PCI Express® 2.0 Software and Configuration Updates

Wesley Shao

Sun Microsystems, Inc

Agenda

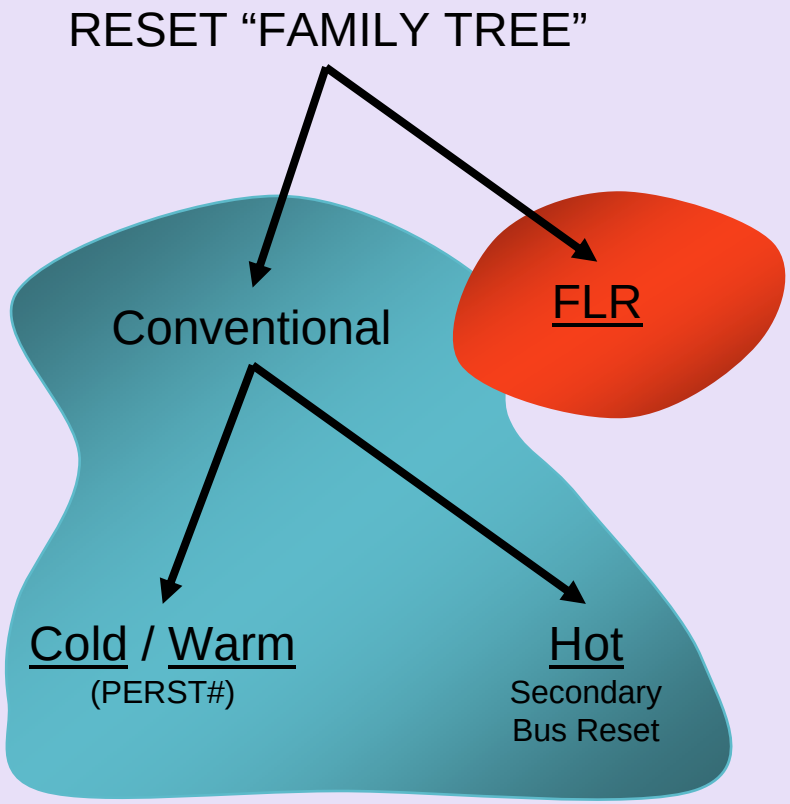
- Introduction
- FLR – Function Level Reset
- Link Speed Controls & Bandwidth Notification
- TCS – Trusted Configuration Space - removed
- ACS – Access Control Services
- ARI – Alternative Routing-ID Interpretation
- Completion Timeout
- Power budgeting and 5GT/s Link Controls

Introduction

- Introduced in Q4 2006
- Single largest change – 5GT/s Speed
  - ✓ Optional new capability
- Integrates all ECNs & Errata
- Errata highlights
  - ✓ Root Complex Event Collector Base Class Code conflict
  - ✓ Clarifications of uncommon error cases
  - ✓ Register bit clarifications on defaults, implementation requirements
  - ✓ Interrupt Disable bit consistency w/conventional PCI
- ECNs **Today's Focus**

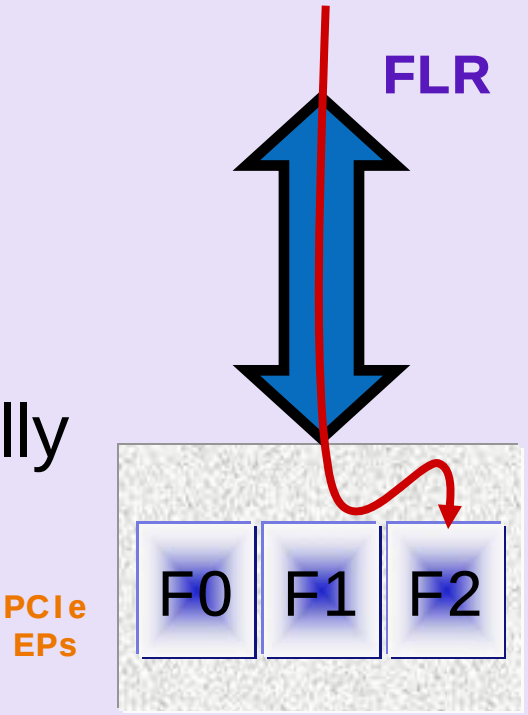
Function Level Reset (FLR)

- Returns Function specific registers to initialized state
- Existing resets may (but not required to) reset function internals
- Preserved configurations
  - ✓ Clock, Link, Power and Power Management
  - ✓ Sticky bits
- Stops and makes device quiescent
  - ✓ BME, MSI
  - ✓ De-asserts INTx
- Virtualization friendly



Function Level Reset (FLR)

- Endpoints only
  - ✓ Legacy
  - ✓ Native
  - ✓ Integrated
- Register interface simple
- Implementation & effects potentially complex
  - ✓ Resets internal function-specific state
- Not all architected registers are reset
  - ✓ Hardware initialized (HwInit), BIOS set, etc.



Link Speed Control & Bandwidth Notification

- Link Speed Control
  - ✓ By default, hardware automatically trains to highest common speed
    - Software can set upper bound
    - Hardware can always limit speed for link reliability reasons
  - ✓ By default, hardware is permitted to change speed autonomously for other purposes
    - Part of power management
    - Software can disable it
- SW controllable entering/exiting Compliance
- Bandwidth Notification
  - ✓ Notify software speed/width changes
    - Due to hardware autonomous link training
    - Required for >x1 or multiple link speed downstream ports

Link Speed Control & Bandwidth Notification

Register	Bit	Width	Speed
Link Capability Register	Link Bandwidth Notification	★	★
	Supported Link *Speeds*		★
Link Control Register	Link Autonomous Bandwidth Interrupt Enable	★	★
	Link Bandwidth Management Interrupt Enable	★	★
	Hardware Autonomous *Width* Disable	★	
	Retrain Link	★	★
Link Status Register	Link Autonomous Bandwidth Status	★	★
	Link Bandwidth Management Status	★	★
	Current Link *Speed*		★
Link Control 2 Register	Hardware Autonomous *Speed* Disable		★
	Target Link *Speed*		★

Access Control Services (ACS)

- Peer-to-peer traffic can be
  - ✓ Disabled
  - ✓ Isolated
- Applicable to switch and multi-function devices
- Configuration should be consistent across fabric
- Error handling
  - ✓ Handling of Completer Abort needs to be enhanced
- ACS implementation is *Optional*

Access Control Services (ACS)

	Root Ports	Switches		Devices		Condition
		Downstream Ports	Upstream Ports	Single Function	Multi-Function	
Source **V**alidation Checks requester bus# within aperture	Required	Required	0	0	0	
Translation **B**locking Denies ATS request	Required	Required	0	0	0	
P2P **R**equest Redirect Redirects P2P requests upstream or RP reflection	*Required*	Required	0	0	*Required*	P2P
Completion Redirect Redirects Non-RO completion upstream or RP reflection	*Required*	*Required*	0	0	*Required*	P2P Request Redirect
Upstream Forwarding Same as P2P except being target itself	*Required*	Required	0	0	0	Redirected Request Validation
P2P **E**gress Control Blocks P2P on per root/downstream port basis	Optional	Optional	0	0	Optional	
Direct **T**ranslated P2P Direct path for translated P2P transactions	*Required*	Required	0	0	Required	ATS + P2P

Required: Required if the device implements the ACS option

Alternative Routing-ID Interpretation (ARI)

- Eliminates Device ID
- Number of Functions increased from 8 to 256
- Requires ARI enabled downstream ports
  - ✓ Switches and Root ports
- Multi-function devices vs ARI devices
- ARI cap, extended function discovery (link list)
- MFVC arbitration with Function Groups
- ACS P2P egress control + Function Groups
- OS and software requirements
  - ✓ Firmware should not enable ARI (OS may not be ARI aware)
  - ✓ Configuration, PM, error handling, and INTx complications

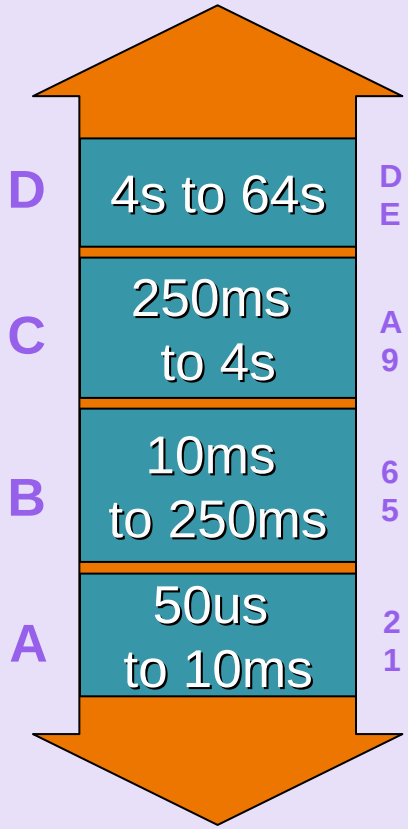
Alternative Routing-ID Interpretation (ARI)

	Category	Multi-function Devices	ARI devices
Configuration	Max Payload Size	Must be the same	Function 0
	Ext-function enumeration	N/A	ARI-aware OS
	Common clock	Must be the same	Function 0
Power Management	L1 entry	All enter non-D0 D-states	Non-D0 or D0$_{uninitialized}$
	ASPM	Most active D-state	Function 0
	Clock PM	Independent	Must be the same
	Enable Clock PM	All must allow	Function 0

Software visible

Completion Timeout

- Required: Architected disable bit
  - ✓ 'Turns off' timeout
  - ✓ Not to be used in normal operation
- Optional: Completion Timeout value programmability
  - ✓ Supported ranges
    - Device Capabilities 2 Register
    - Contiguous selectable ranges
      - A, B
      - AB, BC
      - ABC, BCD
      - ABCD
  - ✓ Two choices within each selectable range
    - Device Control 2 Register



Power Budgeting and 5GT/s Link Controls

- Power budgeting
  - ✓ Optional, via Power Budgeting Capability
  - ✓ 250W, 275W, 300W addition
- Link Controls (Link Control 2 Register)
  - ✓ Selectable De-emphasis
    - – Downstream port only
      - 1 -3.5dB
      - 0 -6dB
  - ✓ Compliance features
    - – Transmit Margin, Enter Modified Compliance
    - – Compliance SOS, Compliance De-emphasis

Thank you for attending the
PCI-SIG Developers Conference 2007.

For more information please go to
www.pcisig.com

PCI Express® 2.0 Software and Configuration Updates

Wesley Shao

Sun Microsystems, Inc