The New Generation of Storage:
From PCI Express® 4.0 to PCI Express 6.0

Dr. Debendra Das Sharma
PCI-SIG® Board Member
Intel Fellow and Director of I/O Technology and Standards
Intel Corporation
Agenda

- Introduction: Evolution of PCI Express® Technology
- PCI Express and Storage
- Form Factors
- Compliance
- Conclusions
Organization that **defines the PCI Express® (PCIe®) I/O bus specifications and related form factors.**

800+ member companies located worldwide.

Creating specifications and mechanisms to **support compliance and interoperability.**

**PCI-SIG member companies support the following usages with PCIe:**

- Virtual reality
- Automotive
- Artificial intelligence
- Telecommunications
- Storage
- Consumer
- Mobile
- Data Center
Adoption is Well Under Way

- **Key Features:**
  - Data Rate 16 GT/s
  - Maintains full backwards compatibility with PCIe 3.x, 2.x, and 1.x
  - Implements:
    - Extended tags and credits
    - Reduced system latency
    - Lane margining
  - Superior RAS capabilities
  - Scalability for added lanes and bandwidth
  - Improved I/O virtualization and platform integration
  - Maximum channel loss is 28dB

- **Compliance Status:**
  - PCI-SIG Launched Official FYI Testing for PCIe 4.0 in December 2018
  - Formal Compliance testing targeted for Q3 2019

- **Adoption:**
  - Numerous vendors with 16GT/s PHYs and controllers in silicon
  - Test equipment from multiple vendors
  - Several member companies have publicly announced & exhibited PCIe 4.0 products
Published in May 2019

- **Key Features:**
  - Data Rate 32 GT/s
  - Maintains full backwards compatibility with PCIe 4.0, 3.x, 2.x, and 1.x
  - Maximum channel loss is 36dB
  - Electrical changes to improve signal integrity and mechanical performance of connectors
  - Advanced test and debug capabilities

- **Compliance Status:**
  - PCIe 5.0 compliance testing is under development
  - **Adoption**
    - Several member companies have publicly announced and are showcasing PCIe 5.0 solutions and interoperable silicon
    - Adoption expected to grow in the next few months due to demand from high performance applications
PCI Express 6.0® Specification Targets
Aiming for completion in 2021

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Rate</td>
<td>64 GT/s, PAM4 (Pulse Amplitude Modulation – 4 level signaling)</td>
</tr>
<tr>
<td>Latency</td>
<td>Low single-digit ns PHY adder w/ Forward Error Correction (FEC) for (Tx + Rx)</td>
</tr>
<tr>
<td>B/W Efficiency</td>
<td>Better than Gen 1-5 due to protocol enhancements even with FEC overhead</td>
</tr>
<tr>
<td>Reliability</td>
<td>$0 &lt; \text{FIT} \ll 1$ (similar to Gen 5) [FIT: Failure In Time ($10^9$ hours)]</td>
</tr>
<tr>
<td>Channel reach</td>
<td>Similar to Gen 5 (max 2 retimers)</td>
</tr>
<tr>
<td>Power Efficiency</td>
<td>Better than Gen 5 (ideally power neutral while delivering 2X b/w)</td>
</tr>
<tr>
<td>Low Power</td>
<td>L1 with entry/exit latency similar to Gen 5</td>
</tr>
<tr>
<td>Plug n Play</td>
<td>Backwards compatible with prior generations (Software, Silicon, and existing Form Factors)</td>
</tr>
<tr>
<td>Other for Gen6</td>
<td>High Volume Manufacturing, cost-effective, scales to hundreds of Lanes in a platform, simple to design and validate</td>
</tr>
</tbody>
</table>
One Interconnect—Infinite Applications

Artificial Intelligence
- High-performance
- High-bandwidth

Automotive
- High-performance
- Reliability
- Availability
- Serviceability

Cloud
- Scalable architecture
- Increased performance
- Reduced TCO

Enterprise Servers
- Redundancy/failover
- Ubiquity
- Power savings

PC/Mobile/IoT
- Faster performance
- Power efficiency
- Low latency

Storage
- Faster data transfer
- Better user experience
- Ubiquity
Agenda

- Introduction: Evolution of PCI Express Technology
- PCIe and Storage
  - Form Factors
  - Compliance
  - Conclusions
PCI Express is a great interface for SSDs

- Stunning performance
- Lane scalability
- Lower latency
- Lower power
- Lower cost
- CPU-integrated PCIe lanes

With Next Gen NVM, the NVM is no longer the bottleneck

Source: FMS 2013
“NVMe Express Overview & Ecosystem Update”
Growth of PCIe® Technology in Storage

- **Data explosion is driving SSD adoption**
  - SSD market CAGR of 14.8% during 2016-2021 *Source: IDC*
  - PCIe SSD market to surpass a CAGR of 33% during 2016-2020 *Source: Technavio*

- PCIe technology is outpacing other interconnect technologies in both units and bandwidth/capacity

*Source: SSD Insights Q1/18, Forward Insights*
PCIe® Features useful for Storage

- Low-latency, High Bandwidth, Scalability, and predictable cadence of speed increase with backwards compatibility
- In addition, PCIe technology offers the following value-add essential for storage
  - Reliability, Availability and Serviceability (RAS)
  - I/O Virtualization
  - Multitude of form factors including cabling support
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
(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
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
Inexpensive Cabling = Independent Clock + Spread Spectrum (SSC) (SRIS)

- **Challenge: PCIe® specification did not support independent clock with SSC initially**
  - SATA* cable ~ $0.50
  - PCIe cables include reference clock > $1 for equivalent cable
  - Routing reference clock across the chassis to front of the rack for storage access is a challenge

- **PCIe base specification has included support since PCIe 3.1**
  1) Requires use of larger elasticity buffer
  2) Requires more frequent insertion of SKIP ordered set
  3) Requires receiver changes (CDR)
  4) Model CDRs

- **SRIS enables a number of form factors for PCIe technology**
  - OCuLink
  - Lower cost external/internal cabled PCIe technology

Separate Refclk Modes of Operation: 5600ppm (SRIS) for 2.5, 5.0, 8.0, and 16.0 GT/s Data Rates and 3600 ppm for 32.0 GT/s; 600ppm (SRNS)
Agenda

- Introduction: Evolution of PCI Express Technology
- PCIe and Storage
- **Form Factors**
  - Compliance
  - Conclusions
PCIe® Form Factors

**BGA**
- 11.5x13 & 16x20mm small and thin platforms

**M.2**
- 30, 42, 80, and 110mm Smallest footprint of PCIe connector form factors, use for boot or for max storage density

**U.2 2.5in**
- Majority of SSDs sold Ease of deployment, hotplug, serviceability 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

Majority of SSDs sold

Ease of deployment, hotplug, serviceability

Single-Port x4 or Dual-Port x2

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

High B/W with PCIe 3.0
Prevalent in hand-held, IoT, automotive
SFF Form Factors

(SFF TA 1002) (Up to 36 Modules)

(SFF TA 1006 – SSD) (Up to 32 Modules)

(SFF TA 1007 – SSD)
Agenda

- Introduction: Evolution of PCI Express Technology
- PCIe and Storage
- Form Factors
- **Compliance**
- Conclusions
PCIe® Compliance Process

PCI-SIG® Specs

Describes
Device requirements
- 3.0 Base and CEM specs

C&I Test Spec

Define
Test criteria based on spec requirements
- Test Definitions
- Pass/Fail Criteria

Test Tools And Procedures

Test H/W & S/W
Validates
Test criteria
- Compliance
- Interoperability

PASS
FAIL

Clear Test Output Maps
- Directly to Test Spec

Predictable path to design compliance
Conclusions

- 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 six generations
- A robust and mature compliance and interoperability program
Follow PCI-SIG on Social Media

@PCI_SIG  https://www.linkedin.com/company/pcisig/  PCI-SIG