IDF15
INTEL DEVELOPER FORUM

What You Need to Know to Win the Storage Transition – Preparing for NVM Express™ in the Data Center

Jonmichael Hands
Product Manager, Intel Corporation, Non-Volatile Memory Solutions Group

Michael Hall
Director of Technology Solutions Enabling, Intel Corporation, Data Center Group

SSDS003

intel
experience what's inside™

Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

NVMe™ is Transforming the Data Center

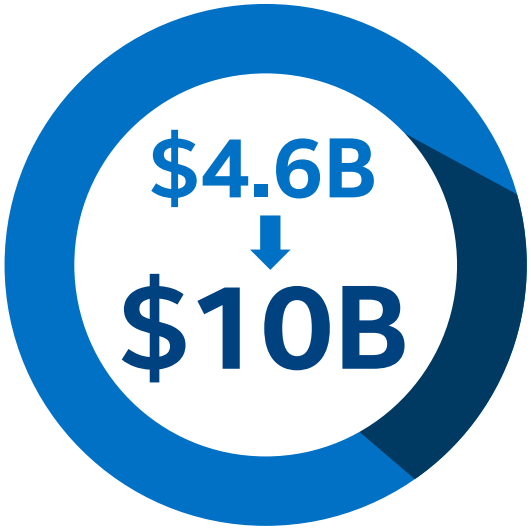
Developed to be lean, NVMe™ delivers high performance with less resources, takes full advantage of multi-core CPUs, and increases storage density to lower TCO.

Scalable performance and low latency of PCIe® with NVMe optimizing the storage stack.

NVMe is bringing PCIe SSDs into the mainstream with industry standard software and drivers and management for the future software defined data center.

Market Overview

Massive data growth is driving SSDs into the data center with NVMe™ as the interface of choice



Data Center SSD Market

Will be approaching $10B in 2018, was $4.6B in 2014

2018 DC Storage TAM

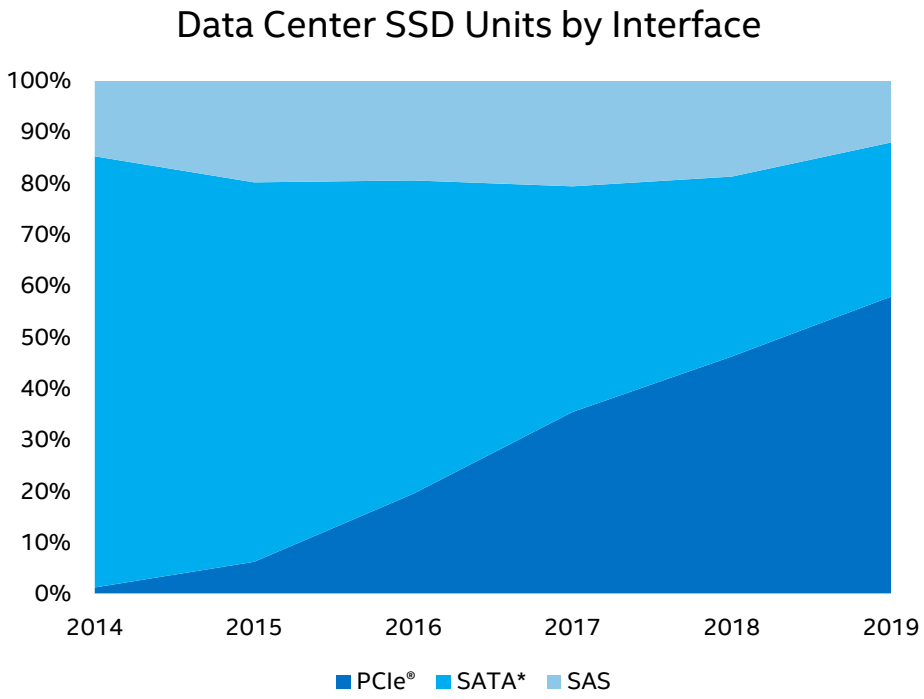
More than 40% of revenue projected to be SSDs, the rest on HDDs

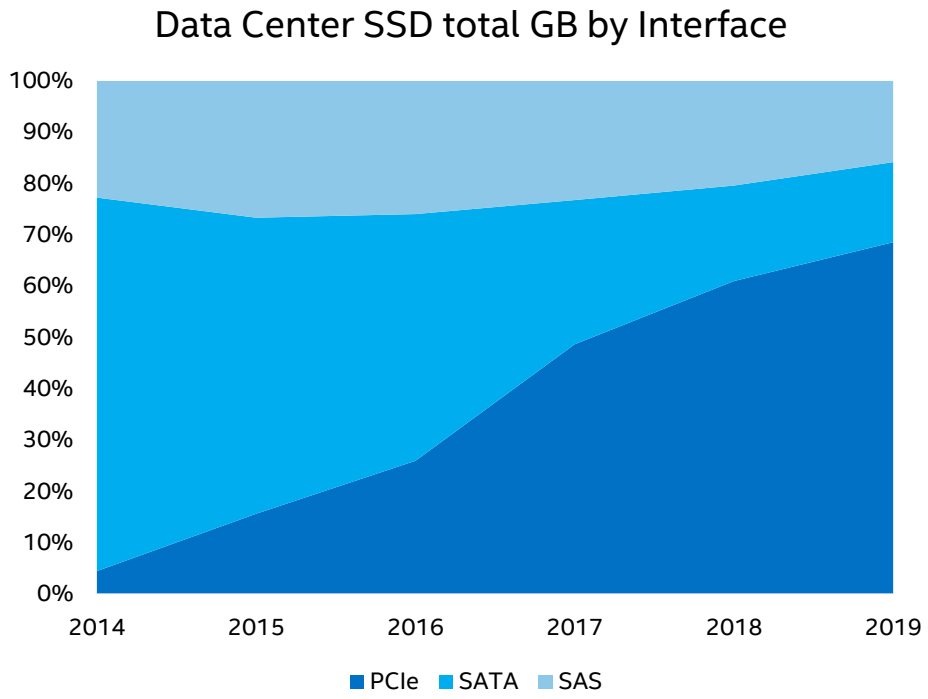
NVMe by 2017

Half the data center SSD market is NVMe by 2017

Source: Forward Insights Q1'15, Intel Q2'15

NVMe™ Driving PCIe® SSDs in the Data Center





Source: Intel Q2'15

Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

Ecosystem

Infrastructure for PCIe®: Cables, connectors, backplanes, switches, retimers

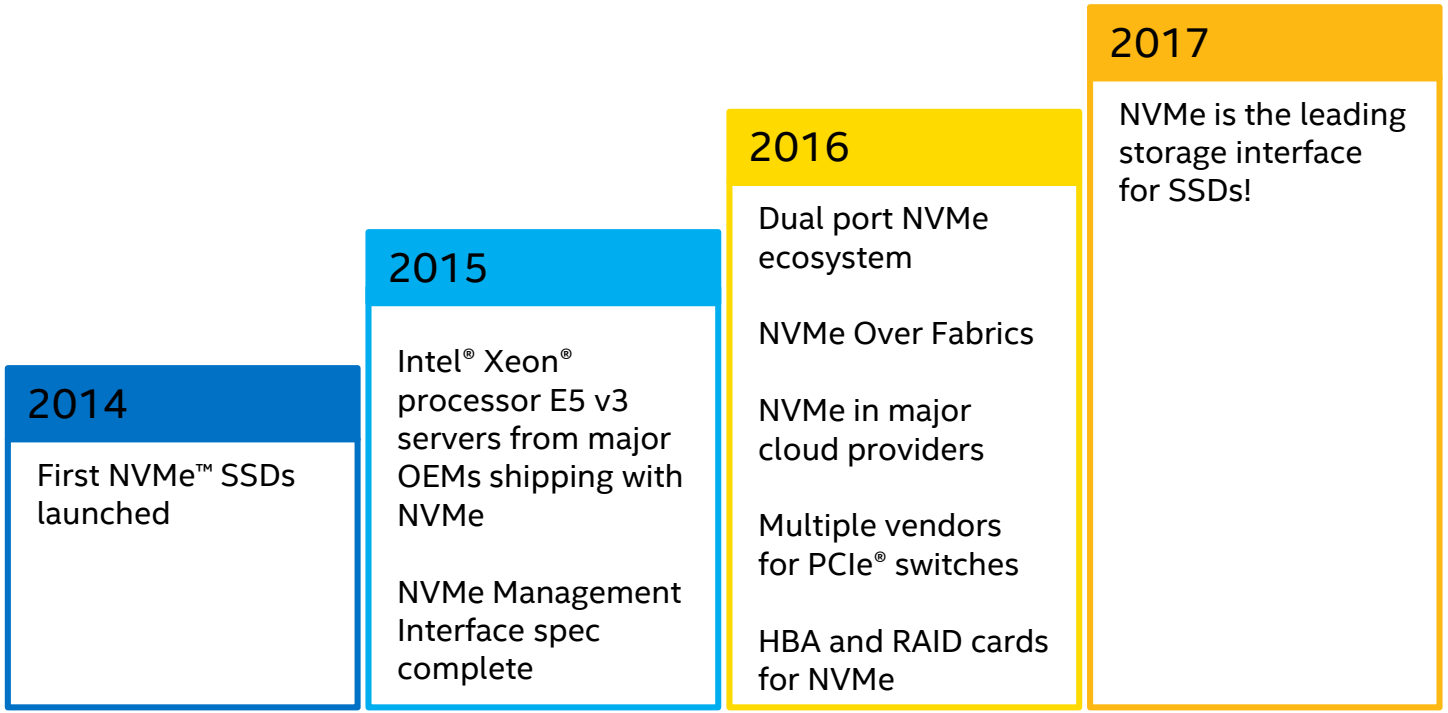
NVMe™ Software, drivers, and management

RAID for NVMe with Intel® Rapid Storage Technology enterprise

IHV and ISV partners

PCIe and NVMe compatibility programs

Data Center NVMe™ Ecosystem Timeline



NVMe™ Driver Ecosystem

Windows Server 2012 R2 Certified

Windows 8 Compatible

Windows 10

Windows 8.1

redhat

6.5, 6.6, 6.7
7.0, 7.1

suse

SLES 11 SP3
SLES 12

ubuntu

13, 14

solaris

FreeBSD

vmware ESXi 6.0

Native / in-box

Windows Server 2012 Certified

CERTIFIED FOR Windows Server 2008 R2

Windows 7

vmware ESXi 5.5

Install NVMe™ driver

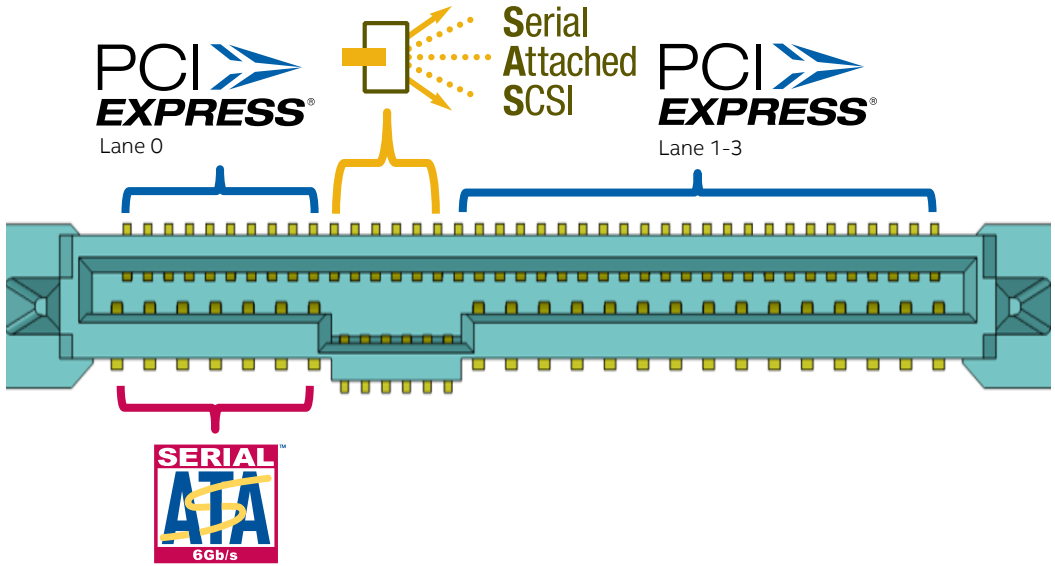
Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

One Connector to Rule Them All

U.2

(Formerly SFF-8639)



Data Center Form Factors for PCI EXPRESS®

2.5in is projected more than 86% of all SSD units sold on all interfaces between 2015-2019

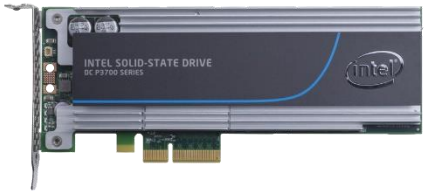
M.2

42, 80, and 110mm lengths, Smallest footprint of PCIe®, use for boot or for max storage density.

U.2 2.5in (SFF-8639)

2.5in makes up the majority of SSDs sold today because of ease of deployment, hotplug, serviceability, and small form factor.

Add-in-card



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

Source: Forward Insights Q2'15

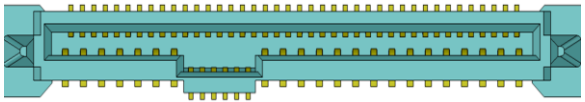
U.2 Naming Conventions

U.2 drive / U.2 SSD

PCIe® 2.0 or 3.0, x1, x2, x4
2.5in x 15mm and 7mm, <10W and 10-25W

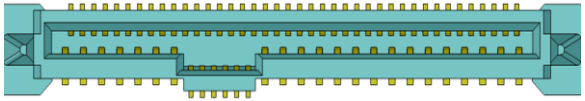
U.2 connector

Connector can be on cable or backplane
Supports PCIe, SAS, and SATA*



U.2 backplane

On a server or workstation



U.2 cable

A cable that connects a U.2 drive

U.2 host connector

Compliance and Interoperability Opportunities

PCI EXPRESS®

Upcoming PCI-SIG* compliance workshops:

- August 11-14th – Milpitas
- October 6-8th – Taipei
- December 1-4th - Milpitas

Compliance program for M.2 form factors is being encouraged by Intel.

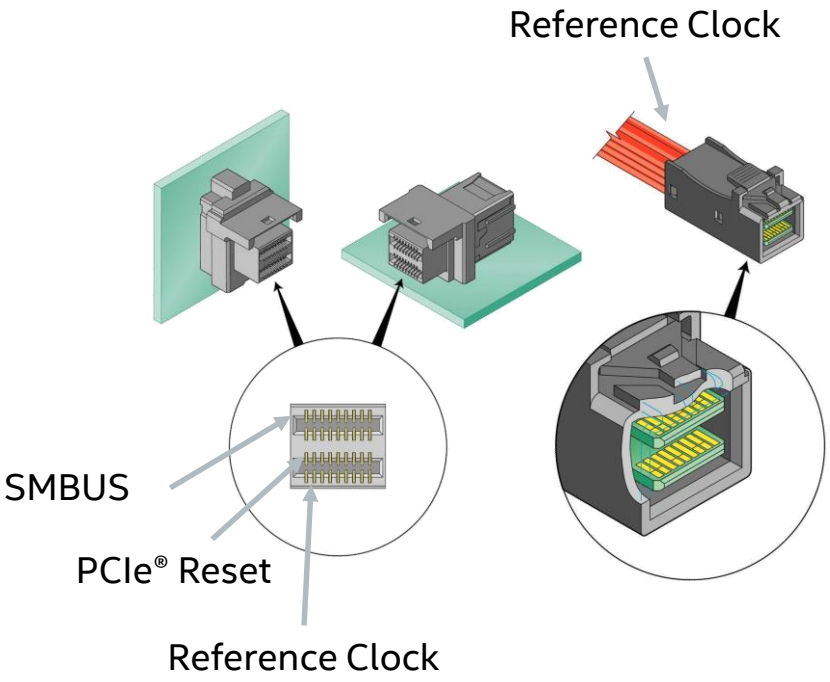
iol University of New Hampshire InterOperability Laboratory

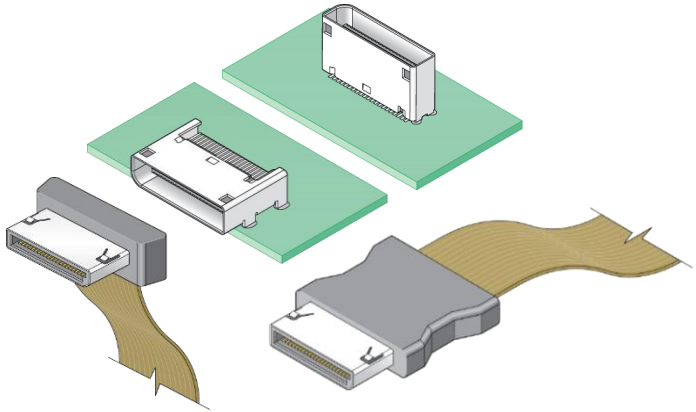
NVMe™ Integrators List testing

- June 6-9th, 2016 – UNH
- Possible event in November 2015 or January 2016

2015 saw the introduction of Hot Plug testing included in UNH testing services.

Cables for PCIe® SSDs – miniSAS HD and OCuLink*



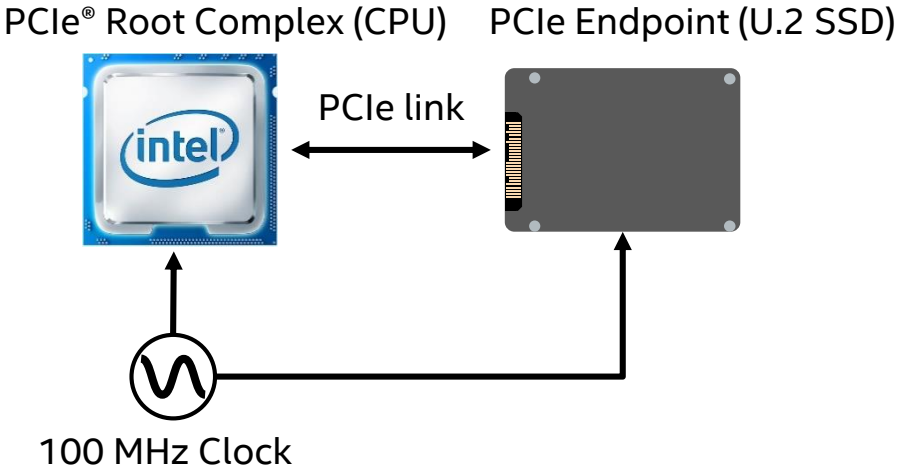


OCuLink* internal cables and connectors

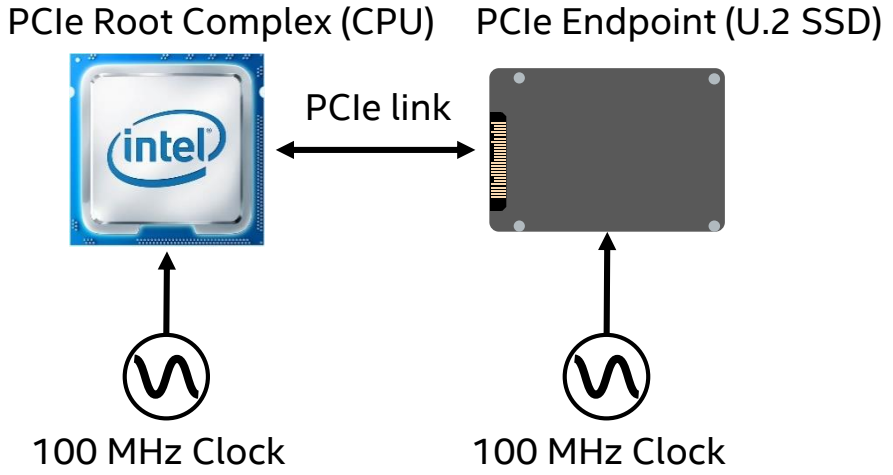
Seeing deployments of miniSAS HD for PCIe in designs today

Comparing Clocking Mechanisms for PCIe®

Reference Clock



Separate Reference Clock with Independent SSC (SRIS)



Tradeoffs of Using SRIS

Save money on clocks

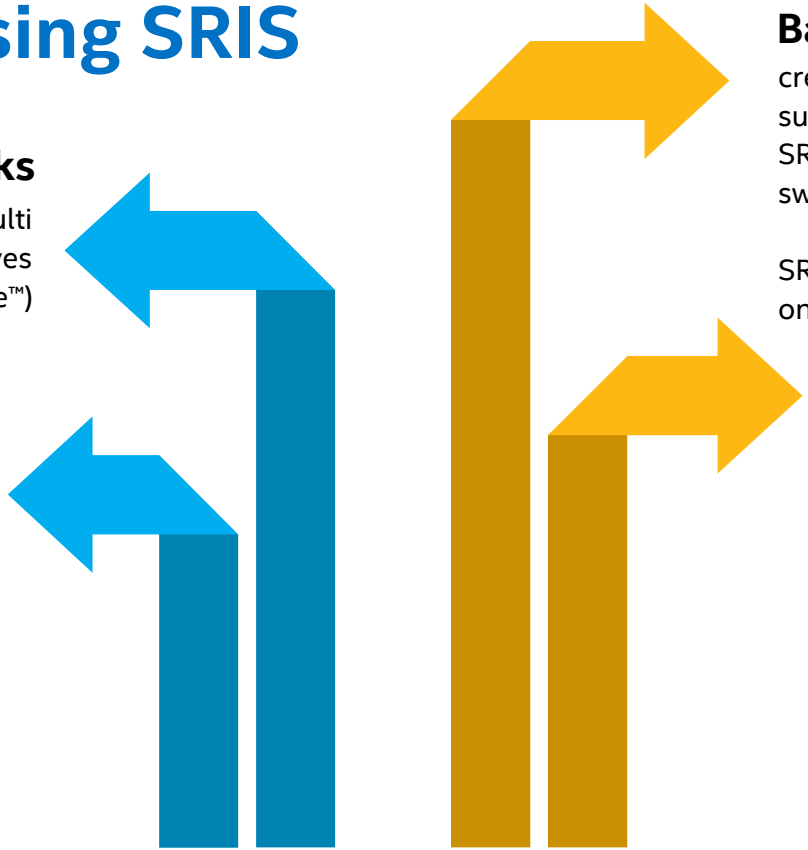
And make routing easier for multi drive systems (specifically >24 drives and dual port NVMe™)

Cables and routing

May lead to less expensive cables (less shielding, fewer wires)

Reduces number of signals to be routed

Lower EMI



Benefits of SRIS **Downside of SRIS**

Backwards compatibility

creates challenges for U.2 SSDs supporting both common clock and SRIS topologies without standard switching mechanism.

SRIS was originally invented for SRIS only topologies!

Hardware support

Requires support on both host root complex (CPU or PCIe® switch) and downstream device (SSD, switch, HBA, etc.)

May reduce performance by 1-3% from extra SKP ordered sets in protocol

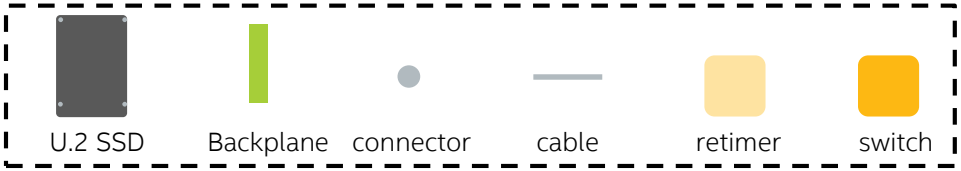
Requires support from link extension devices

Engage with PCI-SIG* to shape how SRIS is used in the future

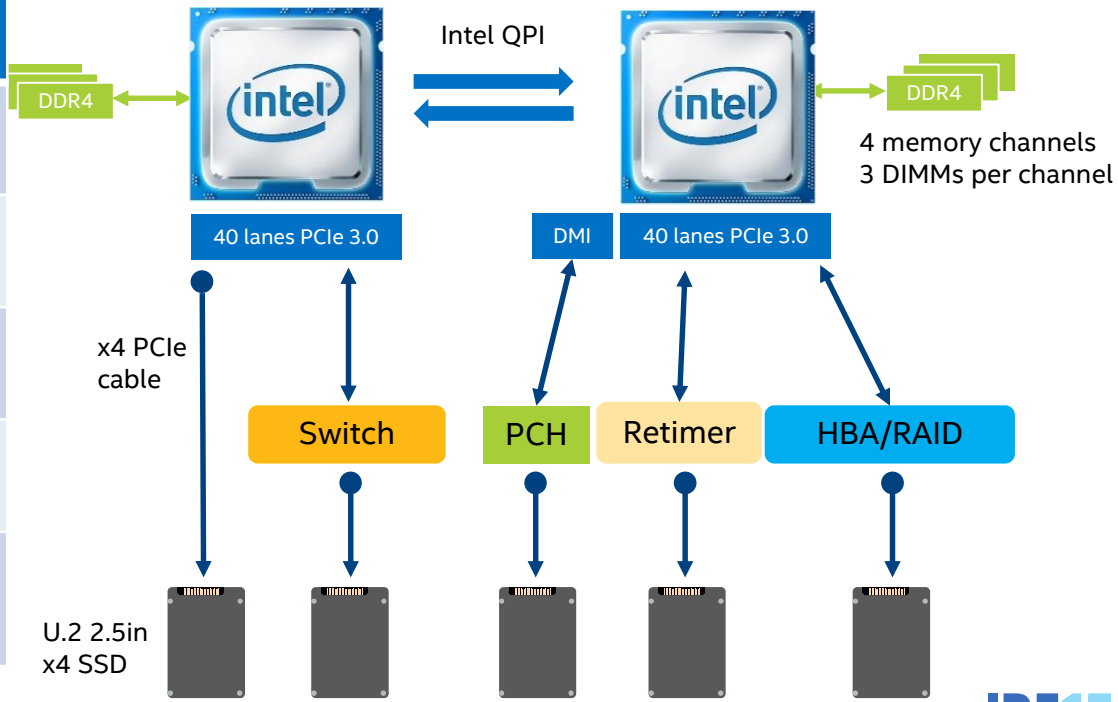
Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

Multiple Topologies for U.2 SSDs

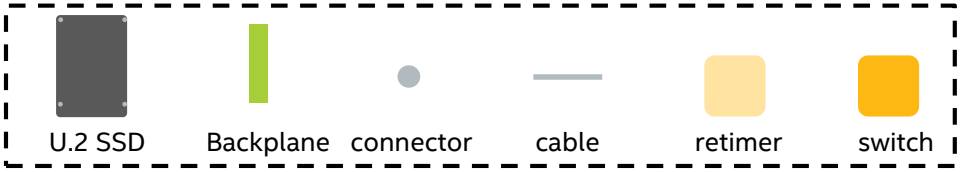


Topology	Key reason to use
CPU attach	High performance, low cost
Switch	Flexibility (x2, x4), high drive count (capacity)
PCH	Cheap attach, x1-x4 PCIe®, don't use CPU lanes
Retimer	Inexpensive link extension of PCIe, routing distance
HBA/RAID card	Hardware RAID of NVMe™ devices, support for multiple protocols (SAS, SATA*, PCIe)



Intel® Quick Path Interconnect (Intel® QPI)

Cabled Topologies



<$ >$

1

1 connector topology used in blade, microserver, compute nodes

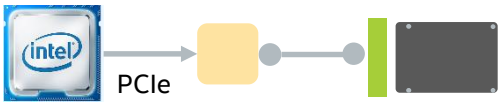
2

2 connector topology used in cloud for high density and inexpensive attach of U.2 SSD at full performance

3

Direct connect to CPU, supports hotplug, mainstream use up to 4 drives

4/5



Use retimer for inexpensive link extension for longer PCIe cabling

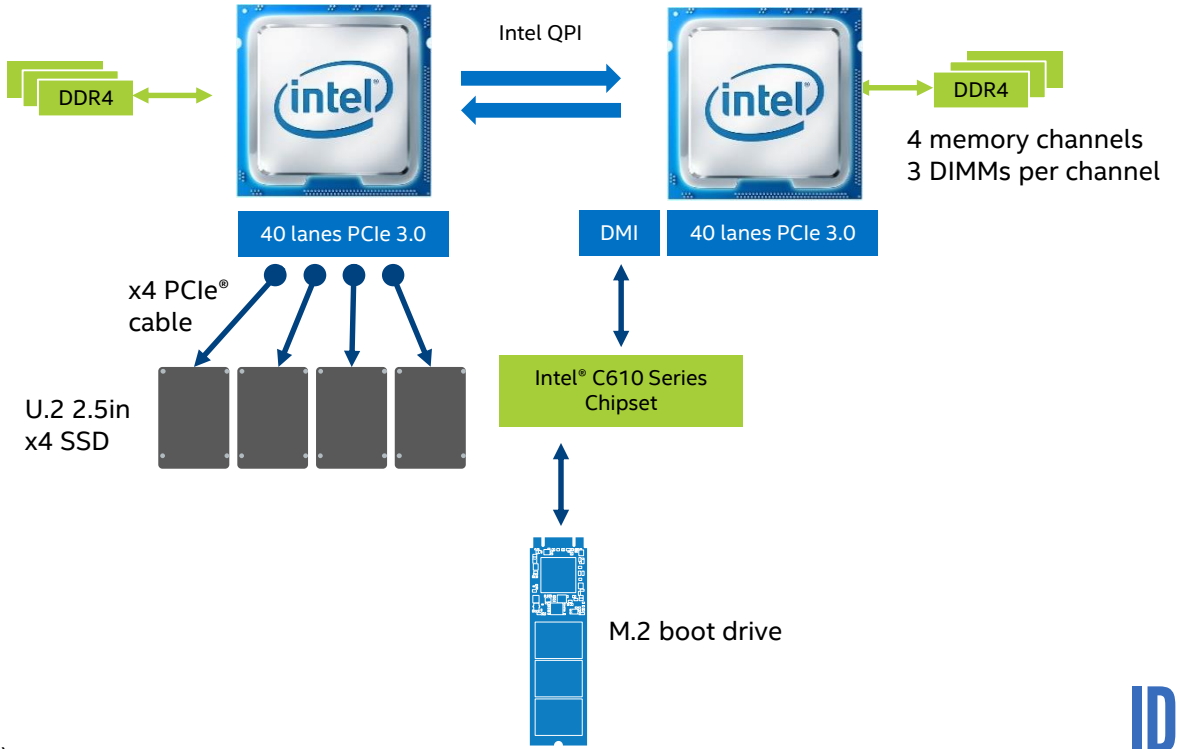
4/5

Common mainstream use, switch topology to support more than 4 U.2 SSDs

Baseline 2.5in PCIe® SSD Server

support four NVMe™ SSDs in cabled PCIe topology directly from CPU

- NVMe™ boot support from CPU
  SATA* boot support from PCH
- Use 16 lanes for four x4 SSDs
- Software RAID 0, 1, 10, 5
  using Intel RSTe
- Hotplug supported with
  backplane
- Cheapest attach of 2.5in. No
  HBA or RAID card required



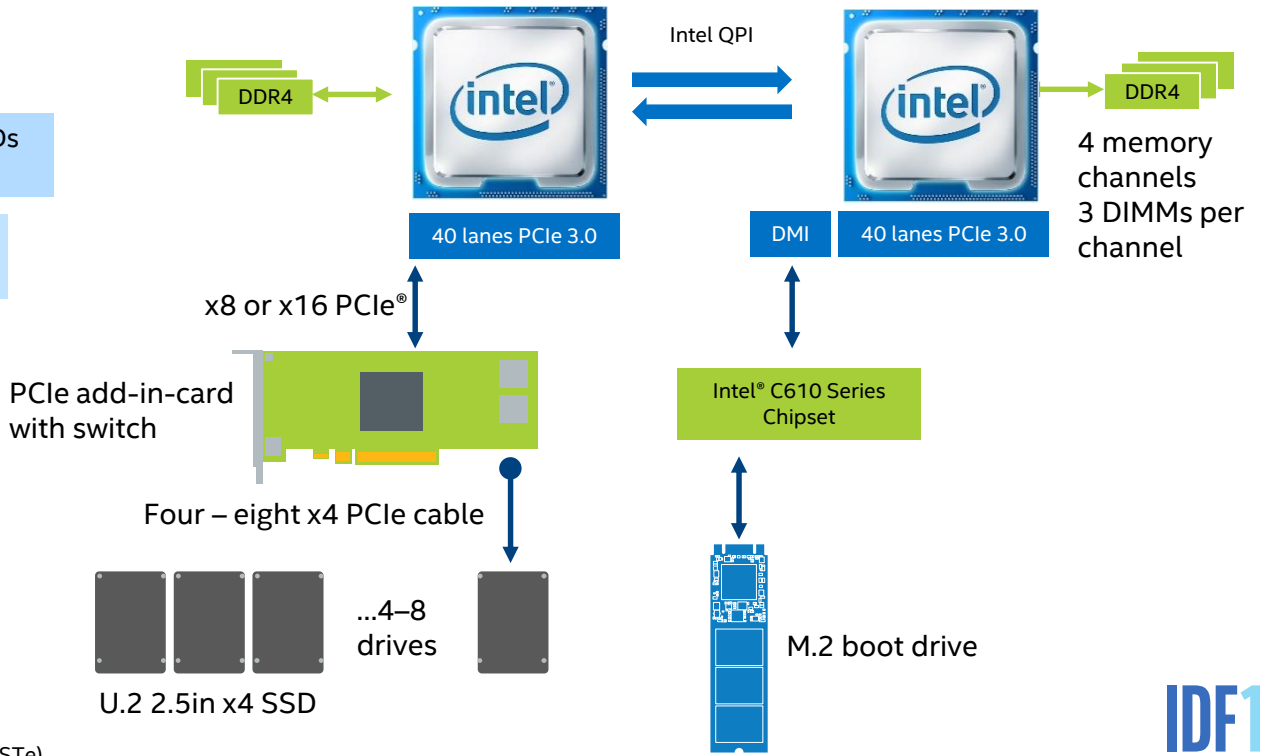
Intel® Rapid Storage Technology enterprise (Intel® RSTe)

Intel® QuickPath Interconnect (Intel® QPI)

Cabled PCIe® Topology with PCIe Switch

support 4-8 NVMe™ SSDs in cabled PCIe topology with switch for slot configurability, link extension and port expansion (multi drive support). Add to existing designs with minimal updates!

- NVMe™ boot support from CPU
  SATA* boot support from PCH
- Use 16 lanes for four - eight x4 SSDs and single add-in-card slot
- Software RAID 0, 1, 10, 5 using Intel RSTe
- Hotplug supported from switch
- Less BIOS development when using PCIe switch



Intel® Rapid Storage Technology enterprise (Intel® RSTe)

Intel® QuickPath Interconnect (Intel® QPI)

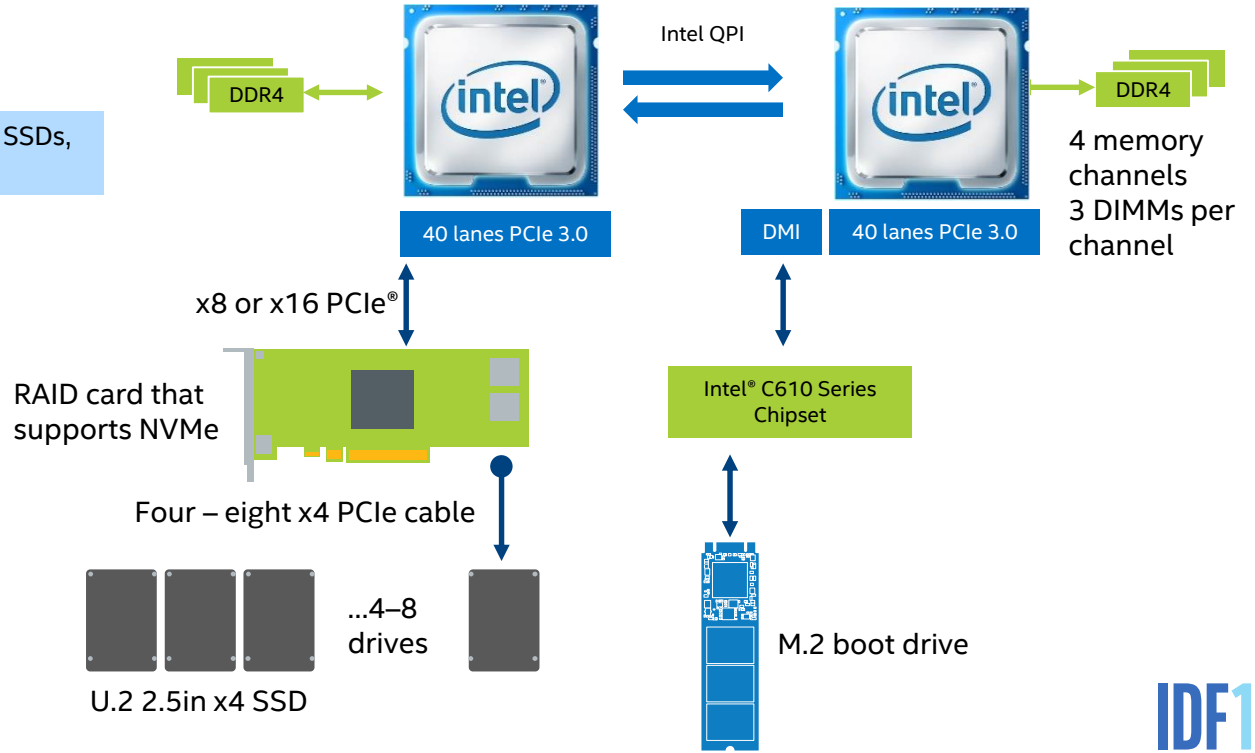
Cabled PCIe® Topology with HBA or RAID Card

Use 8 or 16 lanes for four - eight x4 SSDs, or x2 config (backplane dependent)

Hardware RAID

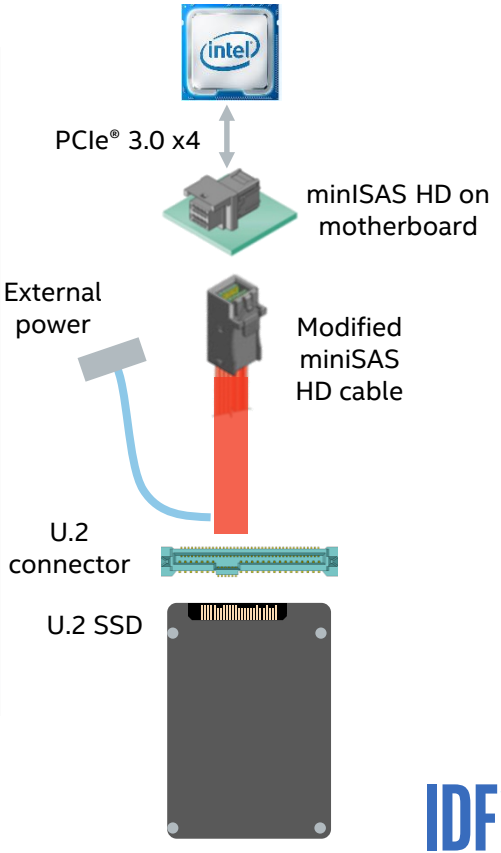
Hotplug supported from IHV

IHV owns compatibility



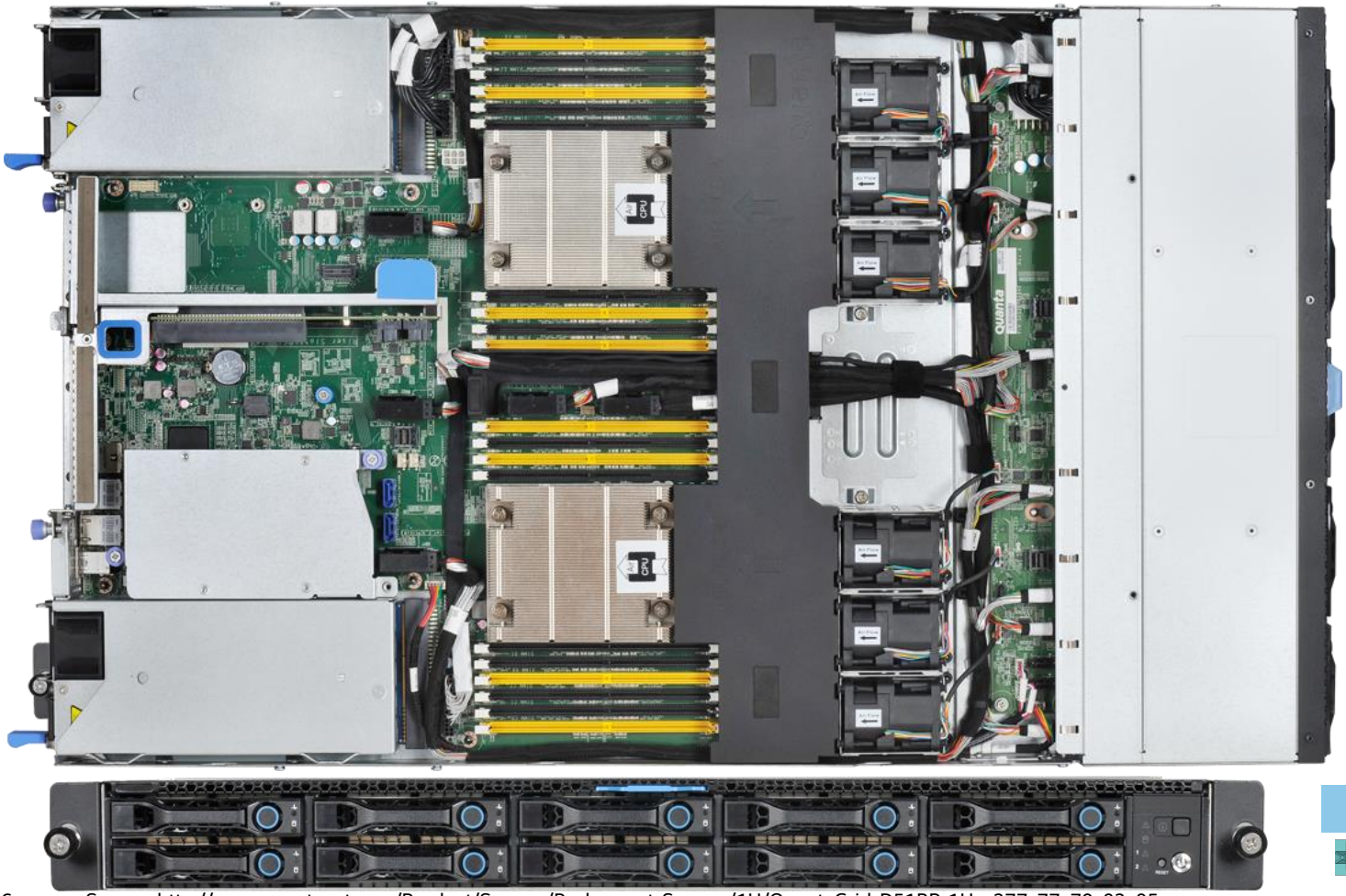
IHV = Independent Hardware Vendor

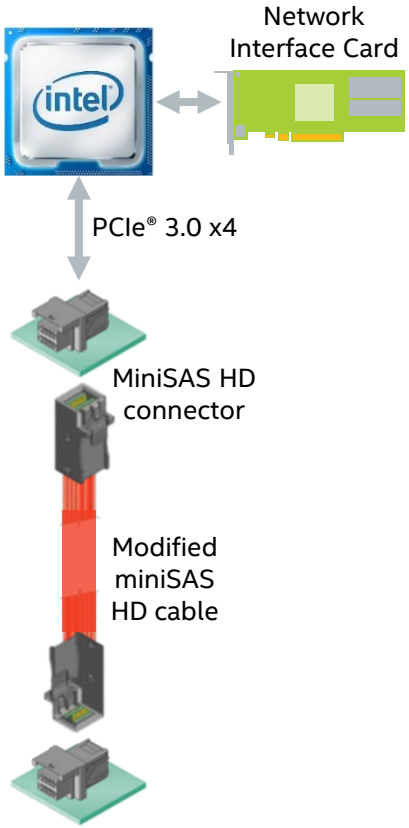
Hyve Solutions* – 2 Connector Topology for Cloud



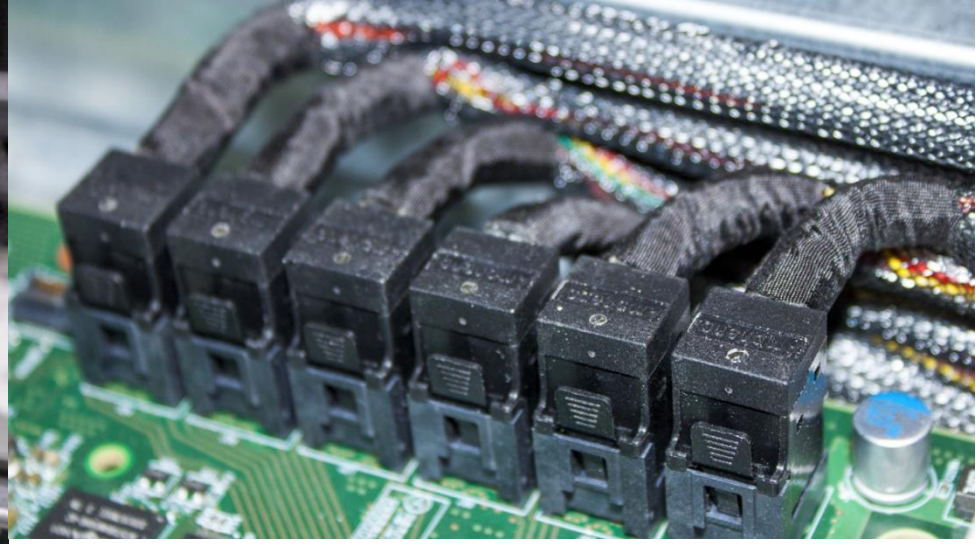
Source: http://hyvesolutions.com/solutions/ambient/

Quanta* – NVMe™ in the Cloud





Source: http://www.quantaqct.com/Product/Servers/Rackmount-Servers/1U/QuantaGrid-D51BP-1U-p277c77c70c83c85



Low Cost Attach of NVMe™ on Mainstream Cloud Designs

IDF15
INTEL DEVELOPER FORUM

Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

Software RAID vs Hardware RAID for NVMe™

RAID is used for: **Performance, aggregation** of storage, and/or **data protection**

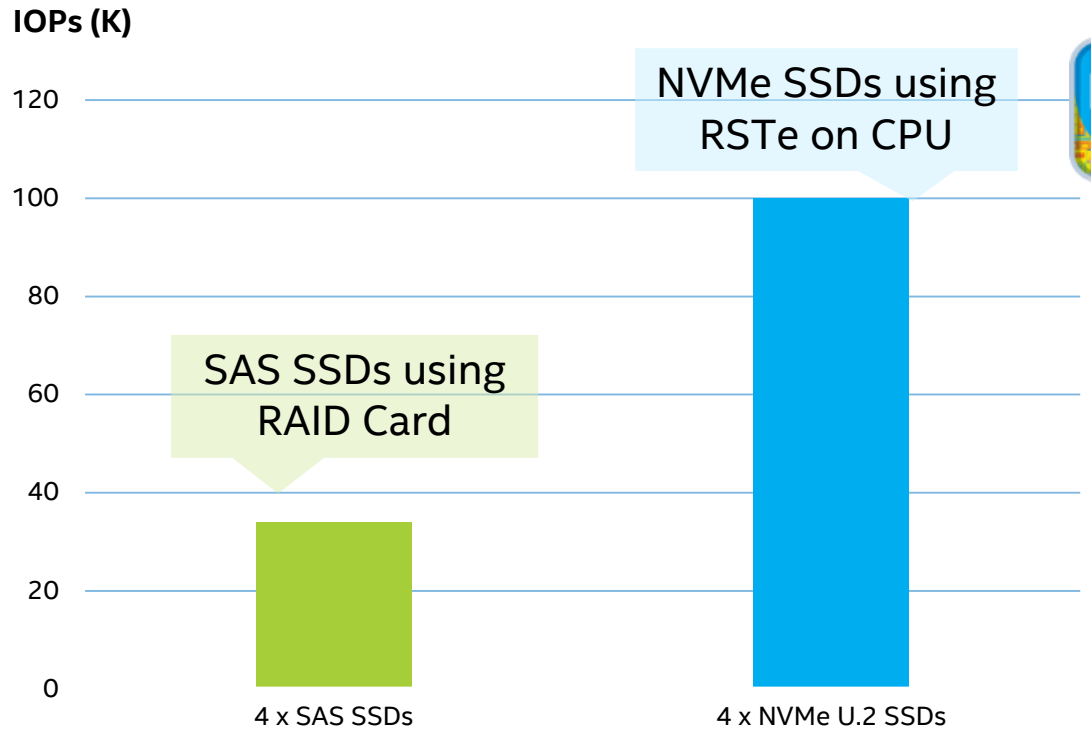
Hardware RAID

- Long history, support from established IHVs
- Dedicated resources for RAID calculations
  - May have lower total IOPS due to hardware limitations
- Software stack from IHVs for management, RAID, hotplug

Software RAID

- Higher performance
- No hardware required (-$)
- Currently lacks support for RAID 5 write hole closure
  - Requires double fault for data loss: power loss and/or drive loss during rebuild
- Not bootable (yet!)

NVMe™ RAID With Intel® Rapid Storage Technology Enterprise



NVMe™ SSDs directly connected to processor

Intel® Rapid Storage Technology enterprise (Intel® RSTe)

3x faster than a RAID Card in RAID 5

Test configuration- Intel GZ2600 Server, dual socket Intel® Xeon® E5-2699 v3, 32GB DDR4 RAM, Hardware RAID controller. **Operating System:** Microsoft Windows Server 2012R2. **Configuration 1- 12G SAS . 4**ea (Seagate 1200 SSD ST800FM0043 SSDs-800GB in RAID5 (LSI MegaRAID SAS-3 3108 Controller). **Configuration 2 – NVMe-RSTe *4ea*** (Intel® DC P3700 800GB SSD) in RAID5 array. Intel RSTe 4.3 BETA software. Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Other names and brands may be claimed as the property of others

Agenda

- NVM Express™ (NVMe) is Transforming the Data Center
- NVMe Ecosystem - Getting Ready for Deployments at Scale
- PCI Express® (PCIe) Infrastructure: Form Factors, Cables, and One Connector to Rule Them All
- Exploring PCIe Topologies
- RAID on NVMe – Putting it All Together
- Summary and Q&A

Summary and Next Steps

- NVMe™ will be the predominant storage interface for data center SSDs as soon as 2016
- The U.2 connector is making the transition happen with flexibility; making many topologies available for designers
- The ecosystem has matured and designs are multiplying today for high drive count deployments
- Design NVMe into your product now to take full advantage of next generation SSDs and 3D XPoint™ technology

Additional Sources of Information

- A PDF of this presentation is available from our Technical Session Catalog: www.intel.com/idfsessionsSF. This URL is also printed on the top of Session Agenda Pages in the Pocket Guide.

- Demos in the showcase: Please visit the NVM Express™ Community, and the Intel Solid-State Drive Pavilion

- Additional info in the NVM Express community

- More web based info: www.nvmexpress.org and www.intel.com/ssd

Other Technical Sessions

Session ID	Title	Day	Time	Room
✓ SSDS001	NVM Express™: The Data Center and Client Storage Transformation	Tues	1:15	2008
✓ SSDS002	SSDs are Here – The Next Wave in Non-Volatile Memory-Driven Storage Modernizations	Tues	2:30	2008
SPCS006	Technology Insight: Intel Non-Volatile Memory Inside. The Speed of Possibility Outside	Tues	5:15	3016
SSDL001	Hands-on Lab: How to Unleash Your Storage Performance by Using NVM Express* based PCI Express* Solid-State Drives	Wed	1:15; 4:00	2010
SSDC001	Tech Chat: Benchmarking Data Center Solid-State Drives – Insights Into Industry-Leading NVM Express* SSD Performance Metrics	Wed and Thurs	10:30 Wed 9:30 Thurs	Tech Chat Station 1
SSDC002	Tech Chat: Insights into Intel® Solid-State Drives Data Retention and Endurance	Wed and Thurs	10:30 Wed 9:30 Thurs	Tech Chat Station 2
SSDC003	Tech Chat: NVM Express* Features for High Availability and Storage Eco-System	Wed and Thurs	10:30 Wed 9:30 Thurs	Tech Chat Station 3
SSDS004	The Future of Storage Security	Thurs	1:00	2006
SSDS005	New Software Capabilities and Experiences Through Innovation in Storage Architecture	Thurs	2:15	2006

✓ = DONE

Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, Xeon, 3D XPoint, and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the second quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as "anticipates," "expects," "intends," "plans," "believes," "seeks," "estimates," "may," "will," "should" and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be important factors that could cause actual results to differ materially from the company's expectations. Demand for Intel's products is highly variable and could differ from expectations due to factors including changes in business and economic conditions; consumer confidence or income levels; the introduction, availability and market acceptance of Intel's products, products used together with Intel products and competitors' products; competitive and pricing pressures, including actions taken by competitors; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel's gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; and product manufacturing quality/yields. Variations in gross margin may also be caused by the timing of Intel product introductions and related expenses, including marketing expenses, and Intel's ability to respond quickly to technological developments and to introduce new products or incorporate new features into existing products, which may result in restructuring and asset impairment charges. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Results may also be affected by the formal or informal imposition by countries of new or revised export and/or import and doing-business regulations, which could be changed without prior notice. Intel operates in highly competitive industries and its operations have high costs that are either fixed or difficult to reduce in the short term. The amount, timing and execution of Intel's stock repurchase program could be affected by changes in Intel's priorities for the use of cash, such as operational spending, capital spending, acquisitions, and as a result of changes to Intel's cash flows or changes in tax laws. Product defects or errata (deviations from published specifications) may adversely impact our expenses, revenues and reputation. Intel's results could be affected by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel's ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. Intel's results may be affected by the timing of closing of acquisitions, divestitures and other significant transactions. A detailed discussion of these and other factors that could affect Intel's results is included in Intel's SEC filings, including the company's most recent reports on Form 10-Q, Form 10-K and earnings release.

Rev. 4/14/15

Backup

Infrastructure Building Blocks

Switches

Used to expand the ports for PCIe® and critical for deploying multiple U.2 drives. Storage specific will emerge with improved hotplug support, enclosure management, and other features.

Retimers

Used for link extension of PCIe to increase lengths of cabled PCIe. Software transparent but participate in PCIe link training upstream to host and downstream to SSD. The low cost makes these attractive!

HBA and RAID cards

Similar to a PCIe switch to allow port expansion of PCIe SSDs, but allows devices to be aggregated in a RAID volume. Supports storage features such as enclosure management, hotplug.

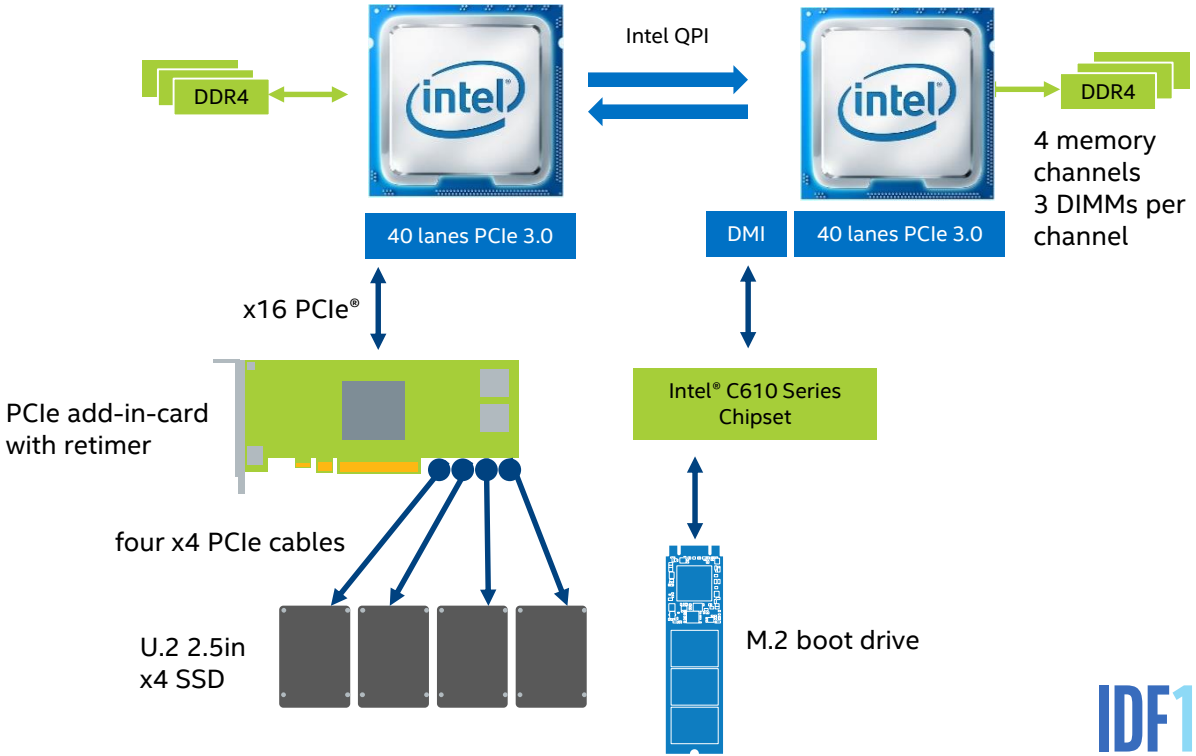
Cables

Choose between miniSAS HD for PCIe and OCuLink*

Cabled PCIe® Topology with PCIe Retimers

Support four NVMe™ SSDs in cabled PCIe topology directly from CPU with link extension, uses add-in-card slot but saves connector cost on board and gains extra cable length

- NVMe™ boot support from CPU
  SATA* boot support from PCH
- Use 16 lanes for four x4 SSDs,
  Requires bifurcation from CPU
- Software RAID 0, 1, 10, 5
  using Intel RSTe
- Hotplug supported with
  backplane
- Retimer card ~$50 BOM cost
  for inexpensive link extension



Intel® Rapid Storage Technology enterprise (Intel® RSTe)

Intel® QuickPath Interconnect (Intel® QPI)

MiniSAS HD for PCIe® Pinout

A0 is an internal only GND Pad (Yellow Circles) **This is NEW!**
A3 is an existing GND Pad (Green Circles)

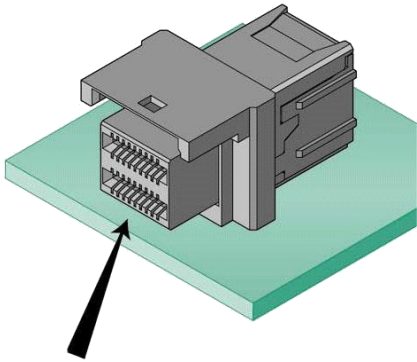


Table14 (Controller/Baseboard/Add-in Card side)

Pin Number	D1	D2	D3	D4	D5	D6	D7	D8	D9
Pin Name	Sideband5	Sideband6	GND	Tx0+	Tx0-	GND	Tx2+	Tx2-	GND
Signal Name	BMC_SMB_DAT	BMC_SMB_CLK	GND	PE_SSD_TX_DP<0>	PE_SSD_TX_DN<0>	GND	PE_SSD_TX_DP<2>	PE_SSD_TX_DN<2>	GND
Signal Name	CPU_SMB_DAT	CPU_SMB_CLK	GND	PE_SSD_TX_DP<1>	PE_SSD_TX_DN<1>	GND	PE_SSD_TX_DP<3>	PE_SSD_TX_DN<3>	GND
Pin Name	Sideband4	Sideband2	GND	Tx1+	Tx1-	GND	Tx3+	Tx3-	GND
Pin Number	C1	C2	C3	C4	C5	C6	C7	C8	C9

Pin Number	B1	B2	B3	B4	B5	B6	B7	B8	B9
Pin Name	Sideband3	Sideband1	GND	Rx0+	Rx0-	GND	Rx2+	Rx2-	GND
Signal Name	PE_RST_N	GND	GND	PE_SSD_RX_DP<0>	PE_SSD_RX_DN<0>	GND	PE_SSD_RX_DP<2>	PE_SSD_RX_DN<2>	GND
Signal Name	CLK_100M_P	CLK_100M_N	GND	PE_SSD_RX_DP<1>	PE_SSD_RX_DN<1>	GND	PE_SSD_RX_DP<3>	PE_SSD_RX_DN<3>	GND
Pin Name	Sideband7	Sideband0	GND	Rx1+	Rx1-	GND	Rx3+	Rx3-	GND
Pin Number	A1	A2	A3	A4	A5	A6	A7	A8	A9

Table15 (HSBP side)

Pin Number	D1	D2	D3	D4	D5	D6	D7	D8	D9
Pin Name	Sideband6	Sideband5	GND	Tx0+	Tx0-	GND	Tx2+	Tx2-	GND
Signal Name	BMC_SMB_CLK	BMC_SMB_DAT	GND	PE_SSD_RX_DP<0>	PE_SSD_RX_DN<0>	GND	PE_SSD_RX_DP<2>	PE_SSD_RX_DN<2>	GND
Signal Name	CPU_SMB_CLK	CPU_SMB_DAT	GND	PE_SSD_RX_DP<1>	PE_SSD_RX_DN<1>	GND	PE_SSD_RX_DP<3>	PE_SSD_RX_DN<3>	GND
Pin Name	Sideband2	Sideband4	GND	Tx1+	Tx1-	GND	Tx3+	Tx3-	GND
Pin Number	C1	C2	C3	C4	C5	C6	C7	C8	C9

Pin Number	B1	B2	B3	B4	B5	B6	B7	B8	B9
Pin Name	Sideband1	Sideband3	GND	Rx0+	Rx0-	GND	Rx2+	Rx2-	GND
Signal Name	GND	PE_RST_N	GND	PE_SSD_TX_DP<0>	PE_SSD_TX_DN<0>	GND	PE_SSD_TX_DP<2>	PE_SSD_TX_DN<2>	GND
Signal Name	CLK_100M_N	CLK_100M_P	GND	PE_SSD_TX_DP<1>	PE_SSD_TX_DN<1>	GND	PE_SSD_TX_DP<3>	PE_SSD_TX_DN<3>	GND
Pin Name	Sideband0	Sideband7	GND	Rx1+	Rx1-	GND	Rx3+	Rx3-	GND
Pin Number	A1	A2	A3	A4	A5	A6	A7	A8	A9

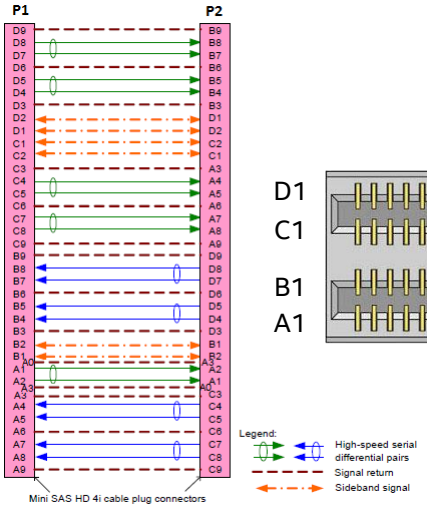
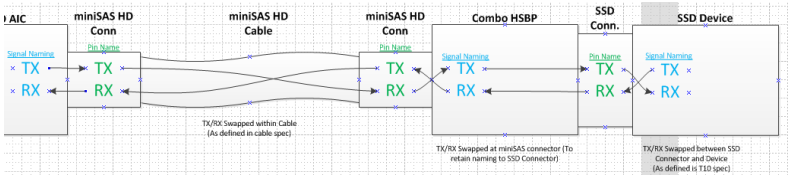
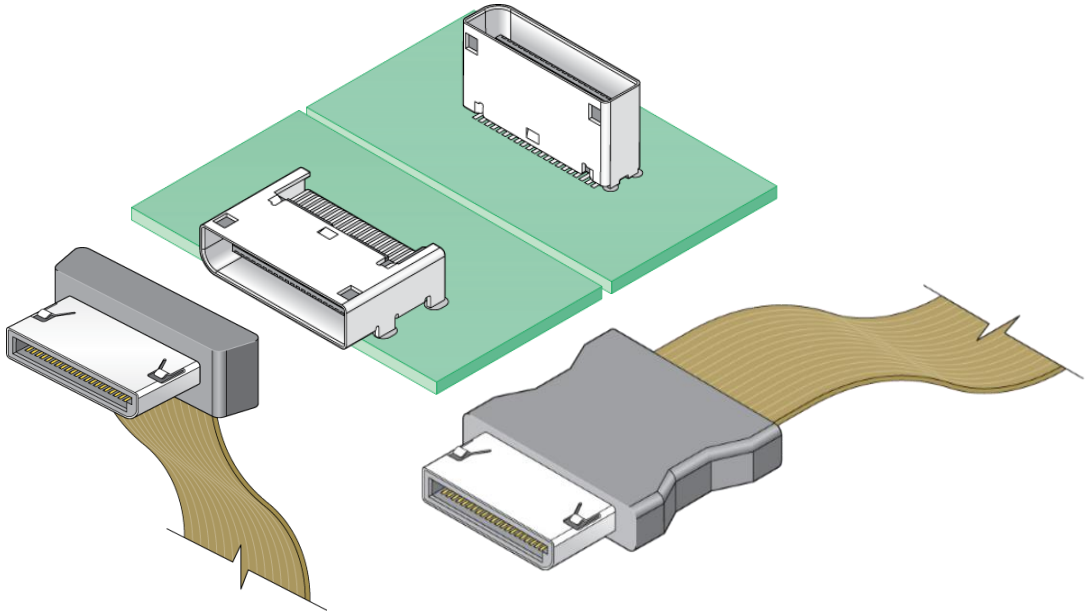


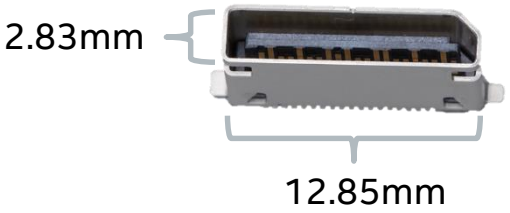
Figure 69 — SAS internal symmetric cable assembly - Mini SAS HD 4i



PCI Express® Cabling for Future Topologies – OCuLink*

Category	OCuLink*
Standard Based	PCI-SIG*
PCIe® Lanes	X4
Layout	Smaller footprint
Signal Integrity	Similar on loss dominated channels
PCIe 4.0 ready	16GT/s target
Clock, power	Supports SRIS and 3.3/5V power
Production Availability	Mid 2015





OCuLink internal cables and connectors

molex® one company › a world of innovation MSX Link

IDF15 INTEL DEVELOPER FORUM