PCIe® 2.0 Server Validation Challenges

Will Atherton

Tools / Test Strategy Lead

IBM

Topics

- Servers
- Areas of Testing
- PCI Express® 2.0 Updates
- RAS
- Recommendations

Server

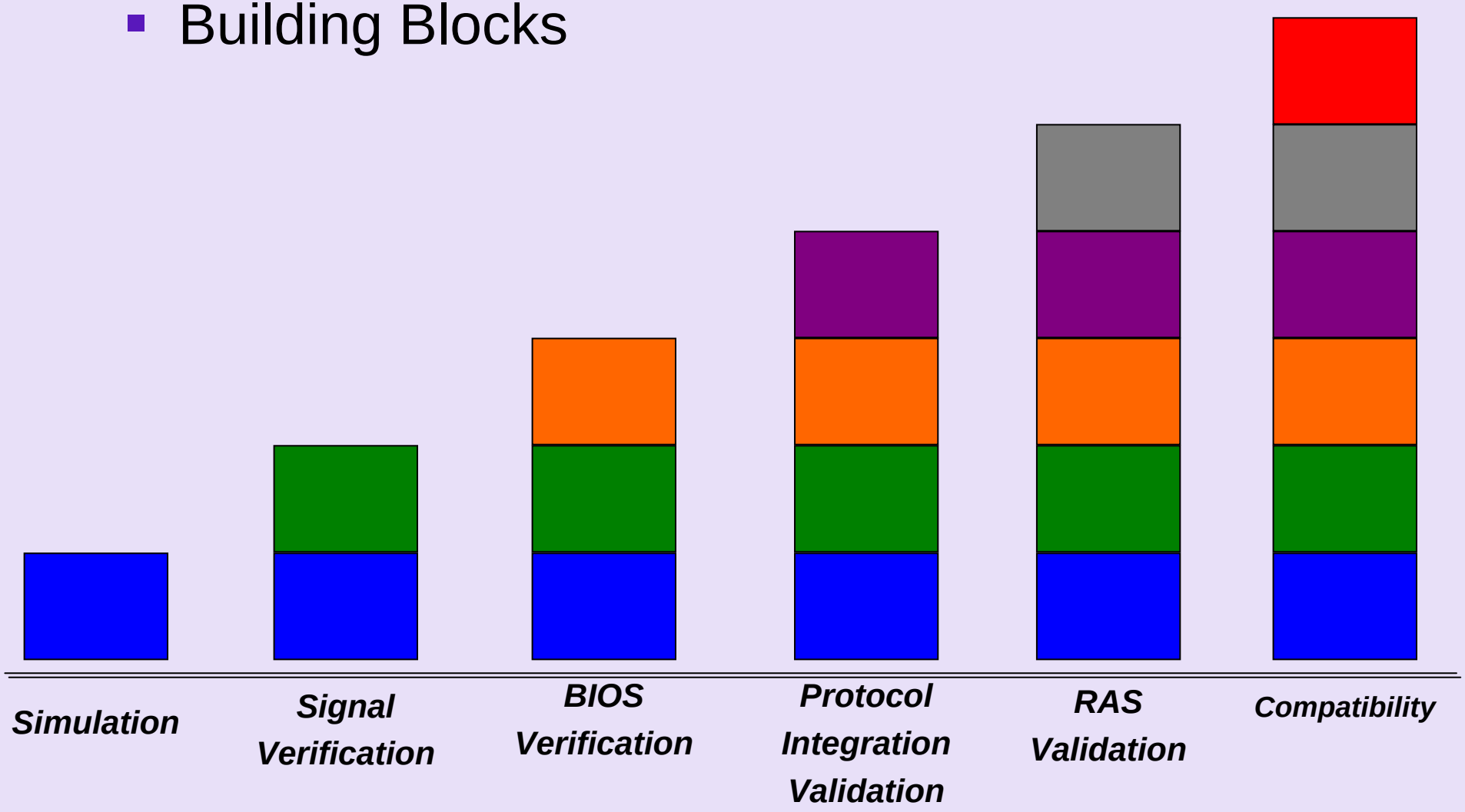
Server?

Servers

- A server is a pool of high-end resources (memory, disk, processor, I/O, etc..) for clients / customers
- Servers range from uni-processor to 32-processor Multi-Node NUMA (Non Uniform Memory Addressing) scaleable servers
- Requires High RAS (Reliability, Availability, & Serviceability)
  - ✓ Low tolerance for persistent recoverable and non recoverable events
- All servers need high performance I/O enablement
  - ✓ Compatibility and compliance are key!

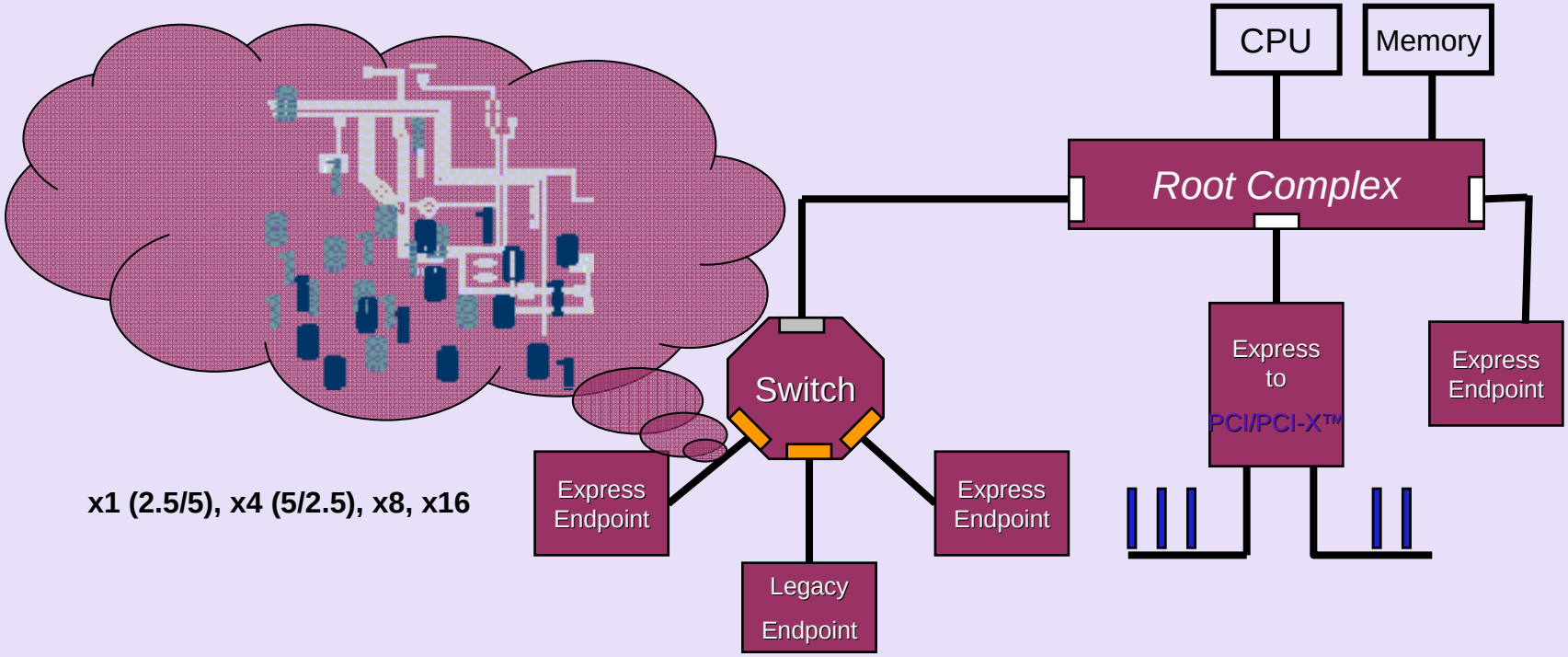
Areas of Testing

- Building Blocks



Areas of Testing

- Simulation
  - ✓ Pre-Hardware / Logic testing
  - ✓ Can simulate needed dynamic capabilities of PCIe 2.0
  - ✓ Can be for hardware or testing tools



Areas of Testing

Signal Verification

- Physical Bring-up
- Signal integrity measured
  - ✓ Methods
    - PCI-SIG® - PCI Express Electrical Test Fixtures (eye diagrams)
      - New Registers (Link Control 2 Register)
        - Enter Compliance, Enter Modified Compliance, etc..
      - Compliance Load Board (CLB)
      - Compliance Base Board (CBB)
    - Custom Analysis Tools
      - Create soft Insertion points for verification
      - Allows check point verifications
      - Trigger unique events (2.5/5)

Areas of Testing

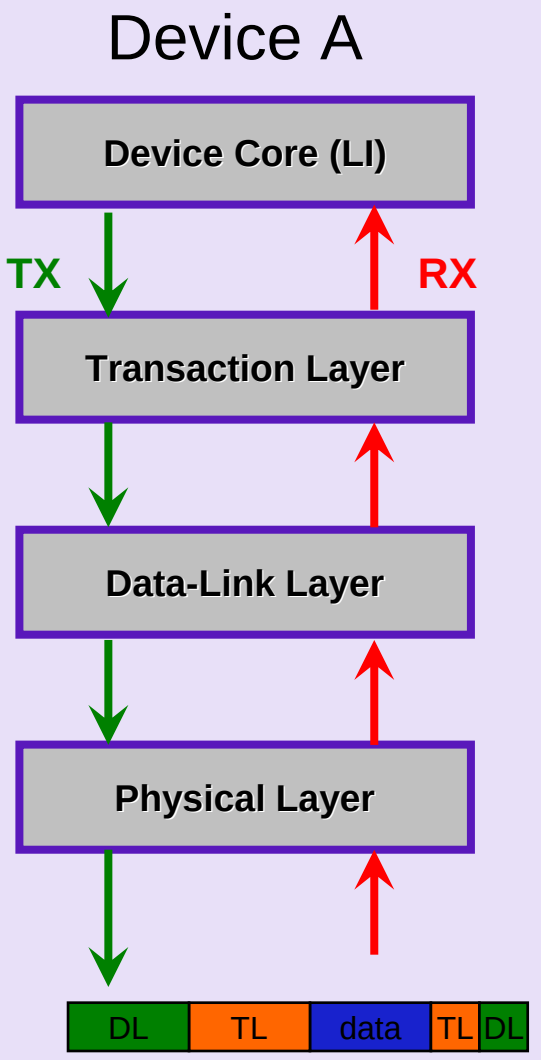
BIOS

- System / Device Initialization
  - ✓ Methods
    - PCI-SIG - PCI Express Platform BIOS test
      - Uses Protocol Test Card + External Software
      - Simulates various devices
        - Good events
        - Bad events
    - Custom Tools
      - Specific BIOS call testing
        - Software
      - Specific BIOS handling
        - Hardware
  - ✓ Enable static PCI Express values
    - Slot by slot Configurations
      - Lane width / Frequency

Areas of Testing

Protocol

- System / Device Interaction
  - ✓ Methods
    - – Custom Tools
      - • Software / Hardware
        - • Fabric Management
      - • Extended verification
        - • Vendor Specific
  - ✓ Different Levels of the Protocol
    - – Several layers for validation
    - – PCIe 2.0 focus on Physical
  - ✓ Compatibility / Add in-cards
    - – Several approaches
      - • Integrators list
      - • OS Selection / Software stack



PCI Express 2.0 Updates

New Speed and Lane Width Controls

- 5.0 GT/s and 2.5 GT/s Link speeds supported
  - ✓ 5.0 GT/s Not required
- Dynamic ***link speed management*** allows developers to control the speed at which the link is operating
  - ✓ Requires both system and device to insure capabilities and reporting
- Link bandwidth ***notification alerts*** platform software (operating system, device drivers, etc) of changes in link speed and width
  - ✓ Gives you verification points and enhanced RAS

PCI Express 2.0 Updates

- New card and system Interoperability
  - ✓ Allowed → now required
- All PCI Express add-in cards ***must be able to negotiate and operate in all smaller Link widths*** from the full Link width down to x1. The x2 and x12 Link widths are optional.
- The upstream PCI Express components ***on a system board must be able to negotiate and operate in all smaller Link widths*** from the full Link width down to x1. The x2 and x12 Link widths are optional.

CEM Spec 1.1

Table 6-2: Card Interoperability

Slot / Card	x1	x4	x8	x16
x1	Required	Required	Required	Required
x4	No	Required	Allowed	Allowed
x8	No	No	Required	Allowed
x16	No	No	No	Required

CEM Spec 2.0

Table 6-2: Card Interoperability

Connector / Card	x1	X4	X8	x16
x1	Required	Required	Required	Required
x4	No	Required	Required	Required
x8	No	No	Required	Required
x16	No	No	No	Required

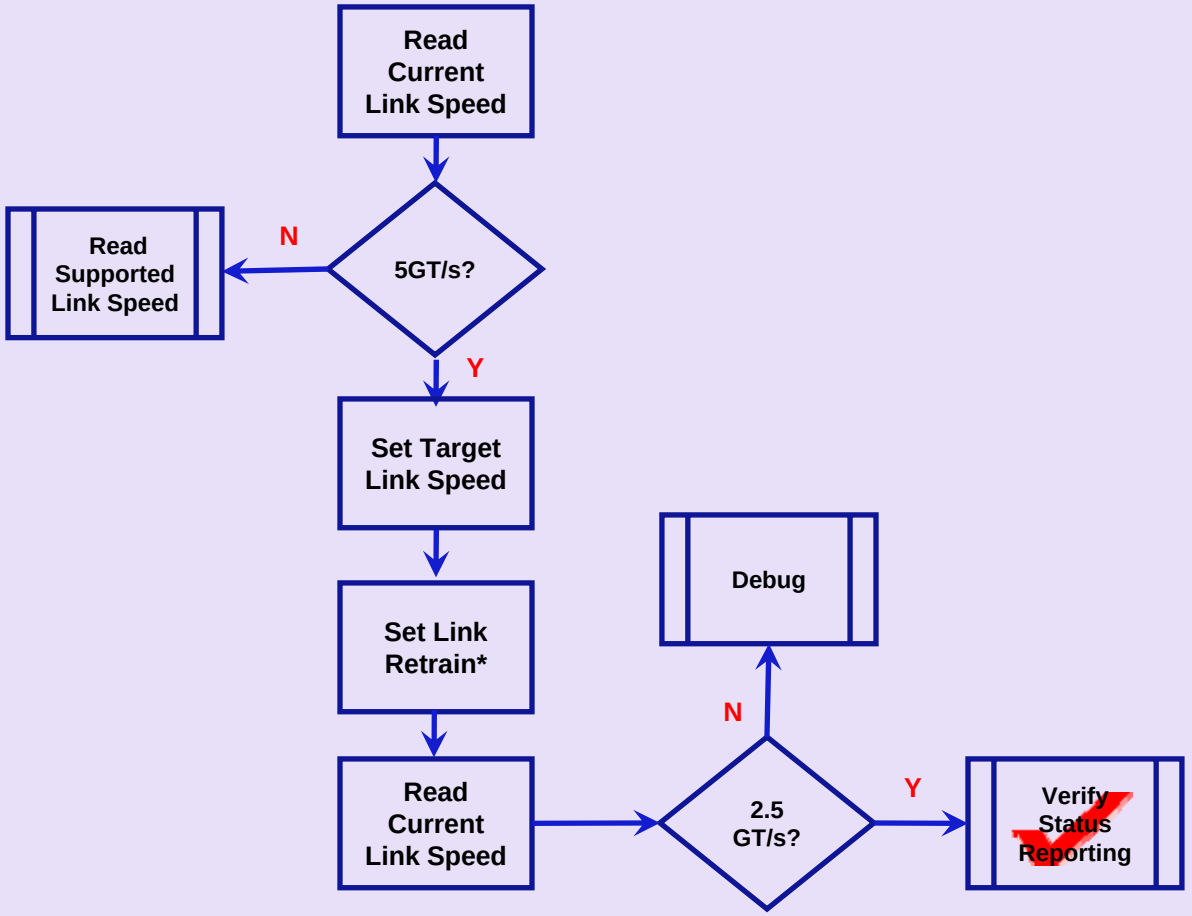
Slot bandwidth Testing

Slot bandwidth testing via ***Software***

- ✓ Use a device that support 5GT/S
- ✓ Use new and existing registers!

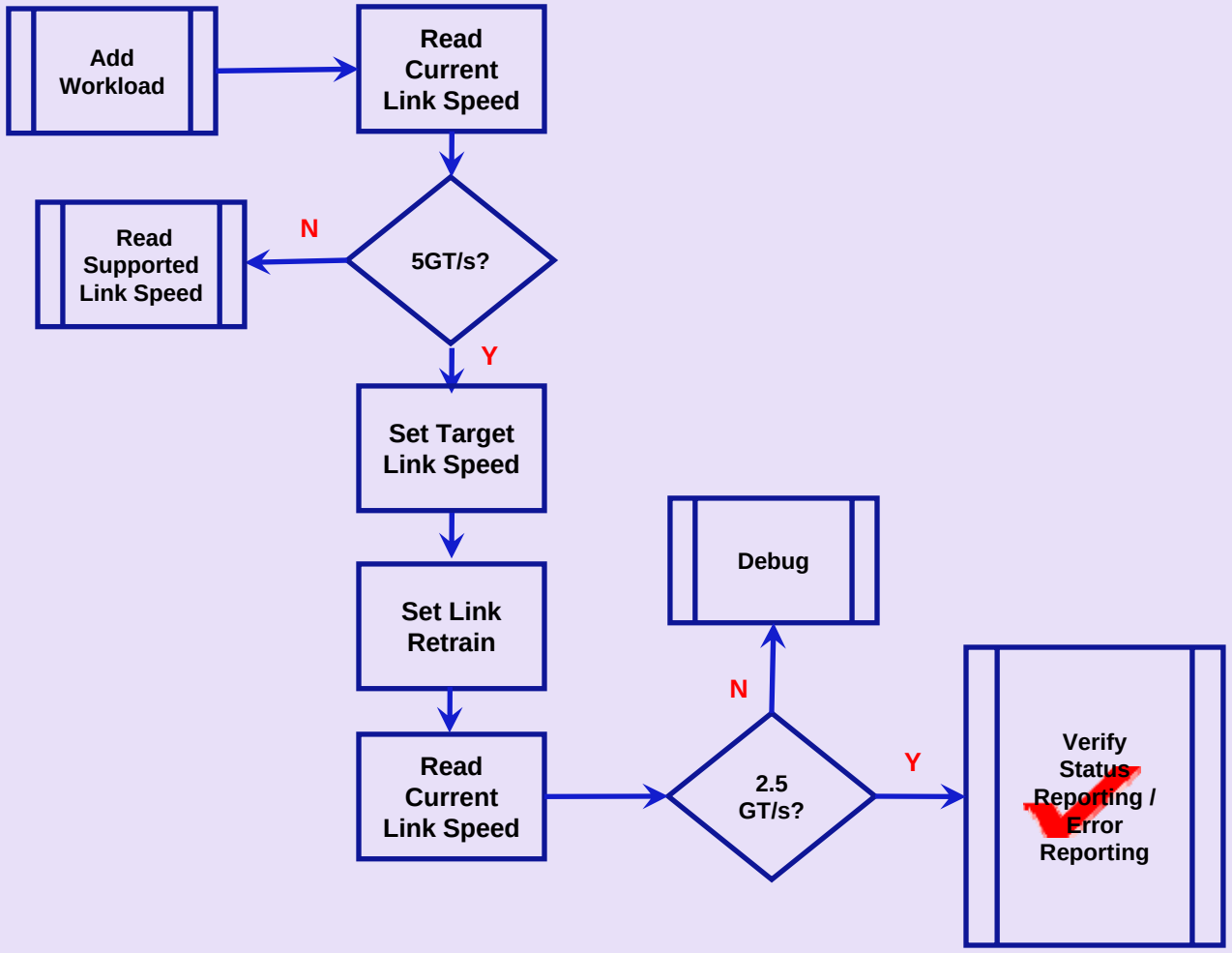
- Link Capabilities Register
  - ✓ Supported Link Speeds
  - ✓ Maximum Link Width
- Link Status Register
  - ✓ Current Link Speed
  - ✓ Negotiated Link Width
  - ✓ Link Bandwidth Management Status
  - ✓ Link Autonomous Bandwidth Status
- Link Control Register
  - ✓ Retrain Link bit
  - ✓ Link Disable
- Link Control 2 Register
  - ✓ Target Link Speed

Slot bandwidth Testing Flow



*** You could use Secondary Bus Reset (from Bridge Control Register) on the up stream port but you must use caution.**

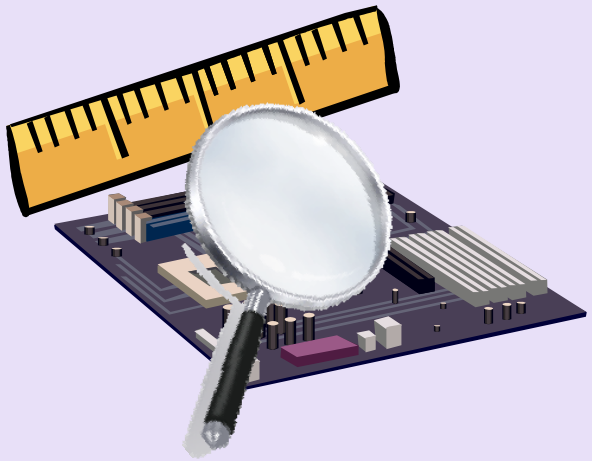
Slot bandwidth Testing Flow



Dynamic Slot Link Width Testing

Dynamic Slot Link width testing via ***Software***

- ✓ Use a device that support 5GT/S
- ✓ Use new and existing registers?



- Link Capabilities Register
  - ✓ Maximum Link Width
- Link Status Register
  - ✓ Negotiated Link Width
  - ✓ Link Bandwidth Management Status
  - ✓ Link Autonomous Bandwidth Status (covers width too)
- Link Control Register
  - ✓ Retrain Link bit
  - ✓ Link Disable
- Set Link Width???

Dynamic Slot Link Width Testing

No defined standard “Registers”

…Then how can it happen?

It’s only occurs during the ***“Link Initialization and Training”*** initiated by a device or System

Dynamic Slot Link Width Testing

It is VERY unlikely that **systems** will not implement this!

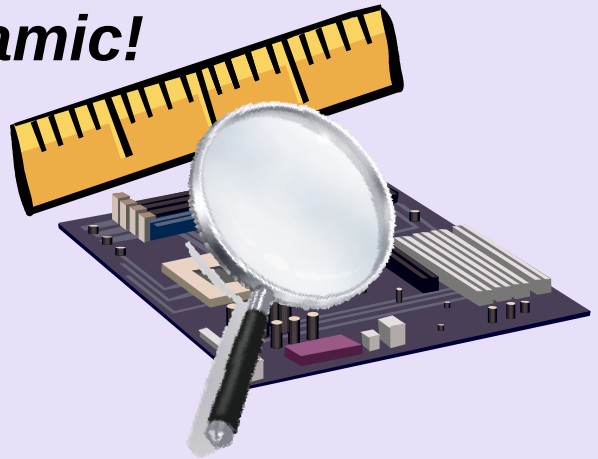
- If you have to support…you need a way to test and benefit
- Early indication are through bridge registers and BIOS

Card vendors could expose this as well

- Could be used for DIAGS, performance, and testing
- Card vendors need to verify they can handle as well

Static Slot Link Width Testing

Static has the Same Problem as Dynamic!



System Vendors

- Could enable via BIOS
- Could enable via board jumper
- Could enable via “bridge register” warm reset

Card Vendors

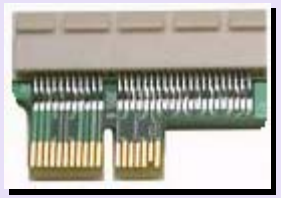
- Could enable via jumpers
- Could enable via Firmware / Device Driver
- Could enable via Diags

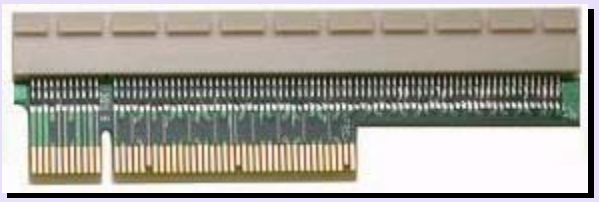
Static Slot Link Width Testing

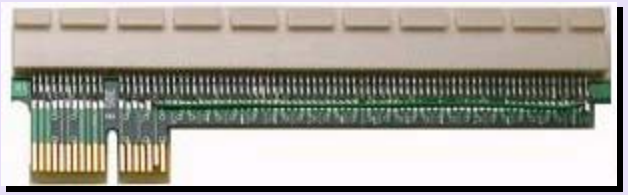
Alternate approach!

Lane Converters / Bus Extenders / Riser

- Must use caution as not to create new problems







PCI Express 2.0 Continued

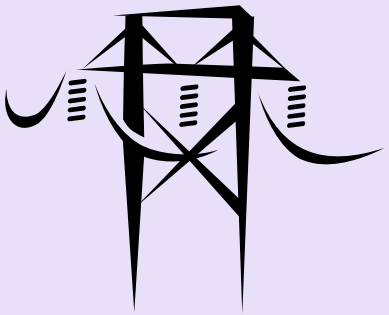
- Function-level reset provides an **optional** mechanism to reset functions within a multifunction device
  - ✓ Key new feature for resetting Multifunction device
  - ✓ Future IOV usages
  - ✓ RAS / Error injection testing
  - ✓ **Test device usage and system test**

- Completion timeout control allows developers to define a required disable mechanism for transaction timeouts
  - ✓ Device Capabilities 2 Register
    - **Completion Timeout Disable Supported**
    - **Completion Timeout Value**
    - **Completion Timeout Disable**
  - ✓ Internal fabric testing
  - ✓ **Key for debugging and testing tools**

PCI Express 2.0 Continued

Most servers to not support full Power requirements for all slots

- Very expensive to support max power
  - ✓ Cooling
  - ✓ Power supplies
- New testing to utilize power registers
  - ✓ Real vs. Reported Power limits
  - ✓ Does the system over commit power does the device scale well
  - ✓ Does the device not power on or can it be scaled



Methods for testing

- Loader cards / Hardware exercisers
  - ✓ Set different values to insure system resiliencies
- System policies and interoperability
  - ✓ Negative testing
  - ✓ How do real devices react?

As system become larger and customers push “greener” systems, power control is a key area

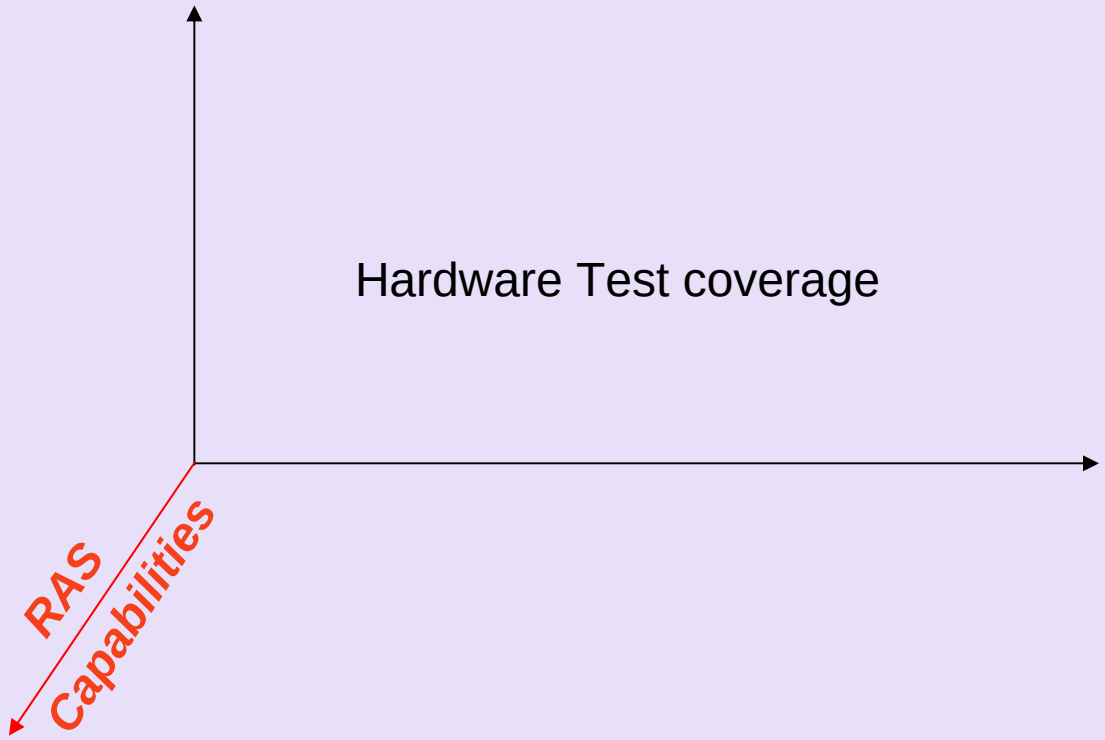
RAS

Reliability, Availability, & Serviceability

RAS - New Challenges

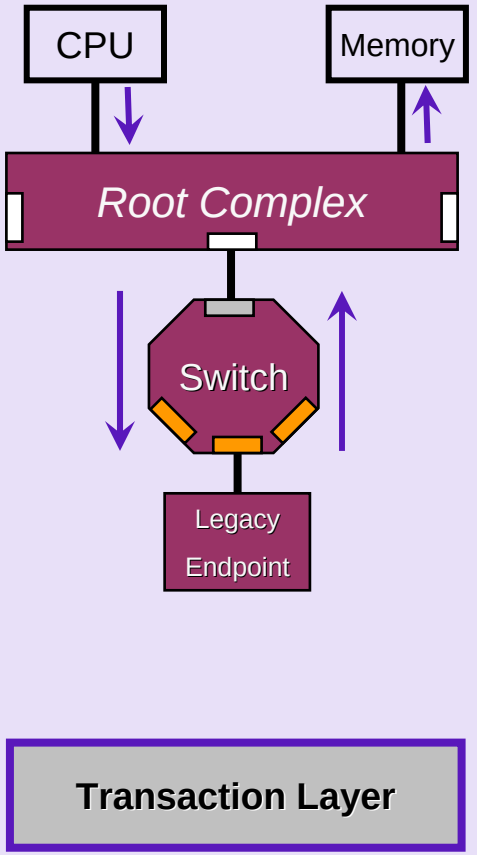
RAS Capabilities

- ✓ Error Generation
- ✓ Error Reporting
- ✓ Error Recovery
- ✓ Provisioning
- ✓ Optimization
- ✓ Validation



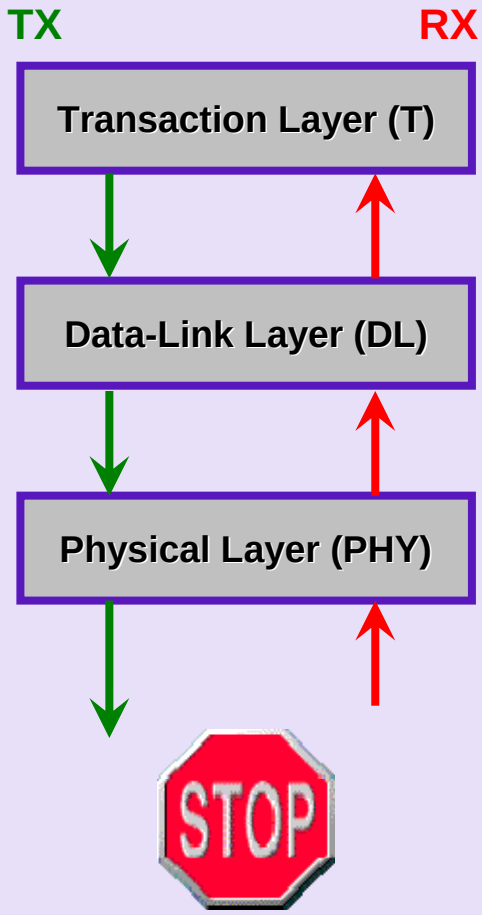
RAS - New Challenges

- RAS Capabilities
  - ✓ Run negative testing
    - – Uncorrectable Non-Fatal
      - • Unexpected Completion
      - • Unsupported Request
      - • Completion Abort
      - • Completion Time-out
      - • Poisoned TLP Received
      - • ECRC Failed
  - ✓ Verify register status
    - – Unsupported Request Error Status
    - – ECRC Error Status
    - – Malformed TLP Status
    - – Link Autonomous Bandwidth Status
    - – Link Bandwidth Management Status



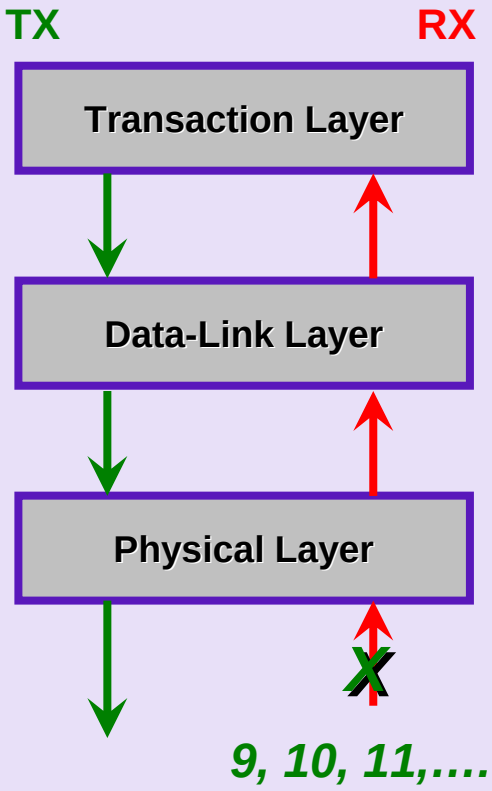
RAS - New Challenges

- RAS Capabilities
  - ✓ Run Negative Testing
    - – Uncorrectable Fatal
      - Flow Control Protocol Error (T)
      - Malformed TLP (T)
      - Training Error - Undefined (PHY)
      - DLL Protocol Error (DL)
      - Receiver Overflow (T)
  - ✓ Verify register status
    - – Malformed TLP Status
    - – Flow Control Protocol Error Status
    - – Receiver Overflow Status
    - – Data Link Protocol Error Status



RAS - New Challenges

- RAS Capabilities
  - ✓ Thresholds / Dynamics
    - – Verify (exceed)
      - • Non-Fatal / Correctable
    - – Action taken
      - • i.e. Retraining?
    - – Reporting / Timing
      - • Location and information
      - • Performance and latency
  - ✓ Systems Responses
    - – Take the system down
    - – Device not powered on
    - – Recovery



RAS - New Challenges

Must support Legacy and Express

- Insures Software and hardware compatibility
  - ✓ Legacy
  - ✓ New dynamic capabilities
- Is advanced error reporting on?
  - ✓ Not a requirement …
  - ✓ New status registers can give you new RAS insight
- No power in legacy devices

Recommendations

- Recommendations
  - ✓ Signal / Simulation Verification feed your testing process
    - Most stressful patterns in different subsystem
    - New and future dynamic features need early validation
  - ✓ Dynamic testing can be offset with software testing
    - Look at device testing (compatibility test)
    - Status register are key
  - ✓ Touch each layer of the protocol
    - But don’t get lost in the combinations
  - ✓ Compatibility is a joint partnership
    - Out of the box testing
    - Can’t replace real hardware interactions
  - ✓ RAS is key to success
    - Detection is only one aspect
    - Positive and Negative testing insure overall quality
    - Reporting and verification helps you in development and in the field

Recommendations

- Recommendations
  - ✓ No easy straight solution
    - – Most solutions are a combination
      - • Software and Hardware
    - – Can use existing equipment / software
      - • Register and ….
  - ✓ Compliance allows a common measurement
    - – Helps reduce chances of hardware churn
    - – Allows easy comparison
    - – Early compatibility testing
    - – Open collaboration and sharing
    - – Direct ability to test may dynamic configurations
  - ✓ Compare Hardware / Software Test Vehicles
    - – Many are on the market
      - • All have + and –
    - – Some can lock you in
    - – Cost can be a barrier

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

For more information please go to
www.pcisig.com

PCIe 2.0 Server Validation Challenges

Will Atherton

Tools / Test Strategy Lead

IBM