SDC

STORAGE DEVELOPER CONFERENCE

SNIA SANTA CLARA, 2017

PCI Express® Technology: The Ubiquitous I/O Interconnect, Now in its Fifth Generation

Dr. Debendra Das Sharma

Member, PCI-SIG Board of Directors

Sr Principal Engineer and Director of I/O Technology and Standards

Intel Corporation

Agenda

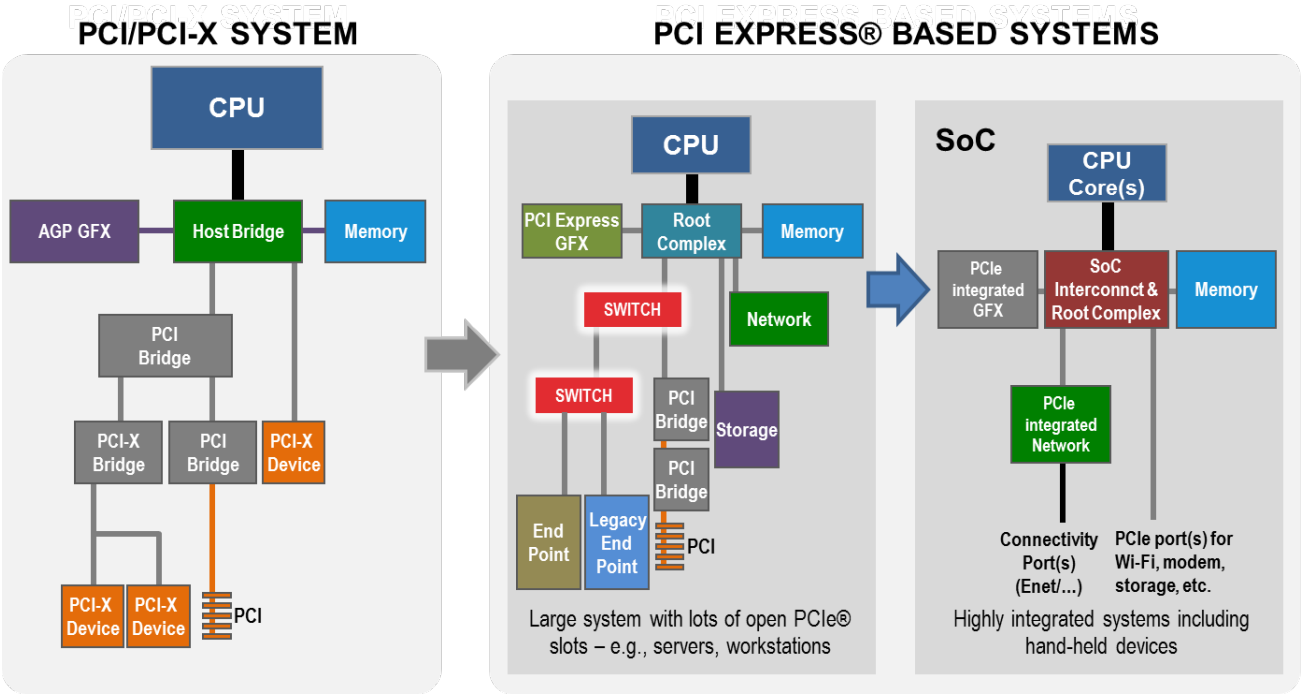
- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

Evolution of PCIe® Technology

- Peripheral Component Interconnect (PCI) started as bus-based PC interconnect in 1992
  - Evolved through width/speed increases
- Moved to link-based serial interconnect with full-duplex differential signaling with PCI Express® (PCIe®) with backwards compatibility for software
  - Currently in fifth generation, with bandwidth doubling every generation
- Evolution from PC to HPC, servers, clients, hand-held, and Internet-of-Things usage over three decades

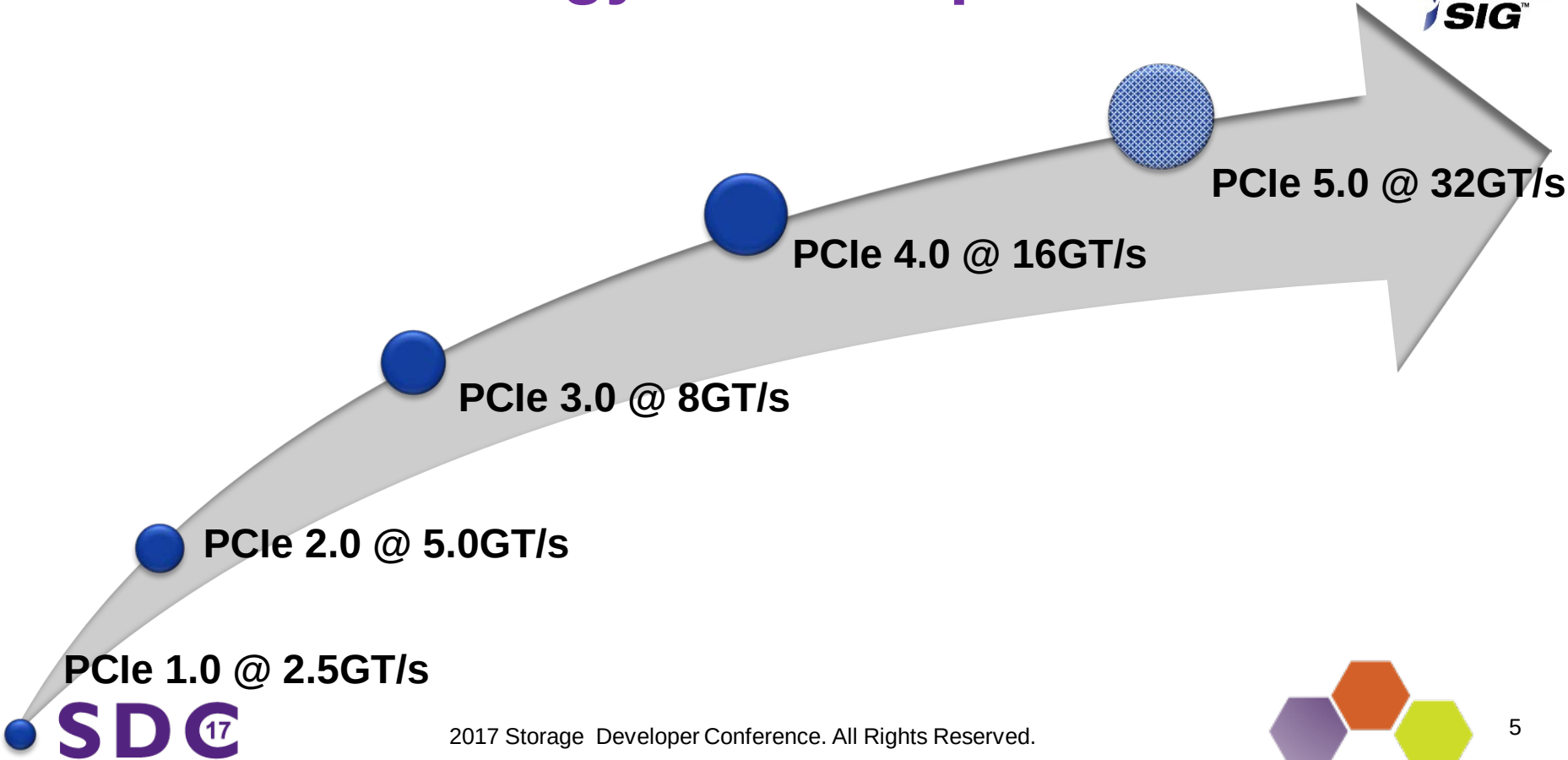
Five generations of PCI Express architecture with backwards compatibility!
Doubling the data rate every generation and going strong

Evolution of PCI/PCIe® Technology

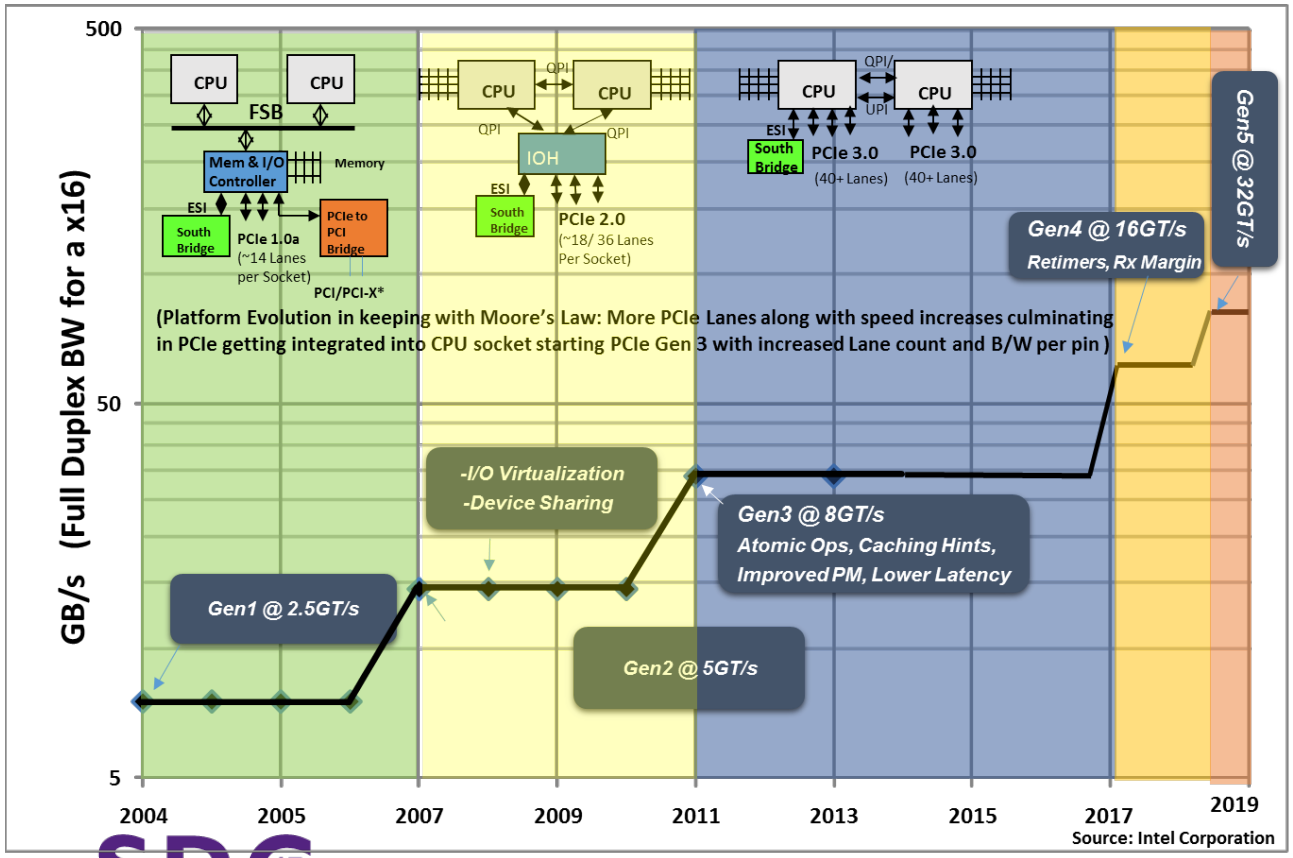


Relevant through evolution of platforms across multiple market segments

PCIe® Technology Roadmap

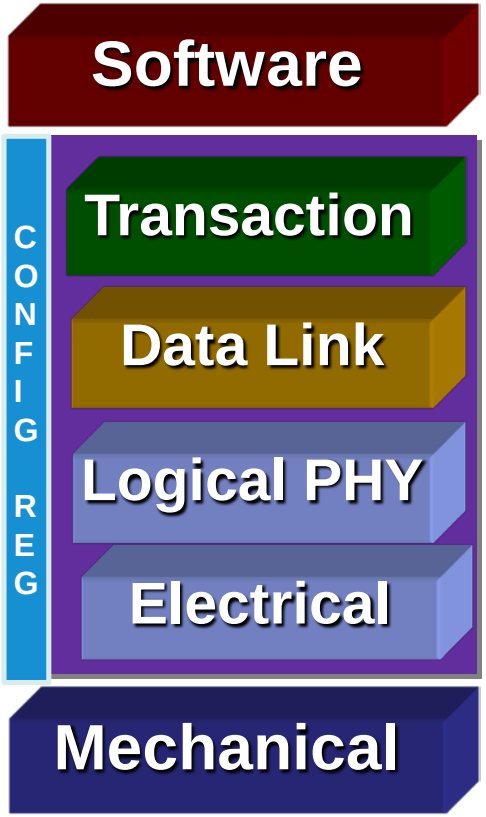


Evolution of PCIe® in Platforms



- *Continuous Improvement: Data Rate, Protocol enhancements, Power enhancements, Form-factor, and Usage Models*
- *Doubling Bandwidth & Improving Capabilities Every 3-4 Years*
- *Relevant through evolution of platforms across multiple market segments*

PCIe® Architecture Layering for Modularity and Reuse



⇦ PCI compatibility, configuration, driver model
⇦ PCIe architecture enhanced configuration model

⇦ Split-transaction, packet-based protocol
⇦ Credit-based flow control, virtual channels

⇦ Logical connection between devices
⇦ Reliable data transport services (CRC, Retry, Ack/Nak)

⇦ Physical information exchange
⇦ Interface initialization and maintenance

⇦ Market segment specific form factors
⇦ Evolutionary and revolutionary

Agenda

- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

Power Efficient Performance

- Delivers scalable performance
  - Width scaling: x1, x2, x4, x8, x12, x16, x32
  - Frequency scaling: Five generations
    - 2.5 and 5 GT/s w/ 8b/10b; 8 and 32 GT/s with 128b/130b encoding

Item	PCIe® 3.0	PCIe® 2.0
Line Speed [Gbps]	8	5
PHY Overhead	128/130, 1[GB/s]	8/10, 500[MB/s]
Active Power [mW]	60 (L0)	46 (L0)
Standby Power [mW]	0.11 (L1.2)	0.11 (L1.2)
MB/mJ (higher = better)	14-18	8-12

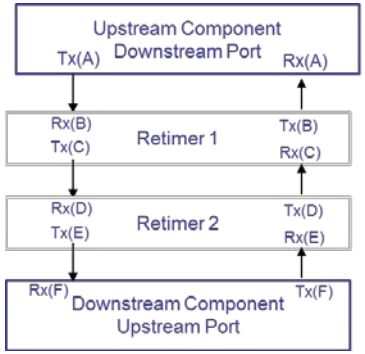
Source: Intel Corporation (IDF, Sept 15)

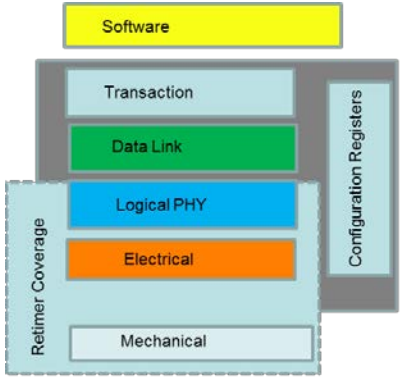
- Low power (active/idle)
  - Rich set of Link (L0s, L1, L1-substates, L2/ L3) and device (D0, D1, D2, D3_hot/cold) states
  - Platform-level power optimization hooks: Dynamic Power Allocation, Optimized Buffer Flush Fill, Latency Tolerance Reporting
  - Active power – 5pJ/b, Standby power: 10 uW/Lane*
- Vibrant ecosystem with IP providers

** http://news.synopsys.com/2015-05-21-Synopsys-Announces-Industrys-Lowest-Power-PCI-Express-3-1-IP-Solution-for-Mobile-SoCs*

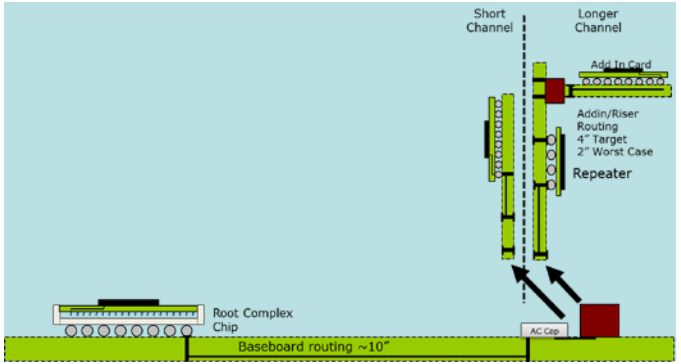
PCIe 5.0 Architecture at 32GT/s

- Backwards compatible with prior generations
- 128b/130b encoding (similar to PCIe 3.0 and PCIe 4.0 architectures)
- Connector improvements to reduce cross-talk and improve insertion loss at 8G/ 16G Nyquist
- 2 connector 20” server PCIe topology needs either re-timer or ultra low-loss PCB to operate at 16 or 32 GT/s
- Retimer part of base specification





Source: Intel Corporation



Source: Intel Corporation

Agenda

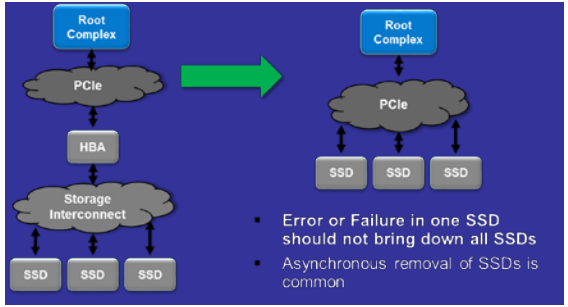
- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

RAS Features

- PCIe® architecture supports a very high-level set of Reliability, Availability, Serviceability (RAS) features
  - All transactions protected by CRC-32 and Link level Retry, covering even dropped packets
  - Transaction level time-out support (hierarchical)
  - Well defined algorithm for different error scenarios
  - Advanced Error Reporting mechanism
  - Support for degraded link width / lower speed
  - Support for hot-plug

DPC/ eDPC Motivation and Mechanism

- Recently added (enhanced) Downstream Port Containment (DPC and eDPC) for emerging usages
- Emerging PCIe usage models are creating a need for improved error containment/recovery and support for asynchronous removal (a.k.a. hot-swap)
- Defines an error containment mechanism, automatically disabling a Link when an uncorrectable error is detected, preventing potential spread of corrupted data
- Reporting mechanism with Software capability to bring up the link after clean up
- Transaction details on a timeout recorded (side-effect of asynchronous removal)
- eDPC: Root-port specific programmable response to gracefully handle DPC downstream

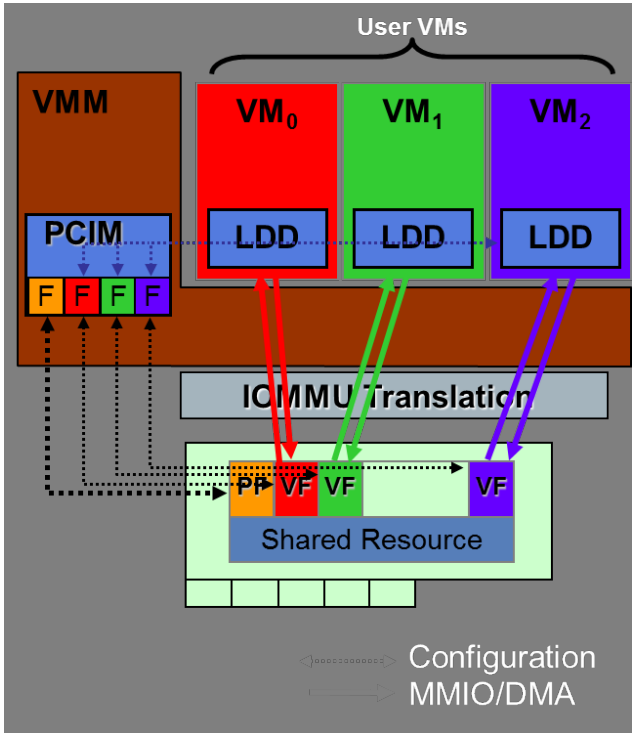


Agenda

- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

I/O Virtualization

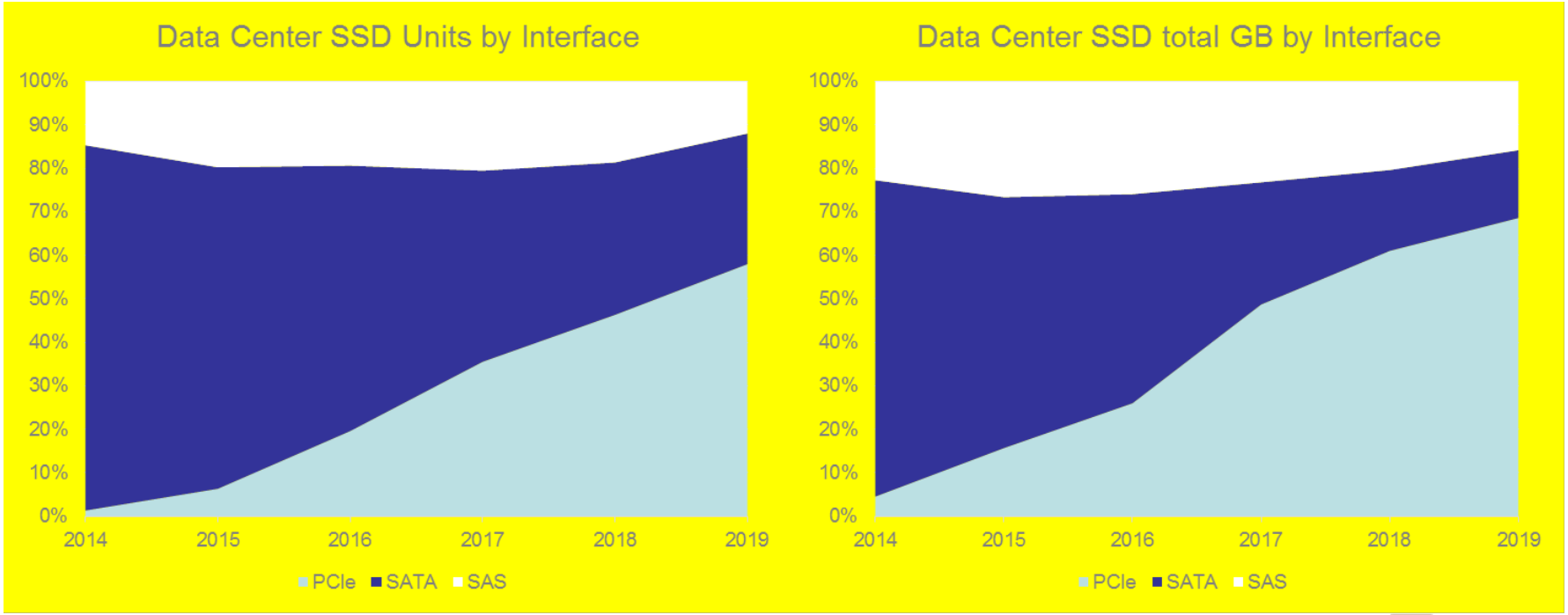
- Reduces System Cost and power
- Single Root I/O Virtualization Specification
  - Released September 2007
  - Allows for multiple Virtual Machines (VM) in a single Root Complex to share a PCI Express* (PCIe*) adapter
- An SR-IOV endpoint presents multiple Virtual Functions (VF) to a Virtual Machine Monitor (VMM)
  - VF allocated to VM => direct assignment
- Address Translation Services (ATS) supports:
  - Performance optimization for direct assignment of a Function to a Guest OS running on a Virtual Intermediary (Hypervisor)
- Page Request Interface (PRI) supports:
  - Functions that can raise a Page Fault
- Process Address Space ID enhancement to support Direct assignment of I/O to user space



Agenda

- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

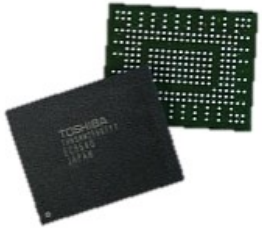
NVM Express™ Driving PCI Express SSDs in Data Center



Source: Forward Insights Q1'15

Data Center Form Factors for PCIe

BGA



16x20 mm ideal for small and thin platforms

M.2

42, 80, and 110mm lengths, smallest footprint of PCI Express® (PCIe®) connector form factors, use for boot or for max storage density

U.2 2.5in (aka SFF-8639)

2.5in makes up the majority of SSDs sold today because of ease of deployment, hotplug, serviceability, and small form factor Single-Port x4 or Dual-Port x2

CEM Add-in-card

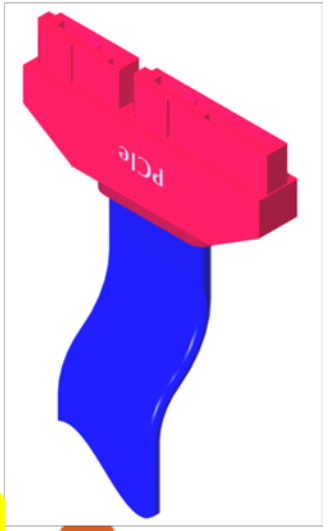
Add-in-card (AIC) has maximum system compatibility with existing servers and most reliable compliance program. Higher power envelope, and options for height and length

Source: Intel Corporation

Inexpensive Cabling = Independent Clock + Spread Spectrum (SSC) (SRIS)

- Challenge: PCIe® specification did not support independent clock with SSC
  - SATA* cable ~ $0.50
  - PCIe cables include reference clock > $1 for equivalent cable
- PCIe base specification 3.0 ECNs approved
  1) Requires use of larger elasticity buffer
  2) Requires more frequent insertion of SKIP ordered set
  3) Requires receiver changes (CDR)
  4) Second ECN updates Model CDRs
- SRIS will create a number of new form factor opportunities for PCIe technology
  - OCuLink*
  - Lower cost external/internal cabled PCIe technology
  - Next generation of PCI-SIG® cable specification

Example of Possible PCIe Cable

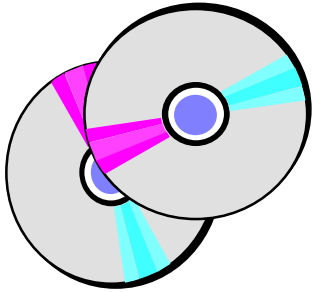


Separate Refclk Modes of Operation: 5600ppm (New - SRIS) and 600ppm (Existing - SRNS)

Agenda

- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

PCIe Compliance Process



Test Tools
And Procedures

Clear Test Output

Maps

- Directly to Test Spec

PCI-SIG® Specs

Describes

Device requirements

- 3.0 Base and CEM specs

C&I Test Specs

Define

Test criteria based on spec requirements

- Test Definitions
- Pass/Fail Criteria

Test H/W & S/W

Validates

Test criteria

- Compliance
- Interoperability

Predictable path to design compliance

Agenda

- Introduction: Evolution of PCI Express Technology
- Power-efficient Performance
- RAS Enhancements
- I/O Virtualization
- Form Factors
- Compliance
- Conclusions

Conclusions

Data Center / HPC | Mobile | Embedded

Source: Intel Corporation

- Single standard covering systems from handheld to data center
- Predominant direct I/O interconnect from CPU with high bandwidth
- Low-power
- High-performance
- Predictive performance growth spanning five generations
- A robust and mature compliance and interoperability program