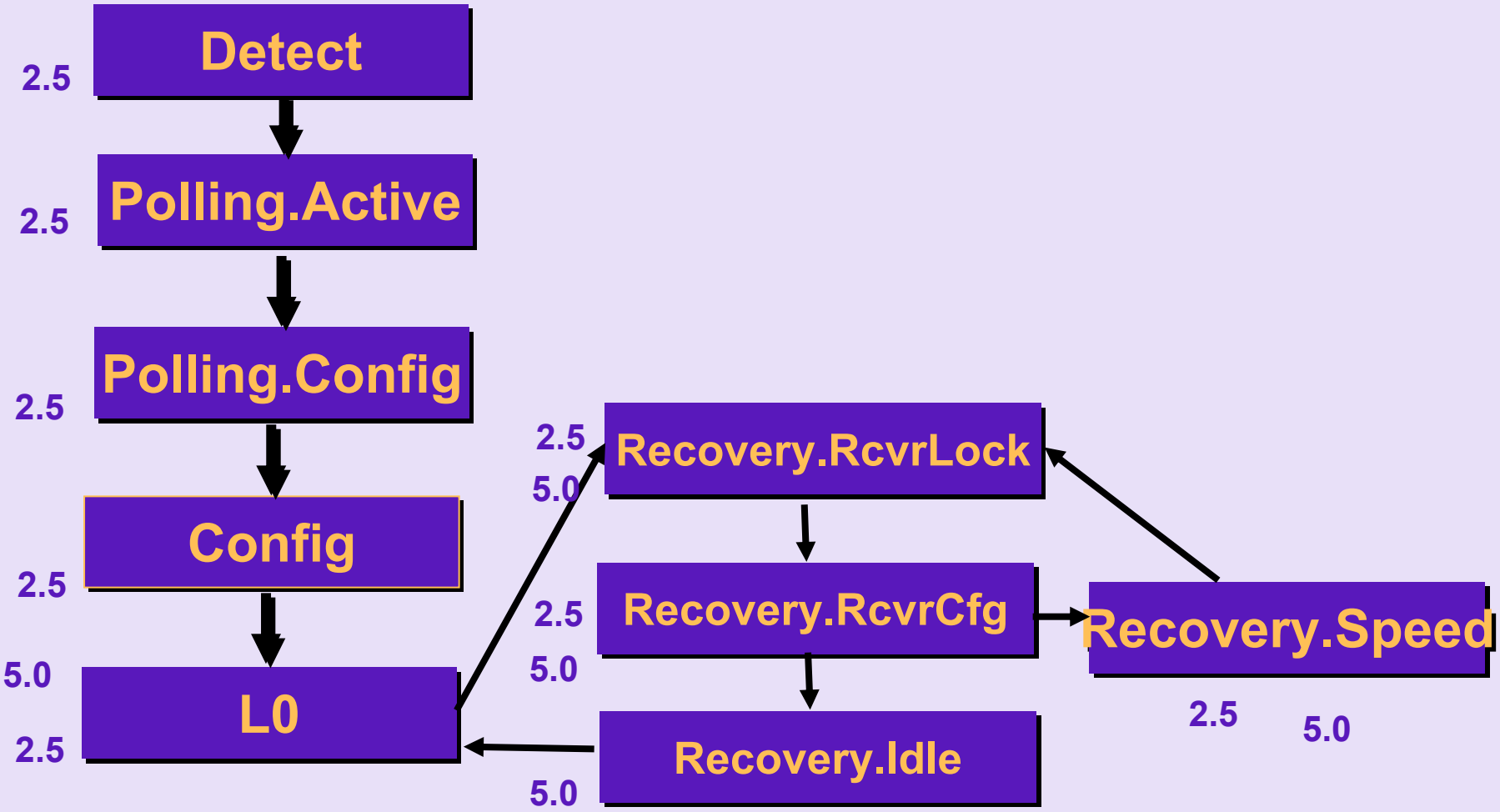
Symbol Number	Encoded Values	Description
4	D2.0, D6.0, D2.4, D6.4, D2.2, D6.2, D2.6, D6.6	Data Rate Identifier Bit 0 – Reserved, set to 0 Bit 1 = 1, 2.5 Gb/s data rate supported Bit 2 = 1, 5 Gb/s data rate supported. Bit 3:5 – Reserved for future gen speeds past gen 2, set to 0 for devices that only support 2.5 Gb/s and/or 5Gb/s speeds. Bit 6: Used for Up_Configure Capability, Autonomous Change (for b/w notification) and Selectable De-emphasis. Different meaning in different states as well as for upstream vs downstream. Bit 7 (speed_change) = 1, Requesting to change the speed of operation. The remaining bits (0-6) specify the highest speed with which we can reliably operate. This bit can be set to 1 only during Recovery.RcvrLock state. All lanes under control of an LTSSM must transmit the same value in this symbol.

LTSSM Speed Change: Example 1

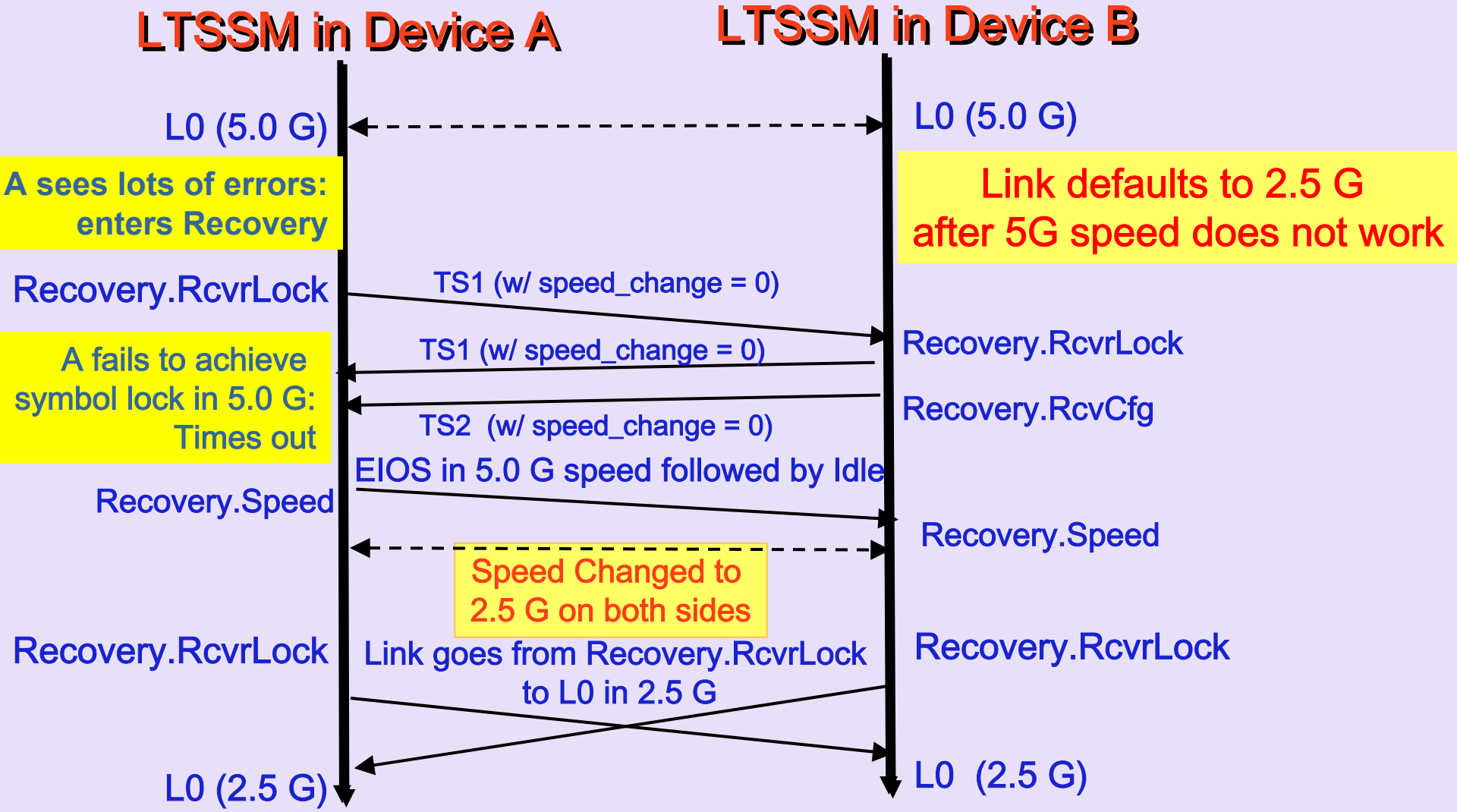


Speed Change : Example 1

LTSSM Speed Negotiation



LTSSM Speed Change: Example 2



Selectable De-emphasis in 5Gb/s

- Under discussion for 0.9 draft of the spec
- Pattern sensitivity to worst case patterns in 5Gb/s
  - ✓ 6.0 dB enables longer loss dominated channels
  - ✓ 3.5 dB enables shorter reflection/crosstalk dominated channels
- CSR to set the de-emphasis level in both upstream and downstream component
- Downstream component may request a 3.5 dB de-emphasis in Recovery.RcvrLock (TS1) when requesting 5Gb/s speed change
- Upstream component sets the de-emphasis level for the channel in TS2 in Recovery.RcvrCfg
  - ✓ Enforced in Recovery.Speed when changing speed to 5GT/s

Config Changes

- In Link Capabilities:
  Maximum Link Speed → Supported Link Speeds
  - ✓ Reports component capabilities – encoding added for new data rate (5.0 G)
- New control field: Target Link Speed
  - ✓ Sets target & upper limit on link operational speed
  - ✓ Set only in upstream component
  - ✓ Speed change occurs by setting link retrain bit
  - ✓ Also sets speed for software initiated Compliance
- In Link Status:
  Link Speed → Current Link Speed
  - ✓ Reports current operational link speed when link is up

Config Changes

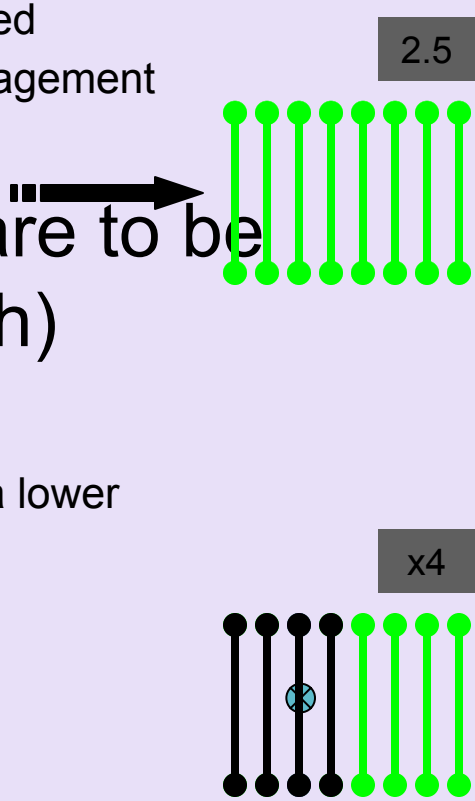
- Hardware Autonomous Speed Control
  - ✓ Controls ability of Hardware to reduce speed
  - ✓ Upper layers in hardware controlling speed changes dynamically use implementation specific policies
  - ✓ Speed reduced for reliability reasons not affected by this bit
- Link Bandwidth change notification mechanism
  - ✓ Link B/W Management Status
  - ✓ Link Autonomous B/W Status
- Software initiated Compliance
  - ✓ Enter Compliance and Enter Modified Compliance
- Selectable De-emphasis (rev 0.9 direction)
  - ✓ Selectable de-emphasis
  - ✓ Enforce de-emphasis (upstream only)

Speed Change Rules

- Components may advertise a subset of speeds supported to cap the post-negotiation link speed
  - ✓ Components may change advertised supported speeds *without* requesting a Link speed change by driving link through Recovery without setting speed change bit in TS's
- 200ms "waiting period" required following a failed link speed change
  - ✓ May try again earlier if other component advertises support for failed speed as described above
- If link reliability unacceptably low, either component permitted to lower link speed
  - ✓ Criteria for acceptable link reliability implementation specific
  - ✓ Not dependent on setting of Hardware Autonomous Speed Control bit

Link Speed Controls

- Hardware automatically trains to the greatest link speed possible
  - ✓ Link trains to 2.5GHz in L0 and then attempts to train higher
  - ✓ Hardware and/or software can place an upper bound on the speed
  - ✓ Hardware and/or software can change the speed for power management
- New mechanism for PCIe-aware software to be notified when bandwidth (speed or width) changes
  - ✓ Hardware notification to software when marginal link retrains to a lower bandwidth



Speed of Polling.Compliance

- Two ways to set the speed
  - ✓ Inband Method: (difference in 0.9 draft direction from 0.7 draft)
    - No compliance comparator to choose between 2.5 Gb/s vs 5 Gb/s
    - De-emphasis levels in 5Gb/s accounted for
    - Cycle through 2.5 Gb/s, 5 Gb/s at -6 dB and 5 Gb/s at -3.5 dB de-emphasis
    - Compliance board will send 100 MHz signals for about 1msec on one leg of differential pair at 350 mV peak-to-peak on any lane to cycle through each of the three settings
  - ✓ CSR method (new method)
    - Link Control2 Register Fields: Target Link Speed, Enter Compliance, Enter Modified Compliance, selectable de-emphasis, and Transmit Margin
    - Procedure:
      - Link trains to L0
      - CSR bits written on both sides to enter compliance with same speed
      - SBRE done in upstream.
      - This causes link to go to Detect and eventually to Compliance at the desired speed, type of compliance (regular vs receiver compliance), transmit margin, and de-emphasis level
- Speed determined prior to entering Polling.Compliance

Polling.Compliance: Transition

- Speed of Polling.Compliance greater than 2.5 G data rate:
  - EIOS needs to be transmitted followed by
  - Electrical idle between 1 ms to 2 ms before changing speed
- Exit condition: Polling.Compliance to Polling.Active:
  - ✓ Electrical idle exit on lanes and enter_compliance =0
  - ✓ Enter_compliance changes from 1 to 0
  - ✓ EIOS detected or electrical idle detected/inferred and enter_compliance is set to 1 in downstream component
- On exit from Polling.Compliance:
  - ✓ If speed if greater than 2.5 G **or** enter_compliance CSR is set
    - Device transmits 8 consecutive EIOS; then
    - Transmitter lanes go electrical idle for 1 to 2 msec; and
    - Prior to entering Polling.Active, 2.5G data rate is selected
- CSR mechanism can be used to cause exit from Polling.Compliance if the entry was through CSR mechanism

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements
- Enhancements for Robustness
- Summary & Call to Action

Electrical Idle Exit Challenges

- Exit from Electrical Idle a challenge in 5.0 G speed
  - ✓ Receiver sensitivity in 5G speed: 120 mV
  - ✓ Idle detection threshold in 5G speed: 175 mV
  - ✓ COM with a run length of 5 does ensure that to cross the idle detection threshold but the frequency of COM is not high enough
- Expect exit EI to be detected by circuits in 2.5 G speed
- Need a low-frequency recurring pattern for 5.0 G:
  - ✓ Electrical Idle Exit Sequence (EIES) defined
    - – COM followed by 14 K28.7 followed by TS1 identifier
    - – Effectively 5 1's followed by 5 0's 13 times => low frequency
  - ✓ Requirements relaxed to detect exit EI in *non-2.5 G* speeds
  - ✓ One EIES sent after every 32 TS1/TS2'es.
    - – Count reset to 0 after receiving first TS2 in Recovery.RcvrCfg
    - – During this transition can go up to 63 TS1/TS2 between two EIES
    - – EIES count reset and not sent when we stop sending TS1/TS2
  - ✓ EIES sent before the first TS1 in a TS1/TS2 sequence.
  - ✓ Sent in Recovery and Config.Linkwidth.Start

EIES Sequence

Symbol Number	Encoded Values	Description
0	K28.5	COMMA code group for Symbol alignment
1-14	K28.7	K Symbol with low frequency components for helping achieve exit from electrical idle
15	D10.2	TS1 Identifier

Electrical Idle Exit Usage

- Exit from EI needed in
  - ✓Detect.Active (TS1)
  - ✓Polling.Active (TS1/ TS2)
  - ✓Polling.Compliance (TS1)
  - ✓ Rx_L0s.Idle (FTS/ TS1 on errors)
  - ✓ L1.Idle (TS1)
  - ✓ L2.TransmitWake (TS1)
  - ✓ Disabled (TS1)
  - ✓Recovery.Speed (5G) for inferring EI (TS1/ TS2)

Changes to ensure EI Exit

- In all states except Rx.L0s, the LTSSM gets the initial EIES to detect an exit EI
- EIES repeated periodically (every 32 TS1/TS2) during TS1/TS2 transmission for robustness
  - ✓ Also helps in inferring electrical idle condition in Recovery for the failure cases
    - E.g., Lack of exit EI in a 16000 UI interval can be inferred as EI
- TS1 ordered set has enough high frequency components to achieve bit lock

Changes to ensure EI Exit

- FTS changes for non-2.5G speeds
  - ✓Prepend the FTS sequence with 4 to 8 K 28.7
  - ✓4 sets of K28.7 helps detect exit from EI
  - ✓FTS used to achieve symbol lock
  - ✓If circuits did not detect exit EI with the prepended sequence (e.g., missed K28.7)
    - FTS has enough low frequency components to cause exit EI
    - Designs can optimize (N_FTS) at non-2.5G frequency if they can exit EI with 4 K28.7

5GT/s: Electrical Idle Entry Challenges

- EI entry detection difficult. Required in:
  - ✓ L0 (for surprise detach)
  - ✓ Loopback.Active (as slave to know when master terminates)
  - ✓ Recovery.Speed (introduced in PCIe 2.0 spec)
- PCIe 2.0 Spec allows to *infer* EI as an alternative to detecting EI in all speeds
  - ✓ L0: EI may be inferred if receiver did not receive COM in 128us. SOS has COM
  - ✓ Recovery.RcvrCfg and on a successful speed negotiation in Recovery.Speed: absence of COM in any configured lane in a 1280 UI interval
  - ✓ Recovery.Speed with unsuccessful speed negotiation: absence of exit from EI in 16000 UI interval
  - ✓ Loopback.Active:Inference (optional) or reset

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements
- Enhancements for Robustness
- Summary & Call to Action

Link Upconfigure Capability

- Added a new capability to upconfigure linkwidth
- One usage model is a device can downconfigure link width when bandwidth requirements are low to conserve power and later upconfigure when bandwidth requirements go higher.
- A link can be upconfigured up to the width that was initially negotiated. Example: a x8 link can go down to a x4, subsequently go down to a x1, and later go back up to a x8 link.
- Link upconfigure capability advertised in the TS2 ordered sets in Config.Complete state only.

Link Upconfigure Capability

- Planned linkwidth downconfigure if both sides advertised link upconfigure capability during the last sojourn in Config.Complete substate
- A component may change its upconfigure capability on a subsequent entry to Config.Complete substate
- Link width upconfigure and planned downconfigure:
  - ✓ L0 -> Recovery.RcvrLock -> Recovery.RcvrCfg -> Recovery.Idle (immediate transition) -> Config substates -> L0
- Inactive lanes continue to be associated with the LTSSM if the lanes were associated with the LTSSM the first time it entered L0 state

Link Upconfigure Capability

- If upstream component intends to upconfigure, it sends TS1's with link and lane numbers set to PAD in Config.Linkwidth.Start initially and later sends the link number after receiving two consecutive TS1'es on all lanes it intends to operate on or after a 1ms timeout
- If downstream component intends to upconfigure, it sends TS1'es on all lanes it intends to operate on in Config.Linkwidth.Start substate. However, it holds off sending the link number until it receives two consecutive TS1 ordered sets on all lanes it intends to operate or after a 1ms timeout
- If any inactive lane receives an exit from electrical idle that the component does not intend to initiate to become active, the component starts sending TS1 ordered sets after receiving two consecutive TS1 ordered sets

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements (0.9 draft direction)
- Enhancements for Robustness
- Summary & Call to Action

Receiver Compliance

- A new mode in Polling.Compliance
  - ✓ 1.1 flavor of Polling.Compliance still there
- Device under test sends out modified compliance pattern
  - ✓ Includes error count for the corresponding receiver
  - ✓ Reflects the number of 8b/10b errors encountered by the receiver
  - ✓ Error count helps to characterize (margin) receiver
- Two methods for entry
  - ✓ TS1 ordered sets with the ‘compliance receive’ bit set
    - Speed and De-emphasis determined by the corresponding fields in the received TS1 ordered sets
  - ✓ Setting the “enter compliance” and “enter modified compliance” bits in L0 followed by an SBRE in the upstream component
    - Speed and de-emphasis determined by the corresponding CSR bits in the Link Control2 register
- Device enters Polling.Compliance from Polling.Active
  - ✓ 5Gb/s operation: Electrical idle; adjust speed and de-emphasis level
  - ✓ Tx level selected based on Link Control2 register bits on entry
  - ✓ Exit from Polling.Compliance if directed

Transmitter Margining

- "Transmit Margin" (3 bits) in the Link Control2 register
- Controls the value of the non-deemphasized voltage levels at the transmitter pins
- Value sampled on entry to: Polling.Compliance and Recovery.RcvrLock
  - ✓ Default values applied from Polling.Active onwards (until sampled in Recovery.RcvrLock)
- CSR value set to 000b on entry to Polling.Configuration
- Encodings:
  - ✓ 000: Normal Operating range (default)
  - ✓ 001: 800 – 1200 mv full swing (400 – 700 mV half swing)
  - ✓ 010 – (n-1): monotonic; non-zero slope; n > 3 and n <7
  - ✓ n: 200 – 400 mV full swing (100 – 200 mV half swing)
  - ✓ (n+1) -111: Reserved
- De-emphasis can be within 1 dB of spec defined range when operating in 5Gb/s

Loopback Enhancements

- A new mode to force slave to loopback even when it does not achieve symbol lock
  - ✓ TS1 control bit : compliance receive
- Speed and de-emphasis change can happen in Loopback.Entry state for a slave on transition from Configuration
  - ✓ Helps test equipment do assumption based training
- Exit from Loopback under “if directed” (reset) or four consecutive EIOS
  - ✓ 8 EIOS needs to be sent by 2.0 spec compliant devices prior to going electrical idle
  - ✓ Still maintain the 1.1 exit under EIOS or electrical idle detected (with extension to inferred) as optional in 2.5 Gb/s speed

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Link Upconfigure Capability
- Testability Enhancements
- Enhancements for Robustness
- Summary & Call to Action

Enhancements for Robustness

- Changes to Polling.Compliance for HA
  - ✓ Entry condition relaxed from any lane that detected a receiver but did not get an exit from Electrical Idle to multiple lanes
  - ✓ Must go to Polling.Compliance if all lanes that detected a receiver did not get an exit from electrical idle
- Electrical Idle Ordered Set extension for >2.5 G speeds:
  - ✓ Two consecutive sets of COM, IDL, IDL, IDL
- Electrical Idle detection (optional):
  - ✓ Received signals switching at a frequency greater than 125MHz

Agenda

- Overview of Architectural Extensions
- LTSSM Speed Negotiation
- Electrical Idle Entry and Exit
- Testability Enhancements
- Enhancements for Robustness
- Link Upconfigure Capability
- Summary & Call to Action

Summary & Call to Action

- Track PCIe 2.0 development work to be ready for PCIe 2.0 components
  - ✓ PCIe 1.1 Components not directly affected
- Take advantage of the flexibility offered by LTSSM speed change and link upconfigure capability to optimize for power and HA
- Take advantage of the flexibility offered by compliance : inband as well as CSR mechanisms
- Take advantage of a simple electrical idle detection circuitry with HVM and low power advantages by adopting the protocol changes associated with electrical idle
- Take advantage of the testability enhancements for TTM advantages
- Make robust designs by adopting HA related enhancements

Thank you for attending the
PCI-SIG Developers Conference 2006.

For more information please go to
www.pcisig.com

PCIe 2.0 PHY Architecture

Debendra Das Sharma

Member, EWG