FPGA-integrated PCIe® 3.0 Digital IP Architecture

Divya Vijayaraghavan

Altera

Disclaimer

Presentation Disclaimer: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Copyright © 2011, PCI-SIG, All Rights Reserved

Acknowledgements

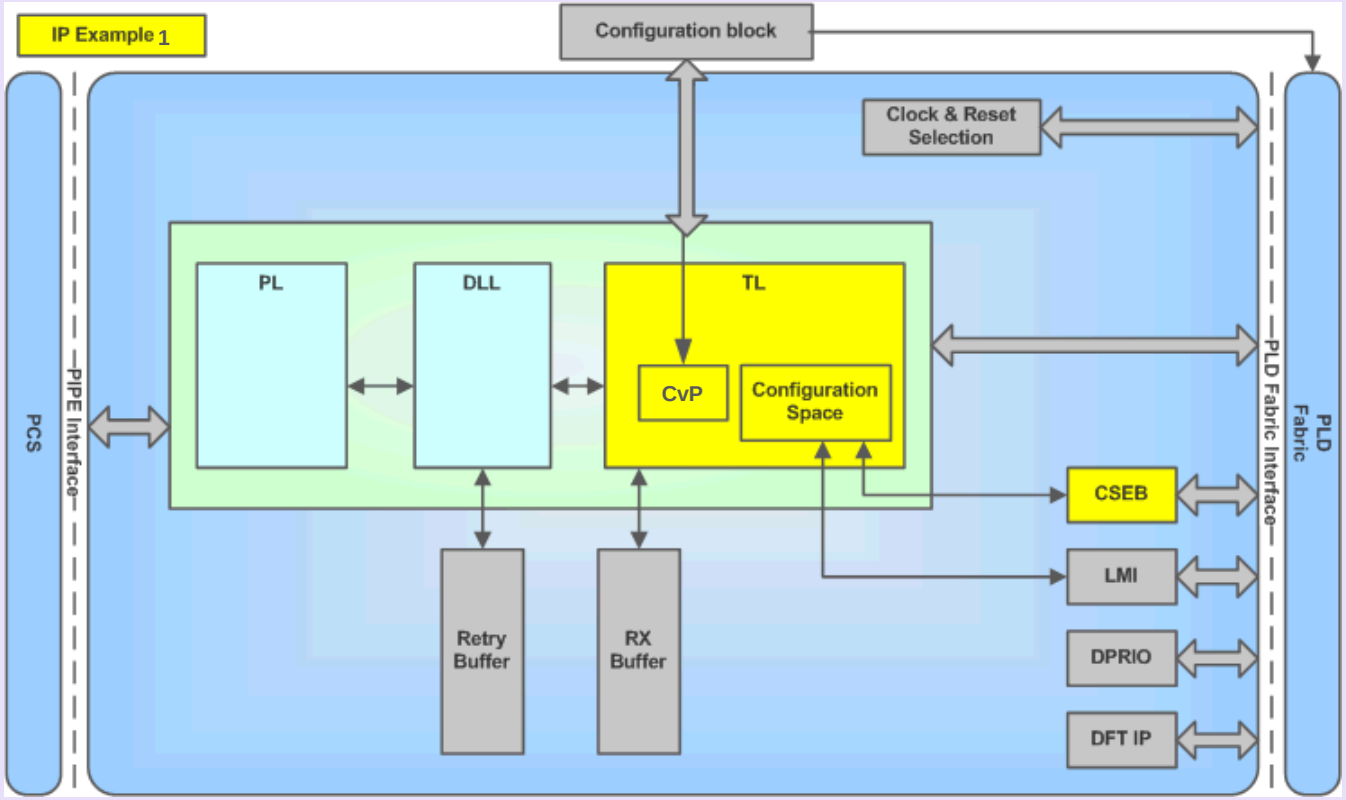
- I would like to acknowledge the contributions of the following individuals
  - ✓ Arye Ziklik
  - ✓ Chong Lee
  - ✓ Chris Finan
  - ✓ Cora Mau
  - ✓ Gopi Krishnamurthy
  - ✓ Ning Xue
  - ✓ Philippe Molson

Overview

- FPGA-integrated PCI Express 3.0 hard protocol stack leverages inherent flexibility of device

- Two examples of highly configurable IP
  - ✓ Transaction Layer IP with Configuration Shadow/Extension Mechanism
  - ✓ Equalization Coefficient Auto-negotiation IP in PCI Express Physical Layer

Transaction Layer IP with Configuration Shadow/Extension Mechanism



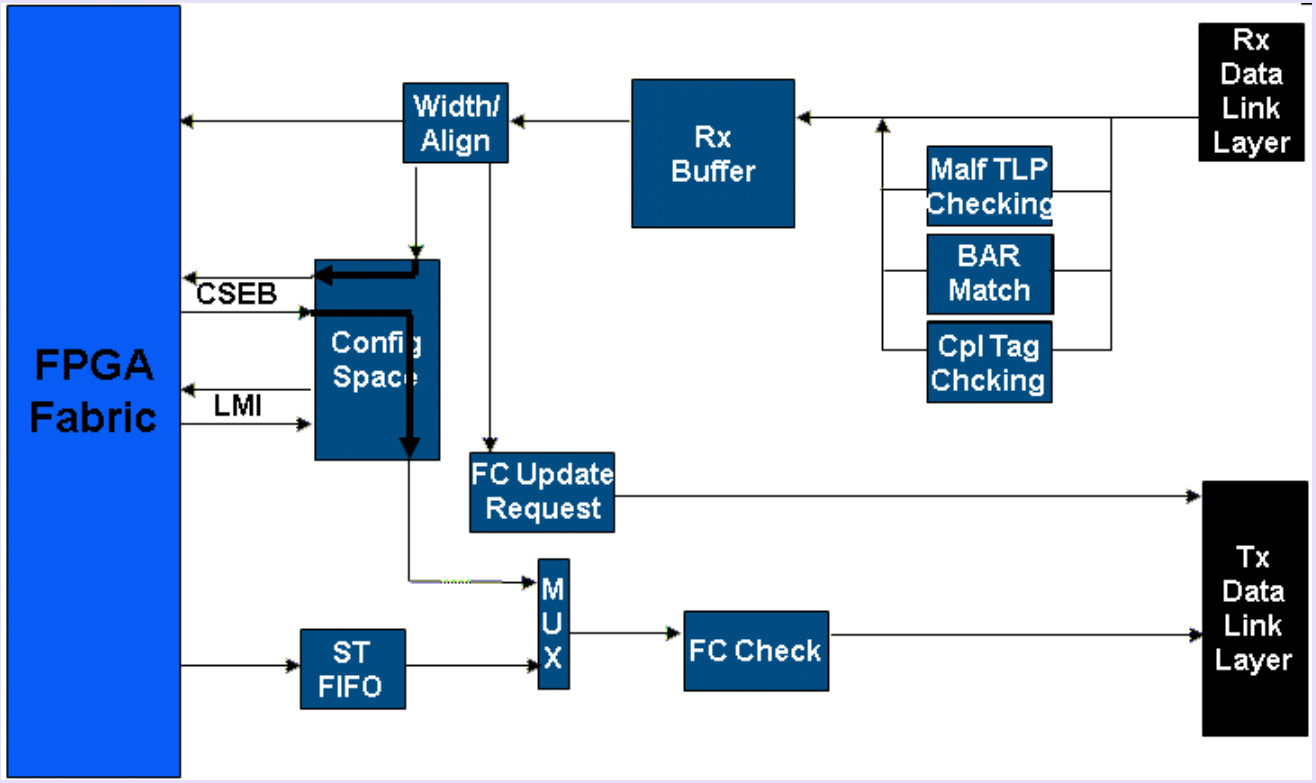
Configuration Shadow/Extension Bus

- Programmable ability to extend or bypass hard configuration space registers to use soft configuration space when required via a Configuration Shadow/Extension Bus (CSEB)
- CSEB may be effectively utilized to implement custom extensions to configuration space
  - ✓ Multiple functions
  - ✓ Single Root I/O Virtualization (SR-IOV)

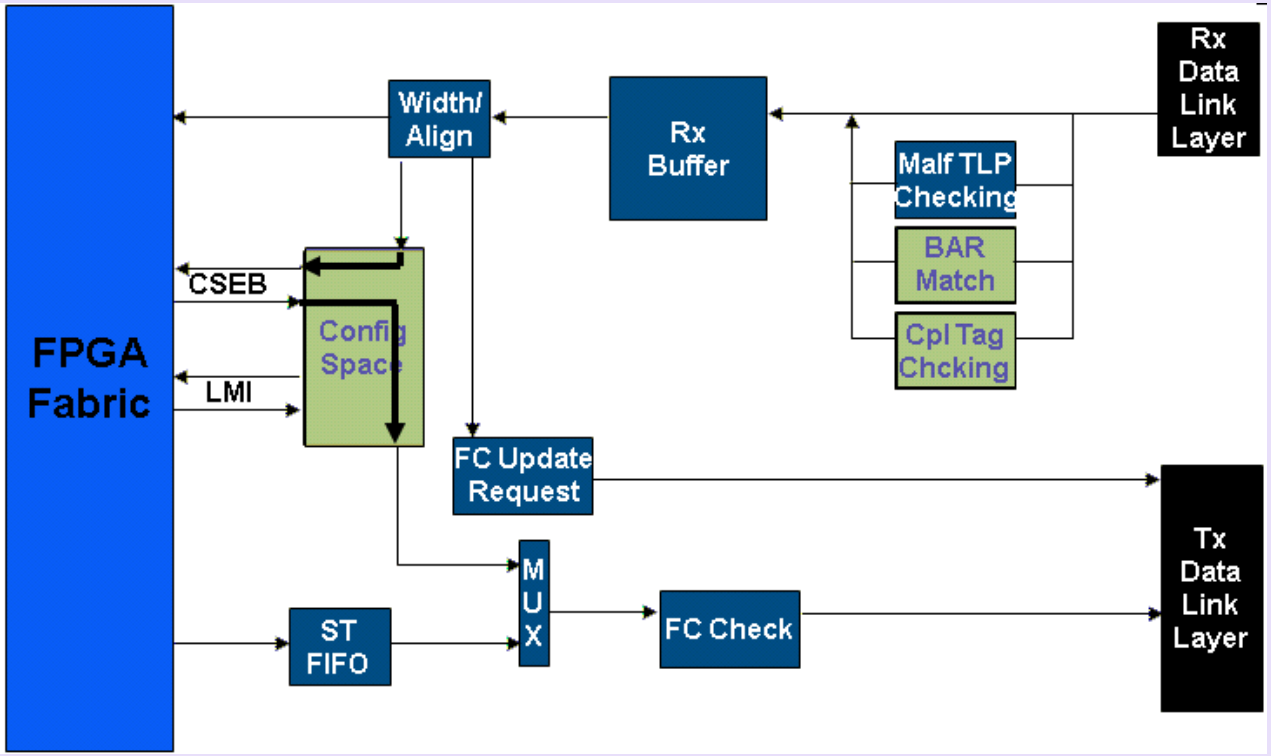
Configuration Bypass

- PCIe 3.0 Transaction Layer's hard configuration space registers bypassed and soft configuration space registers utilized via the CSEB
- All Type 0 configuration transactions redirected to CSEB and responded to by soft logic
- Entire TLP Base Address Register (BAR) matching and completion tag checking logic moved to soft implementation as well
- Generating of unsupported request completions and error logging for requests that do not match a BAR are in soft implementation

Embedded Configuration Space



Bypassed Configuration Space



Transaction Layer Bypass

- Advanced PCIe capabilities addressed by bypassing entire hard Transaction Layer IP
- Enables extension of PCIe configuration space in soft IP
- Provides implementers with ability to interface to underlying hard Data Link Layer IP
- Flexibility of customizing Transaction Layer's functional modules in soft IP
  - ✓ Custom buffering and flow control schemes

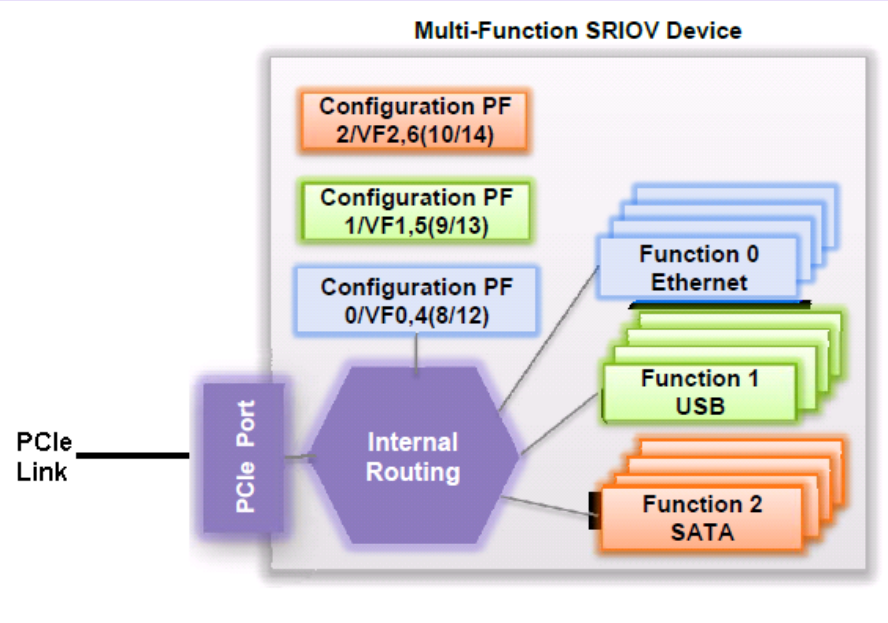
Extended Configuration Space Applications (1/2)

- Configuration space bypass provides programmable logic device with means to address PCIe enhancements
- Multi-function end point
  - ✓ Enables Host CPU behind RC which executes system initialization code and runs software drivers to connect to multiple heterogeneous user devices via PCIe link
  - ✓ PCIe link bandwidth shared by several different devices, each device representing a function and each function requiring a software driver

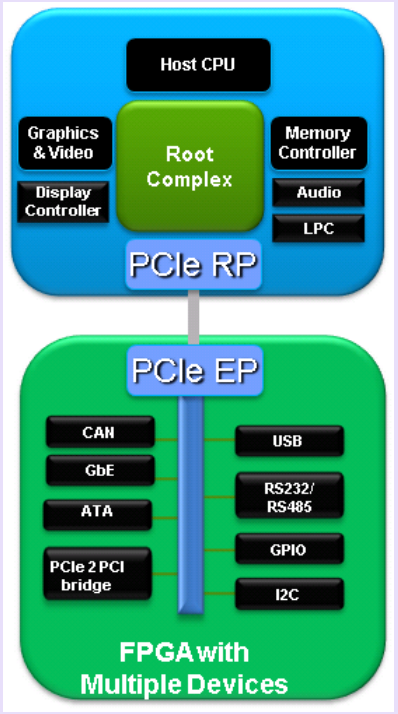
Extended Configuration Space Applications (2/2)

- SR-IOV
  - ✓ Virtualization is the concept of sharing a single platform comprising a physical set of hardware such as CPU, I/O and memory among multiple software operating environments such as operating systems
  - ✓ Multiple physical functions share bandwidth of PCIe link and multiple virtual functions share a single physical function
  - ✓ Application could potentially utilize up to 2^16 physical and virtual functions all of which cannot be implemented in hard IP
  - ✓ CSEB in conjunction with Configuration Bypass provides means to address such applications in soft IP

Devices Utilizing Extended Configuration Space



Multi-function SR-IOV Device



Multi-function End Point

Configuration via PCI Express (CvP)

- CvP enables FPGA fabric configuration to be initially loaded and/or updated though PCIe link
- Made possible by implementing Vendor-specific Extended Capability (VSEC) structures
- Host device or CPU writes to these registers and polls status bits in them to communicate with configuration logic in FPGA
- Loads fabric image using memory or configuration write transactions

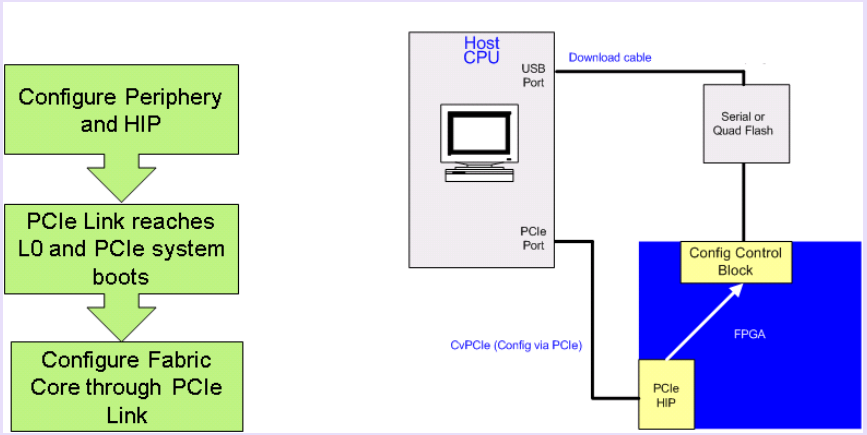
CvP Advantages

- Greatly simplifies user software model for configuration since PCIe protocol and user application topology are utilized to initialize and update FPGA fabric by smart host
- CPU continues to operate while FPGA fabric updates and no host CPU stall or re-boot is required following fabric image updates
- Reduction of dedicated FPGA configuration pins
- Lower cost by eliminating expensive non-volatile configuration devices
- Provides ability to configure multiple FPGAs through single CvP link
- Facilitates user application image protection as image copies are inaccessible to hackers
- Host CPU could save power by loading low power temporary images based on user application profile

CvP Modes of Operation

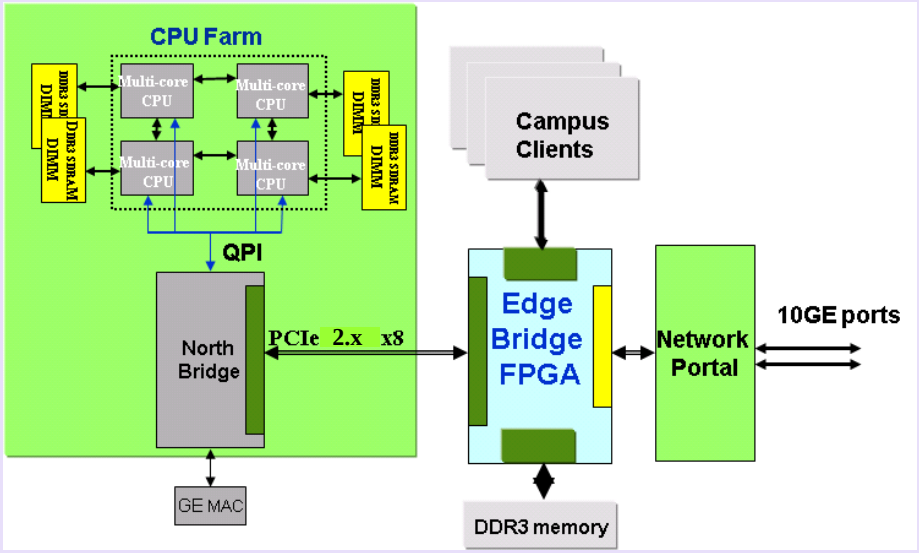
- Single image load scenario
  - ✓ FPGA periphery and hard PCIe IP initially configured utilizing an external device
  - ✓ PCIe link reaches L0 state and PCIe system boots following which FPGA fabric core is configured through PCIe link
- Multi-image update scenario
  - ✓ Configure FPGA fabric as in single image load case and additionally utilize PCIe link to update fabric core
  - ✓ Example: Utilize full-performance FPGA image during work-week hours and use CvP to switch to low-power FPGA image mode at night, with CPU farm and PCIe link remaining active at all times

Configuration via PCI Express Usage

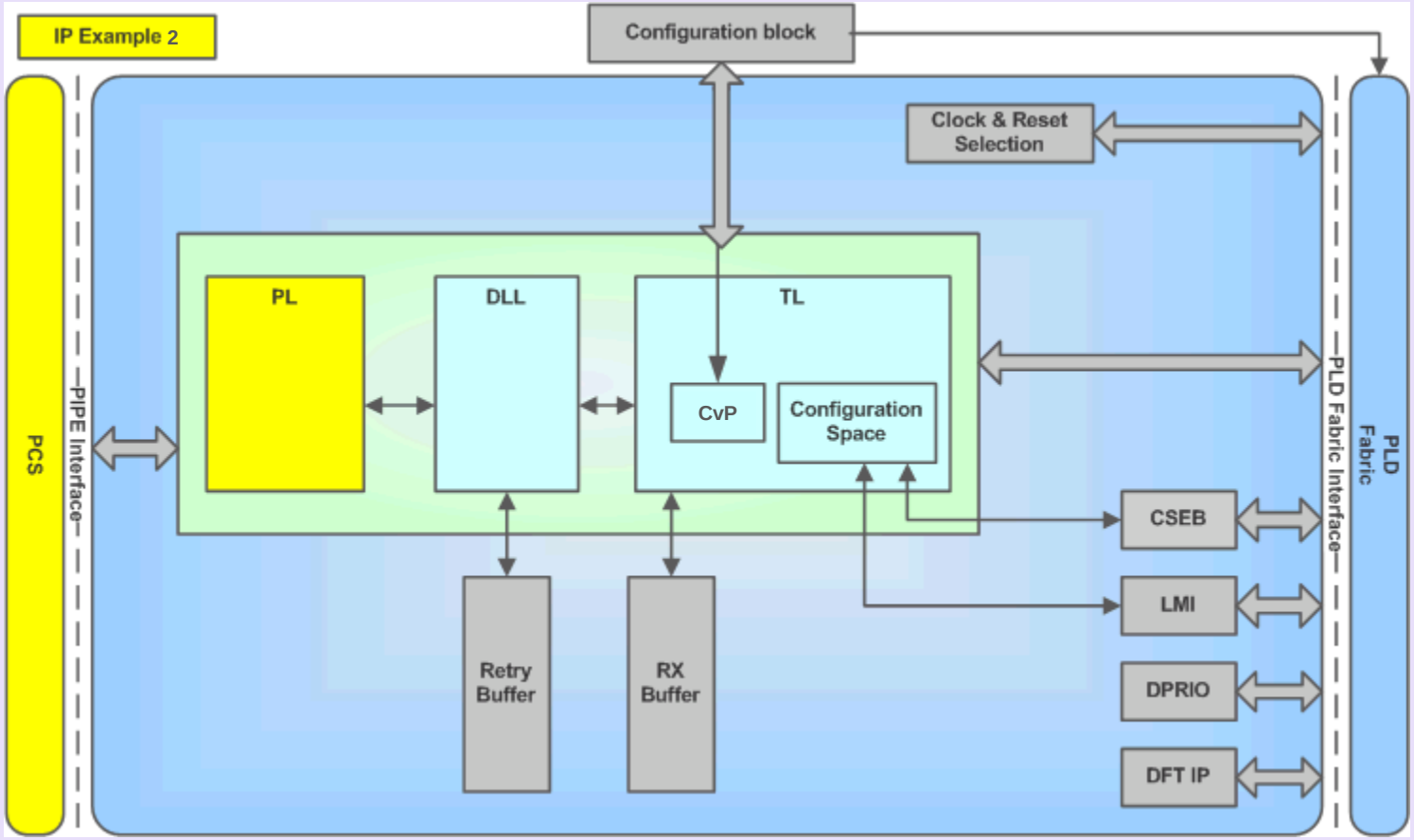


CvP Single Image Load Typical Use

CvP Example: Multi-image Flow



Equalization Coefficient Auto-negotiation IP



Equalization Co-efficient Auto-negotiation

- Impossible to utilize constant pre-determined transceiver equalization settings at 8Gbps and above
- Links must be established with target devices of varying characteristics
- Requirement to dynamically negotiate transceiver's Tx and Rx equalization coefficient settings
  - ✓ Utilize remote receiver's BER determination to achieve optimum eye opening at remote device's receiver and vice versa

PCI Express 3.0 Equalization (1/4)

- New equalization algorithm executed by Recovery.Equalization LTSSM state
- Four distinct phases (0-3) utilized for auto-negotiation of
  - ✓ “coefficients” or transmit finite impulse response filter taps
  - ✓ “presets” which are encoded specification-defined coefficient values

PCI Express 3.0 Equalization (2/4)

- Phase 0 operates at 2.5 or 5 Gbps, other phases at 8 Gbps
- Equalization process may be shortened by bypassing Phases 2 and 3
- At device boot, root complex (RC) transmit preset values and receive preset hints are initialized by firmware

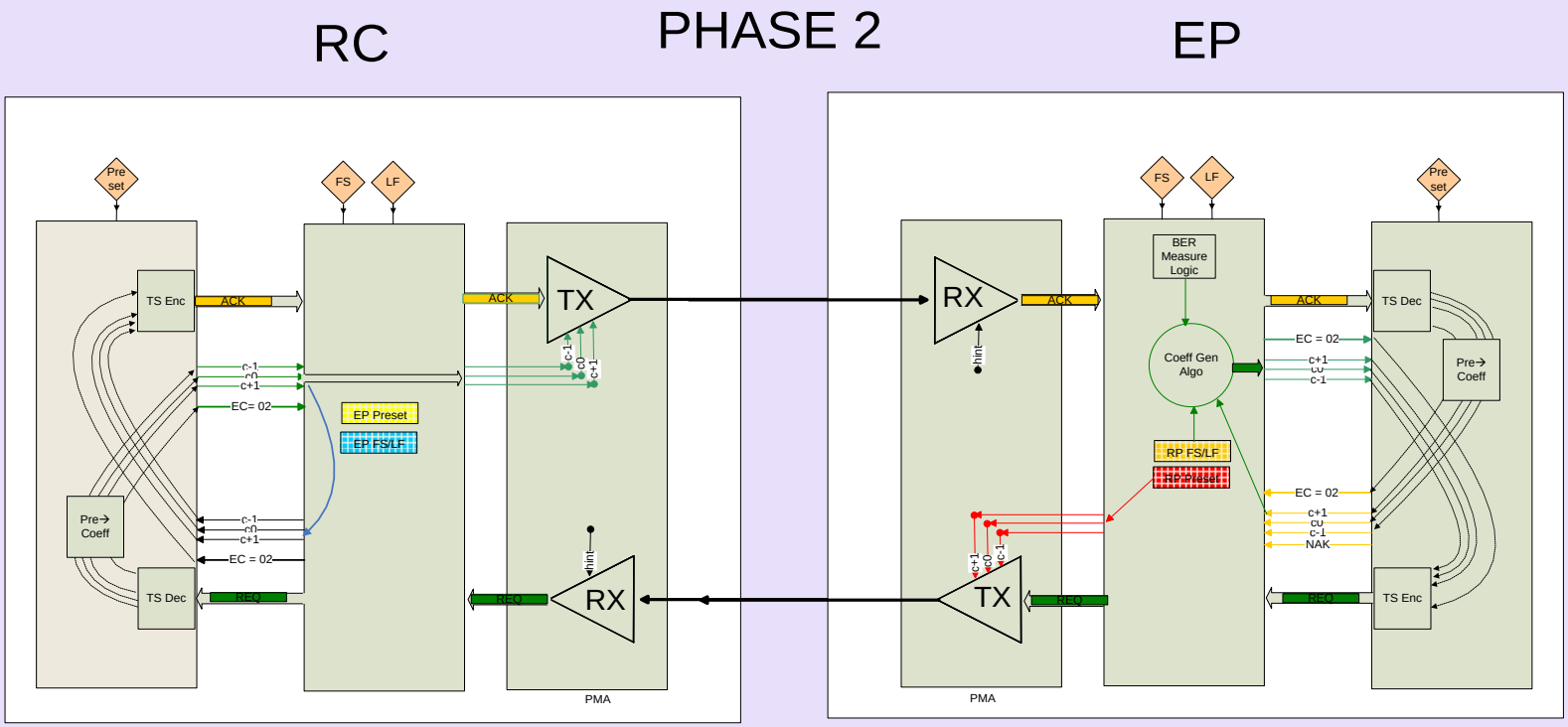
PCI Express 3.0 Equalization (3/4)

- Phase 0
  - ✓ Tx and Rx presets communicated from RC to End Point (EP) at 2.5 or 5 Gbps, to be utilized at 8 Gbps
- Phase 1
  - ✓ Link comes up at 8 Gbps utilizing presets determined during Phase 0
  - ✓ Link partners proceed to exchange full swing (FS) and low frequency (LF) values
  - ✓ Preset values utilized in phase 1 ensure that a BER of 10^{-4} is achieved to ensure that at least two consecutive 8 Gbps PCIe training sequences may be received reliably

PCI Express 3.0 Equalization (4/4)

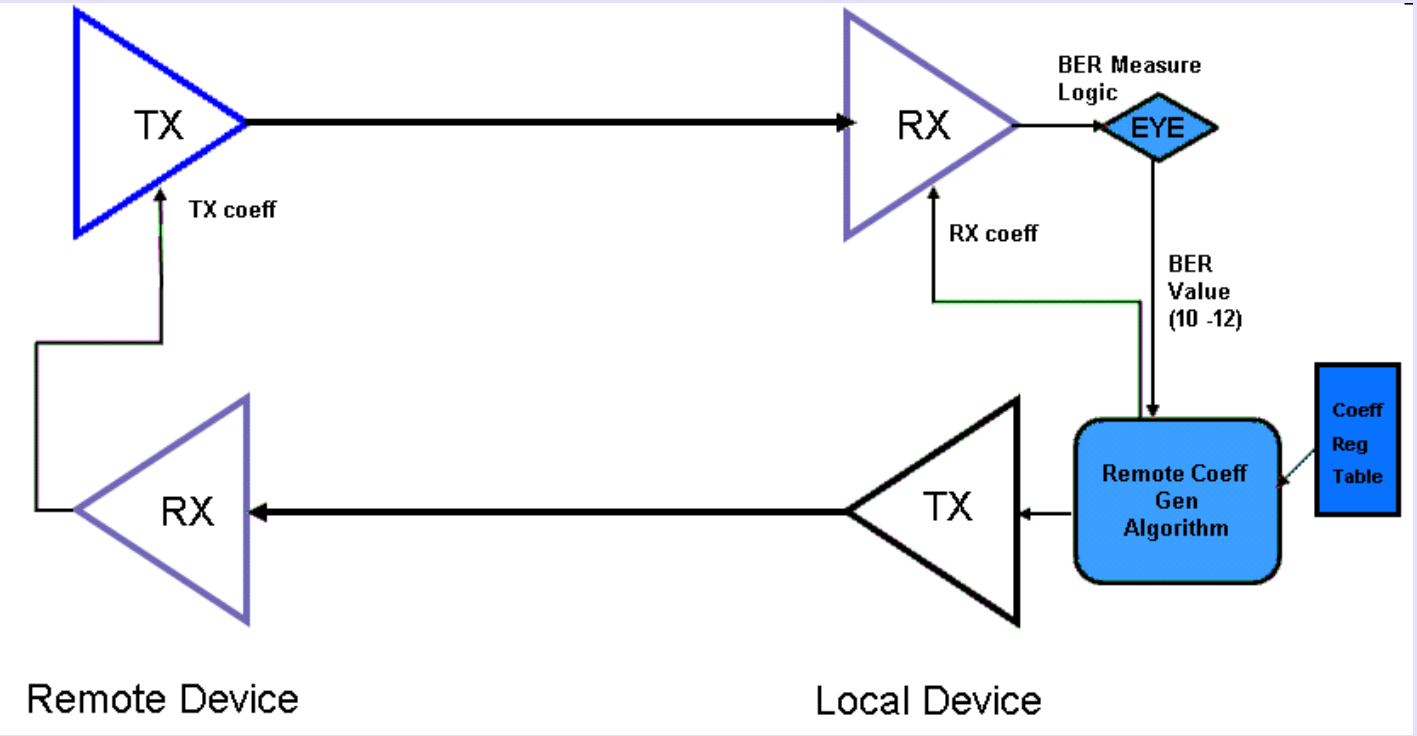
- Phase 2
  - ✓ Local device fine-tunes the remote device's transmit coefficients
  - ✓ EP functions as the master and is allowed up to 32 ms to refine the RC's transmit FIR coefficients and to adjust its own receiver
  - ✓ Achieved by transmitting embedded commands in the TS1 and receiving positive or negative acknowledgments from the RC
- Phase 3
  - ✓ Similar to phase 2
  - ✓ Remote device fine-tunes the local device's transmit coefficients
  - ✓ RC is now the master and is allowed 32 ms to refine the EP's transmit FIR coefficients and adjust its own receiver utilizing the TS-embedded signaling in phase 2

PCI Express Equalization Phase 2



Flexibility in PCI Express Equalization Implementation

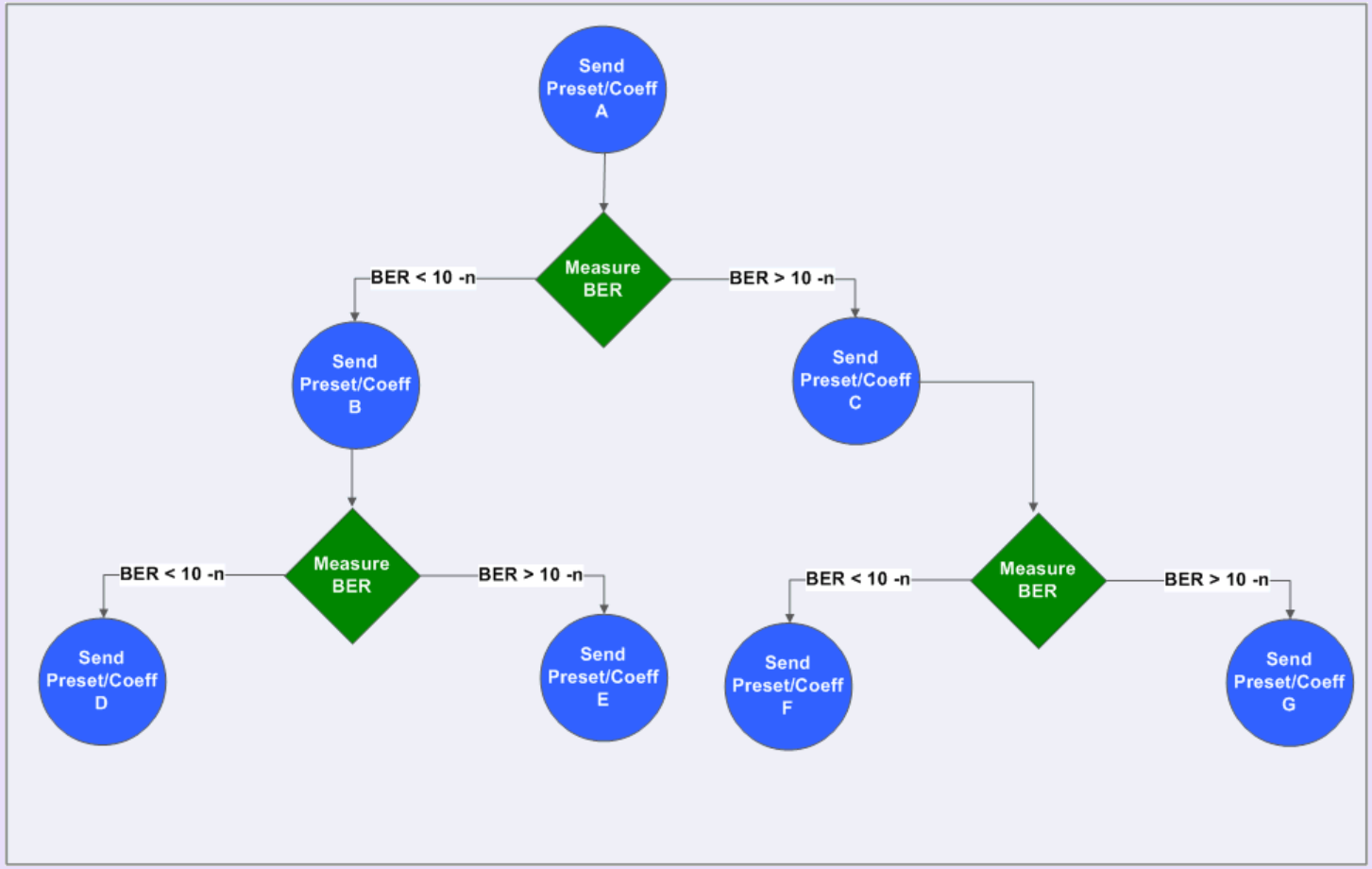
- Phase 2 – Local master, remote slave



Dynamic Equalization Control Digital IP (1/3)

- Configurable state machine
  - ✓ Implements remote coefficient generation algorithm
  - ✓ Each state utilizes a combination of specific Tx and Rx coefficient values which may be represented as presets or coefficients
  - ✓ Values are programmed on device initialization based on anticipated channel characteristics
  - ✓ State transitions determined by coefficient register table
  - ✓ Device converges rapidly to an optimum set of Tx coefficients that would result in a certain programmable BER value
  - ✓ Receiver's adaptive dispersion compensation engine (ADCE) could optionally then be utilized to further reduce the BER to a final programmable value of 10^{-12}, for instance

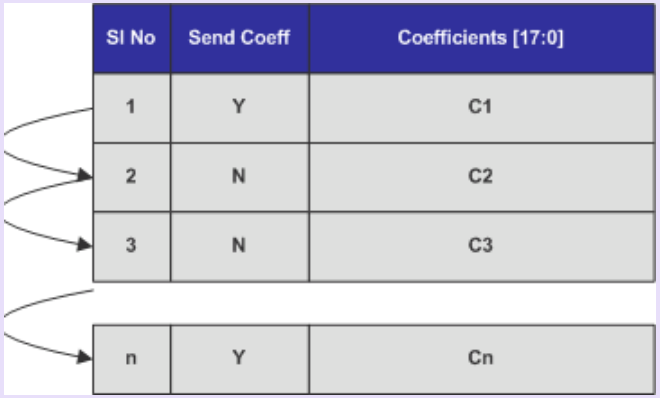
Equalization Control State Machine



Dynamic Equalization Control Digital IP (2/3)

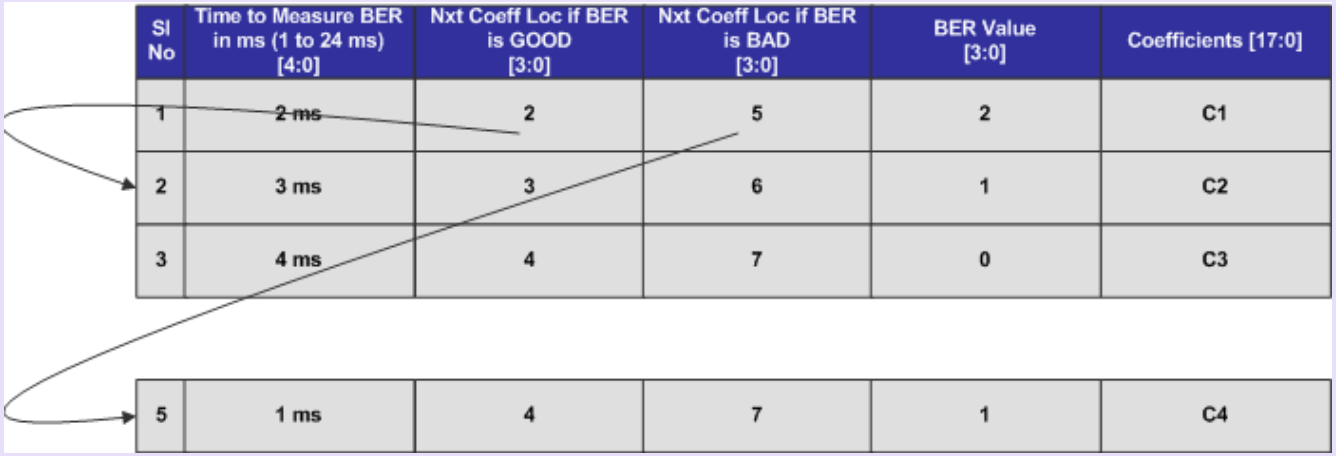
- Configurability to circumvent risk associated with emerging protocol like PCIe 3.0
  - ✓ Coefficient register table may be implemented as programmable linked list or sequential list to sweep coefficients rapidly
  - ✓ State machine may utilize coefficient values or presets only, in a simplified mode
  - ✓ Number of coefficient values supported is programmable

Sweeping Co-efficients



Sl No	Send Coeff	Coefficients [17:0]
1	Y	C1
2	N	C2
3	N	C3
n	Y	Cn

Sequential List



Sl No	Time to Measure BER in ms (1 to 24 ms) [4:0]	Nxt Coeff Loc if BER is GOOD [3:0]	Nxt Coeff Loc if BER is BAD [3:0]	BER Value [3:0]	Coefficients [17:0]
1	2 ms	2	5	2	C1
2	3 ms	3	6	1	C2
3	4 ms	4	7	0	C3
5	1 ms	4	7	1	C4

Linked List

Dynamic Equalization Control Digital IP (3/3)

- BER determination
  - ✓ Utilizes inference mechanism that provides timely indication of errors in incoming traffic without examining every bit
  - ✓ Hooks also provided to indirectly infer the BER by utilizing the PCIe-mandated eye height and eye width for the required target BER

Conclusion

- FPGA-integrated PCIe 3.0 protocol stack facilitates
  - ✓ Configurability
  - ✓ Extensibility to address a wide range of target applications
- Flexibility at all levels
  - ✓ Leverage inherent advantages of FPGA-based implementation
  - ✓ Address requirements of evolving protocol extensions with minimum risk

Patents pending on Transaction Layer Configuration/Extension and Equalization IP

Thank you for attending the
PCI-SIG Developers Conference 2011.

For more information please go to
www.pcisig.com