NVMe™ SSD Form Factor Challenges

NVM Express Developers Day – May 1, 2018

Agenda

- **A little bit of history**
- **Summary of potential NVMe™ SSD form factors**
- **System architectural challenges**
  - **Server platforms**
  - **Hyper-scale platforms**
  - **Client platforms**

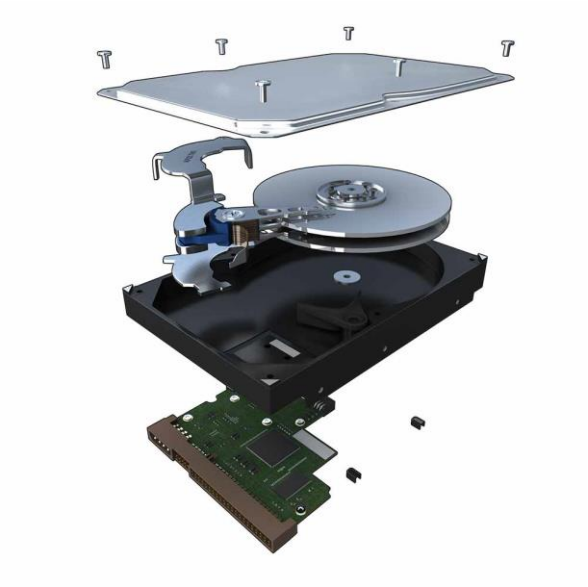
Presenters

–	Bill Lynn	Dell/EMC
–	Michael Krause	HPE
–	Jonathan Hinkle	Lenovo
–	Chris Petersen	Facebook

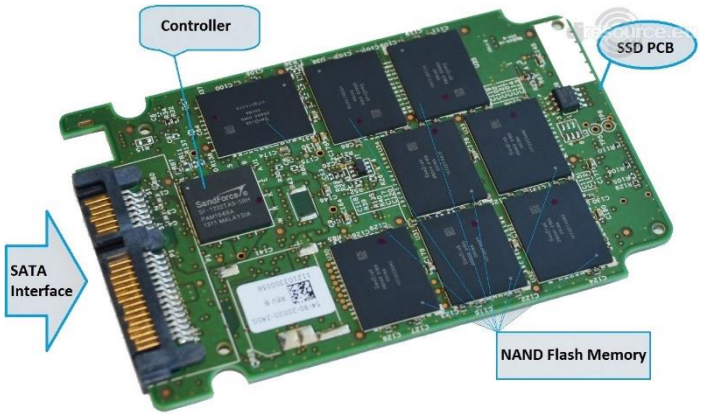
History

HDD's	Form Factor	Introduction		Present
• 1956	Refrigerator	3.75MB		
• 1975	Washing Machine	92MB		
• 1980	5.25"	5MB		
• 1983	3.5"	10MB	→	12TB
• 1988	2.5"	20MB	→	4TB
• 1991	1.8"	21MB	→	250GB
• 1992	1.3"	20MB		
SSD's				
• 2009	mSATA			500GB
• 2013	M.2			1TB

HDD / SSD Architecture

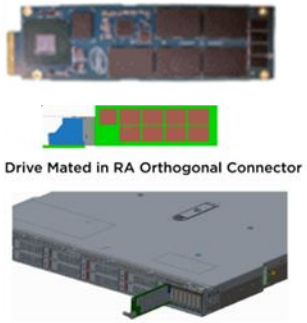


HDD Exploded View

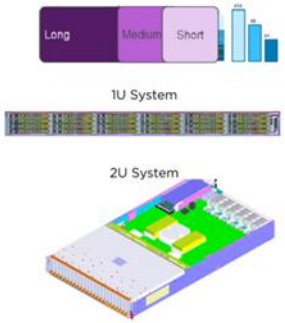


SSD Exploded View

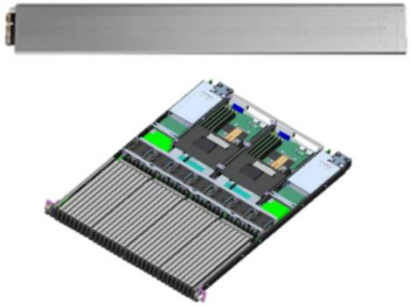
Potential NVMe™ SSD Form Factors



SFF-TA-1006



SFF-TA-1008



SFF-TA-1007

NGSFF

AIC

2.5in U.2

m.2

PCIe M.2

Applicable market segments

- Embedded / T&M, Client, Data center (e.g., SSD carrier)
  - Connector and BGA solutions

Supported Connectors

- x1 / x2: Mid-mount, single-sided, double-sided

M.2 Benefits:

- Large, flexible mechanical form factor family
- Supports multiple protocols: PCIe/NVMe™, SATA, USB, etc.
- Supports 14 vendor-defined pins to customize solutions
- Forward / backward compatibility
- Low power (3.3V & 1.8V using 4 power pins)
- Supports external connectivity: cables, antennae, etc.
- Multiple M.2 can be provisioned on a carrier form factor

U.2 2.5” SSD

Applicable market segments

- Primarily data center with some enthusiast desktop

Size

- 15 x 70 x100.45 mm up to 25W

Supports SFF-8639 (PCIe x4)

U.2 Benefits:

- Adds PCIe support to high-volume 2.5” HDD form factor
- Supports PCIe up to 16 GT/s
- High-capacity / high-density NVMe™ form factor
  - Up to 24 modules in 2U enclosure
  - Up to 14x18 mm packages per module
- Case for ESD protection and hot-plug support

SFF-TA-1006

Applicable market segments

- Data Center (optimized as scalable main storage and acceleration in server and storage systems)

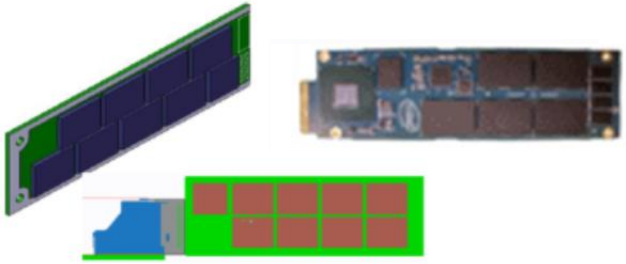
Size

- 31.5 mm x 111.5 mm x 5.75 or 7.55mm, supports up to 12W caseless

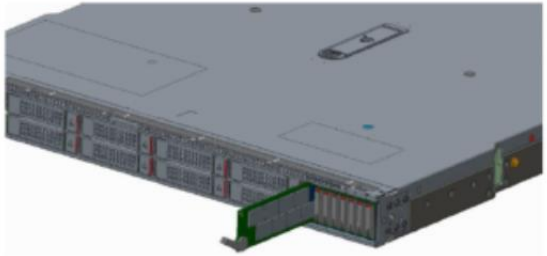
Supports SFF-TA-1002 1C (PCIe x4)

SFF-TA-1006 Benefits:

- PCIe to 32+ GT/s
- High-capacity / high-density NVMe™ form factor
  - Up to 36 modules across 1U rack space
  - Up to 12 14x18 mm packages per module
    - 432 packages / U
- Supports integrated data-centric computation
- Case & Case-less Options



Drive Mated in RA Orthogonal Connector



SFF-TA-1007

Applicable market segments

- Data Center (1U optimized server and storage enclosure)

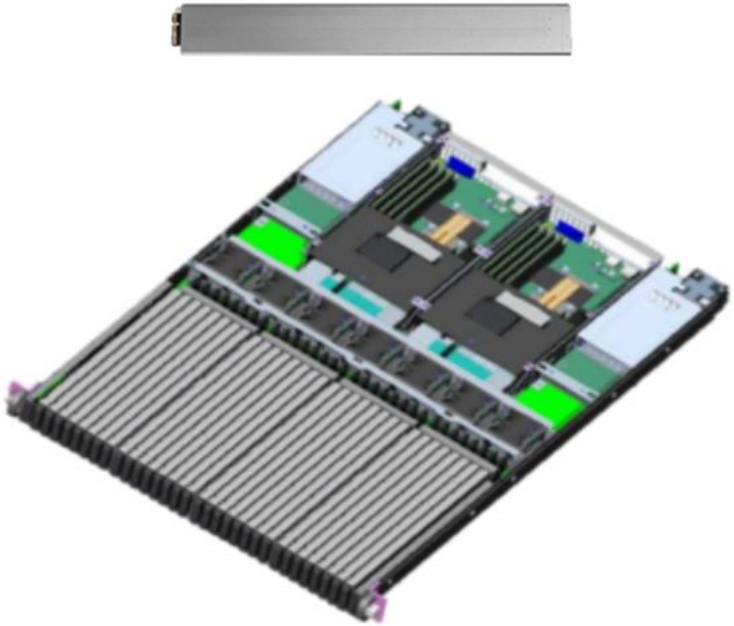
Sizes

- 9.5 x 38.4 x 318.75 mm supports up to 25W
- 18 x38.4 x 318.75 mm supports up to 40W
  - Adds heat sink to improve cooling at system density expense

Supports SFF-TA-1002 2C (PCIe x4-x8)

SFF-TA-1007 Benefits:

- PCIe to 32 GT/s
- High-capacity / high-density NVM form factor
  - 9.5 mm
    - Up to 32 modules
    - Up to 44 14x18 mm packages per module
      - 1408 packages / U
  - 18 mm
    - Up to 18 modules
    - Up to 44 14x18 mm packages per module
      - 704 packages / U
- Supports data-centric computation
  - Reduce packages to free up space for integrated accelerator
- Case for ESD protection, thermal management, & hot-plug



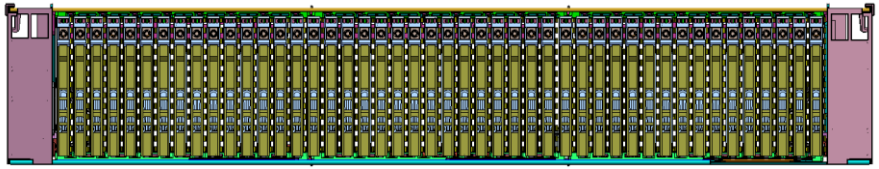
SFF-TA-1008

Applicable market segments

- Data Center (1U and 2U optimized server and storage enclosure)

Sizes

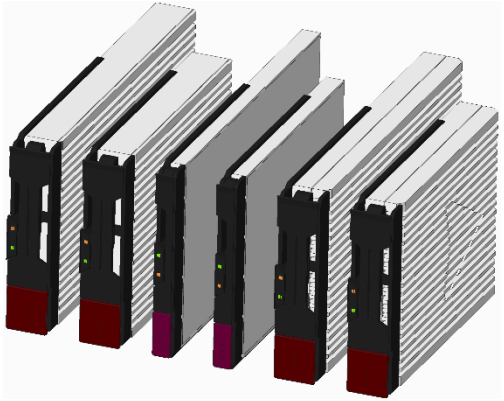
- 7.5 x 76 x 104.9 mm supports up to 25W
- 7.5 x 76 x 142.2 mm supports up to 35W
- 16.8 x 76 x 104.9 mm supports up to 70W
- 16.8 x 76 x 142.2 mm supports up to 70W



Supports SFF-TA-1002 1C, 2C, 4C (PCIe x4-x16)

SFF-TA-1008 Benefits:

- PCIe to 32 GT/s, 802.3 to 112 GT/s
- High-capacity / high-density NVM form factor
  - 7.5 mm
    - Up to 48 modules
    - 104.9 up to 24 14x18 mm packages per module (576 / U)
    - 142.2 up to 48 14x18 mm packages per module (960 / U)
  - 16.8 mm
    - Up to 24 modules
    - Up to 48 14x18 mm flash packages per module (960 / U)
- Supports data-centric computation
  - Reduce packages to free up space for integrated accelerator
- Case for ESD protection and hot-plug support



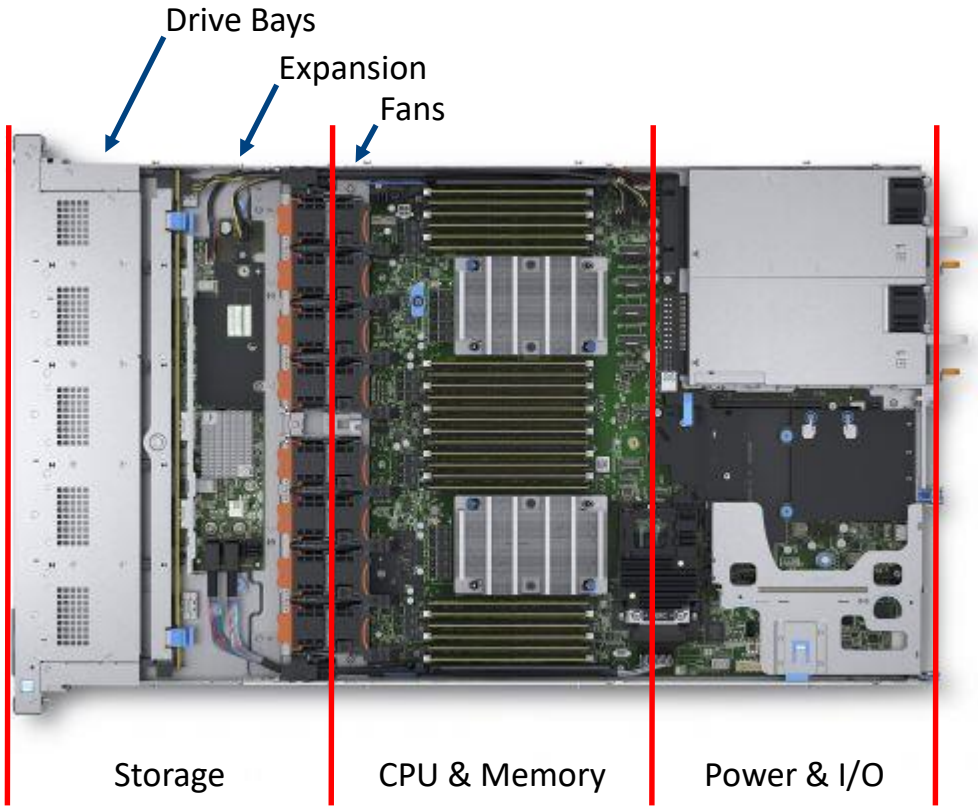
Summary of Proposed Form Factors

Form Factor	Width (mm)	Length (mm)	Thick (mm)	Front Slots	Power (Max)	**Pkgs (14x18)**
2.5” (Baseline)	70	100.45	15	**24**	25	**32**
M.2	22	60/80/110	3.0	**N/A**	8.5	**2/4/8**
SFF-TA-1008* (Short)	76	104.9	7.5	**48**	20	**24**
SFF-TA-1008* (Long)	76	142.2	7.5	**48**	35	**40**
NGSFF*	30.5	110	4.3/4.8	**72**	16	**12**
SFF-TA-1006*	31.5	111.5	5.75/7.55	**72**	12**	**12**
SFF-TA-1007*	38.4	318.75	9.5	**64**	25	**44**

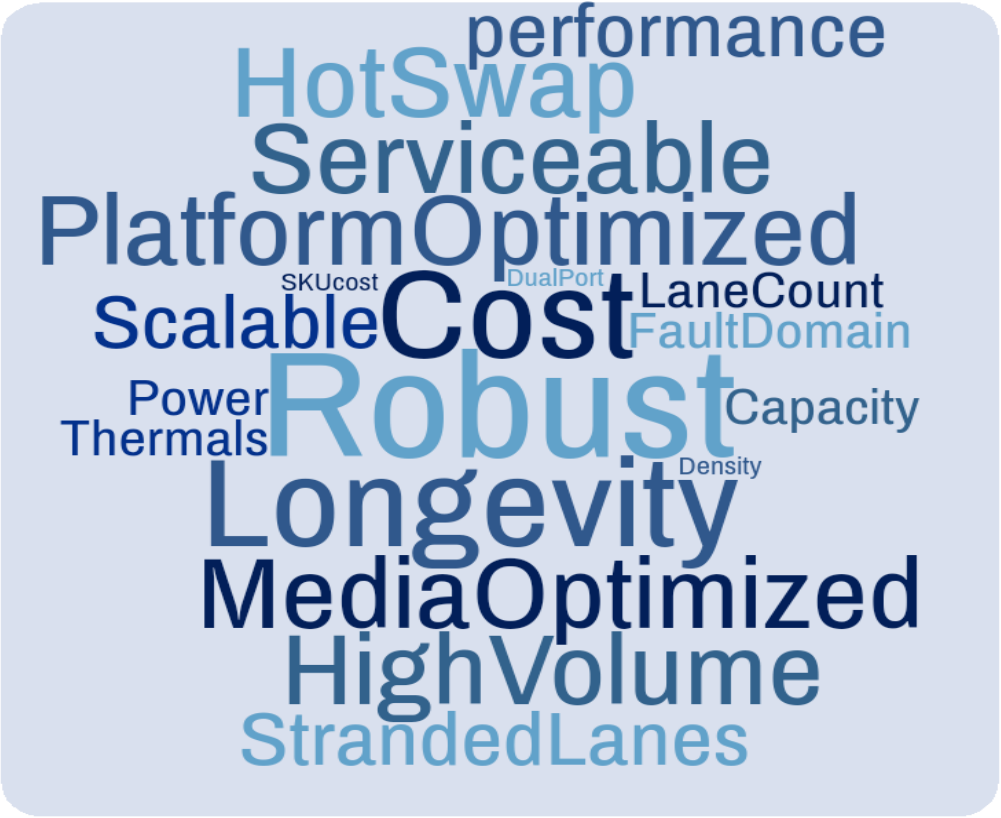
*Newer form factors also support wide versions for higher capacity
**12W TDP supportable for caseless version

Classical 2S Rack Server Architecture

- Typical rack server could be divided into thirds
- Front third typically for storage and supporting infrastructure
- Need to consider:
  - Cooling
  - Power
  - Storage performance
  - Memory performance
  - Networking performance
  - I/O capability

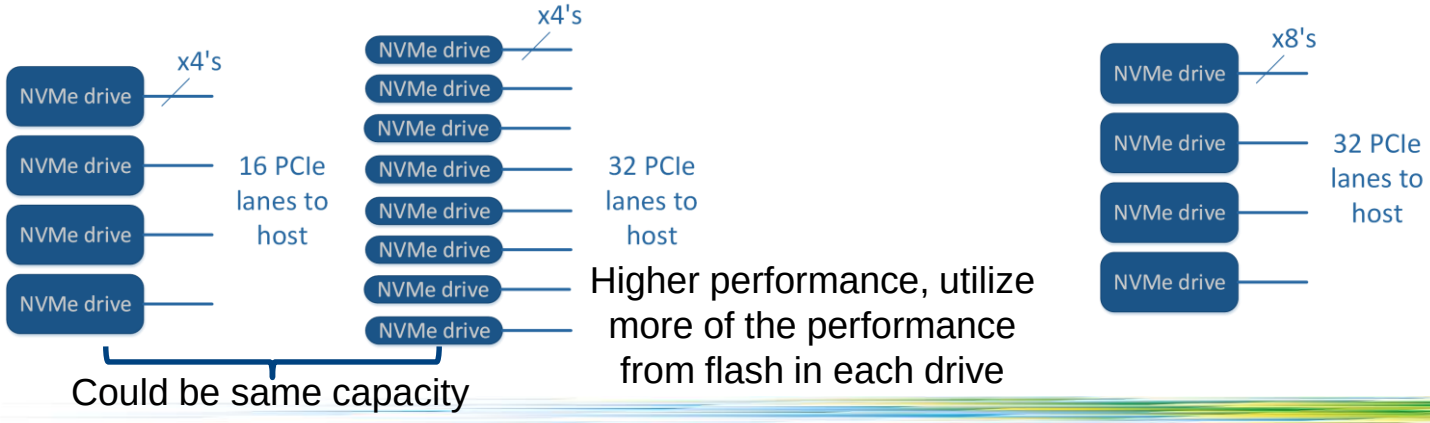


Server Design Considerations



Storage Subsystem Design Considerations

- Must balance drive size and host interface width with performance and capacity
  - PCIe lanes available for NVMe™ storage in systems are going up, but drive performance is bounded by host interface
  - Higher performance PCIe generations (PCIe 3.0 > PCIe 4.0 > PCIe 5.0) may allow drive interface widths to go down (x4 > x2 > x1)
- Scaling storage beyond host PCIe lanes drives significant cost by addition of PCIe switches and additional interconnect



Fewer drives with higher capacity and x8 has no interconnect or switching cost benefits (i.e. 2 x4 = 1 x8)

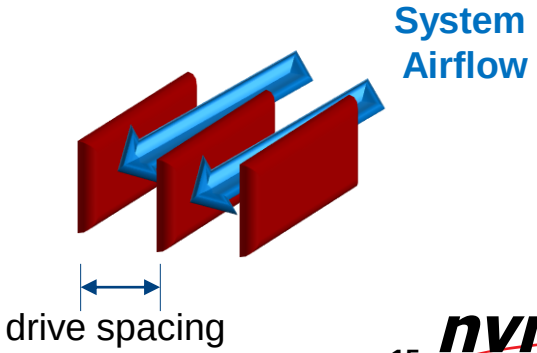
Power and Thermal System Design Considerations

Power delivery

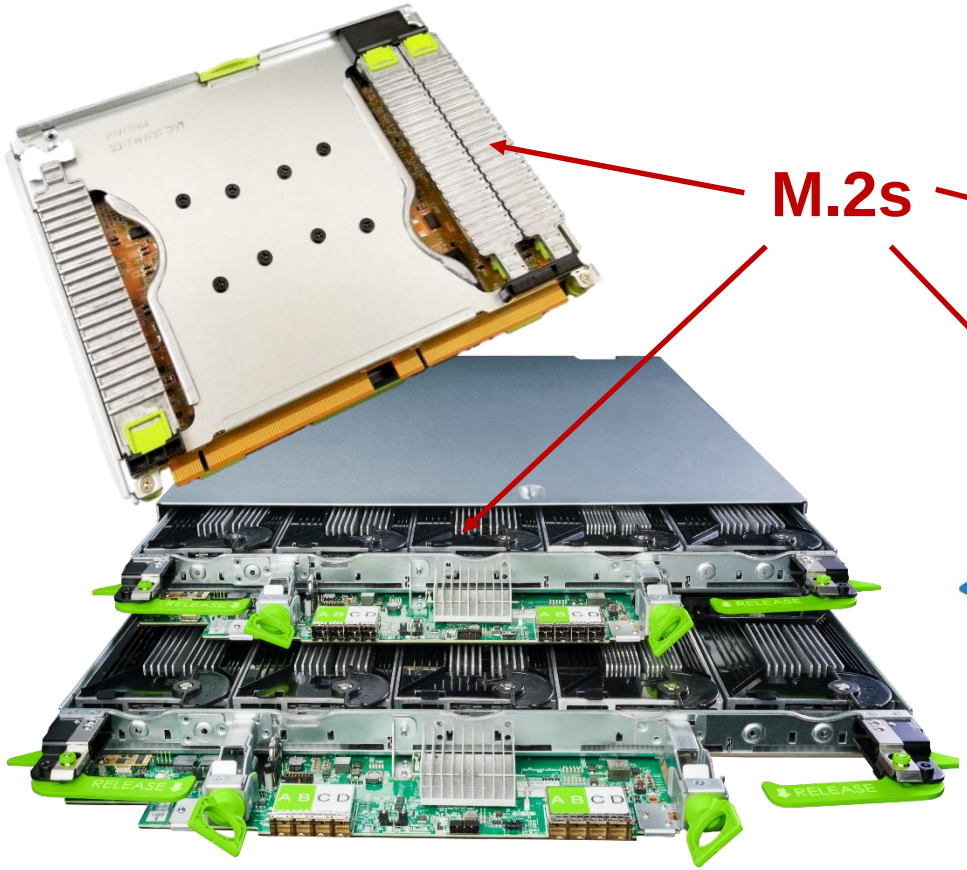
Lower voltage
Higher current density

Higher voltage
Lower current density

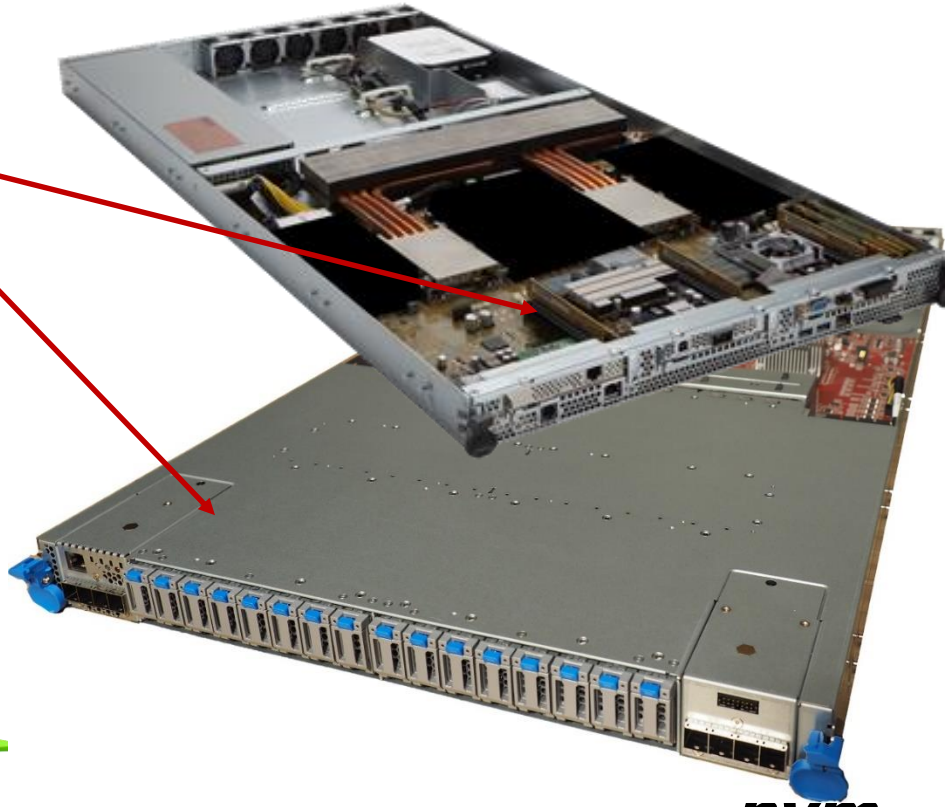
- Drive power
  - Overall system power budget
  - Prefer lower current in system designs: higher voltage like +12V is useful
  - Power density of drive vs. complexity and cost of thermal solution
  - Drive pre-heat vs. keeping downstream system components cool
- Drive dimensions and drive spacing in system
  - Airflow restriction through drives
  - Exposed surface area of drive for heat transfer



How do hyper-scalers use flash today?



M.2s



M.2 Carriers

Hyper-scale NVM Form Factor Characteristics

Important:

- Flexible
- High volume
- Low cost
- Power and thermal efficiency
- Hot swappable
- Serviceable

Less important:

- Backwards compatible
- Support for non-NVM media
- Maximum density

Call to Action: Consolidate to a small number of form factors (<= 4)

Client Challenges

- Size, power, and cost are the biggest issues
- Many OEMs are planning on moving to NVMe™ on client platforms in the near future.
- Dependencies include:
  - Cost
  - Lowering the active idle power

2.5” x 7mm

M.2 2280

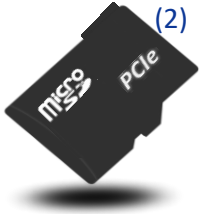
M.2 2230

BGA

Cards beginning to adopt NVMe™

- SDA adding PCIe/NVMe interface to SD card (1)
  - More content (4K/8K content, Social Media)
  - More content generators (Drones, Action Cam, 360, VR
- CFA adding NVMe/PCIe (CFexpress) for high end photography
  - Resolution (push to 8K RAW)
  - Color Gamut – UltraHD Rec 2020
  - Higher Frame Rates (motion blur)
  - Dynamic Range
- Use cases require existing NVMe/PCIe features and will create demand for new features

SD-PCIe (2)



CFexpress

(1) Currently under definition in SDA's Spec WG
(2) The microSD-PCIe card illustration is shown just for this presentation and it does not represent any official SDA standard, yet.

nvm EXPRESS®

Architected for Performance