WHY M.2 IS UNSUSTAINABLE (AND WHAT CAN WE DO ABOUT IT)

LEE PREWITT

PRINCIPAL HARDWARE PROGRAM MANAGER

AZURE CSI - MICROSOFT

M.2 IN AZURE

- Each SKU uses multiple m.2s inside the C2010 Open Compute Chassis
- Each VM gets ~90GB of ephemeral storage space
- Each VM can use up to 4,000 IOPs
  - Any mix of 8KiB IOs (from 100% Random Read to 100% Random Write)

M.2 PROS

- Compact formfactor allows for flexible placement in the chassis
- Ubiquitous use across client and cloud allows for a robust supply chain

- ????

M.2 CONS

- Limited number of NAND die placements (4 with use of PLP)
- Limited power envelope
- Limited thermal envelope
- No hot-plug support
- Connector is not robust
- PCB is thin
  - Increases manufacturing cost and increases risk of PCB failures

So is there a better solution?

WHAT ARE THE OPTIONS?

Adapter Cards

Good: High-perf Cache

Bad: Steals PCIe slots from I/O

Ugly: $$$, Might fit only two?

Consumer SSD

Good: Small and Modular

Bad: Low capacity, no hot-plug, connector less reliable

Ugly: More expensive once adapted to enterprise

HDD Form Factors

Good: Hot-plug, Storage features

Bad: Designed for HDD, not SSD

Ugly: Blocks airflow to the hottest components in server

EDSFF: A BRIEF HISTORY

- Enterprise and Datacenter SSD Form Factor
- Formed in 2017 to rethink the SSD from the ground up with a focus on Cloud and Enterprise use cases

Promoters

DELL EMC · facebook · Hewlett Packard Enterprise · intel · Lenovo · Microsoft · SAMSUNG

Contributors

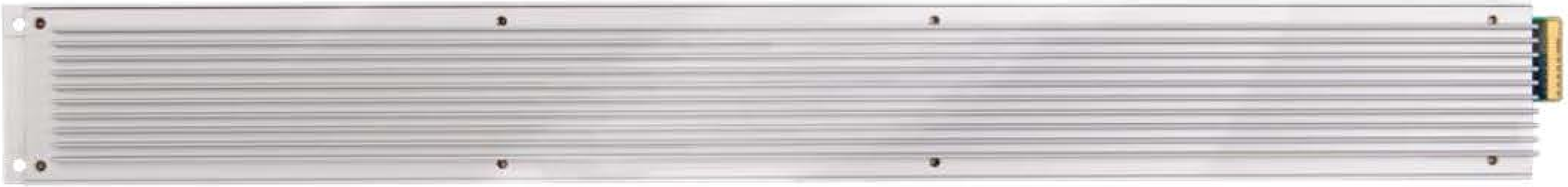
Amphenol · FOXCONN · Micron · molex · SEAGATE

Toshiba Memory · TE connectivity · Western Digital

OF COURSE THERE CAN NEVER BE ONLY ONE…

E1.L (SFF-TA-1007)

- Density Optimized
- 318.75 x 38.4 mm
- Supports > 40W
- Up to 48 Standard NAND sites



E1.S (SFF-TA-1006)

- 111.5 x 31.5 mm
- Up to 12 Standard NAND sites
- Supports >12W

E3 (SFF-TA-1008)

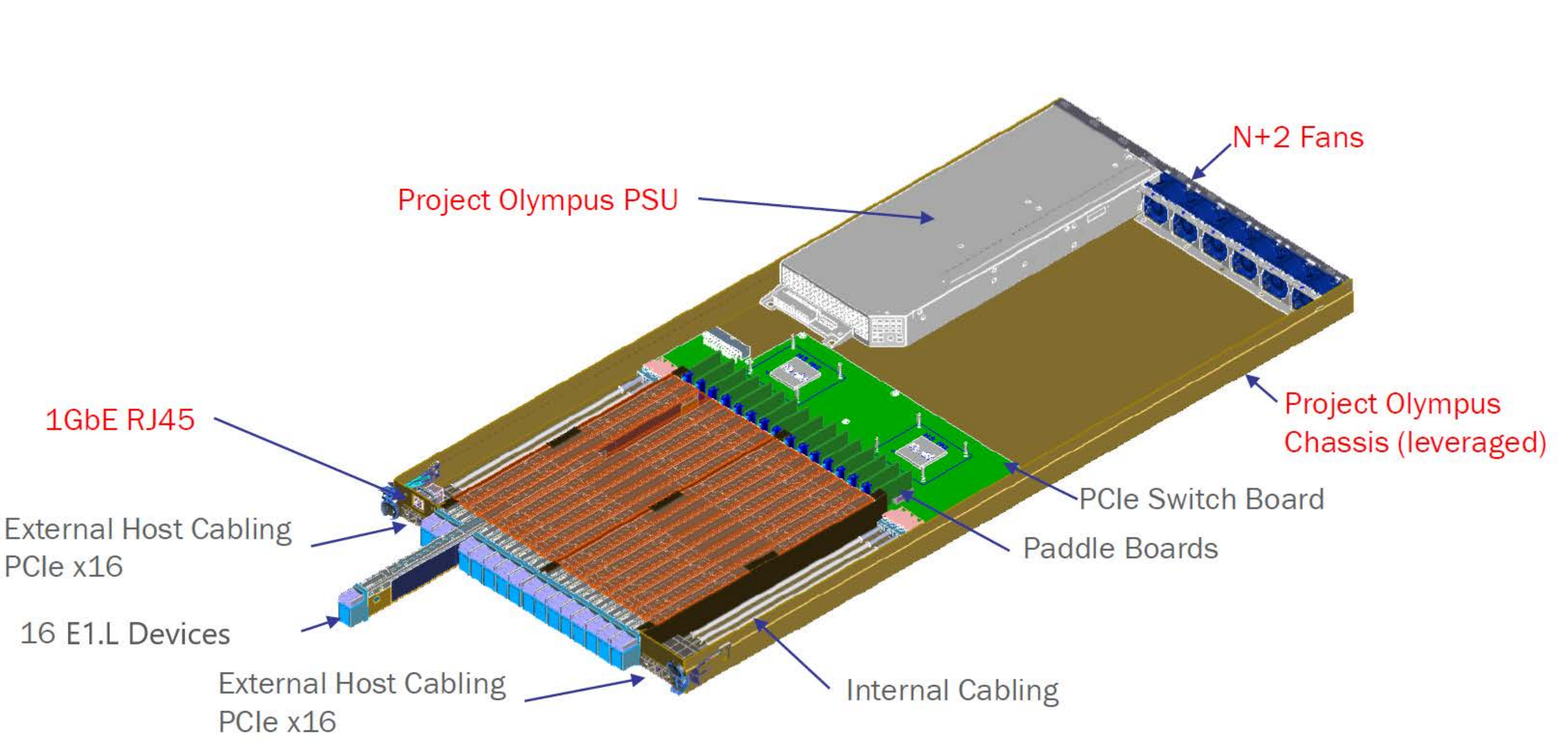
- Ultra high-performance applications
- (104.9/142.2) x 78mm
- Supports up to 70W
- Up to 48 Standard NAND sites

EDSFF Advantages

- **Same** Protocol: NVMe
- **Same** Interface: PCIe
- **Same** Connector: SFF-TA-1002
- **Same** Pinout and Functions

Different Usages
Same Expectations!

E1.L FOR AZURE STORAGE



E1.S FOR AZURE COMPUTE

Enclosure Parameter	5.9mm Device	Device with Heat Spreader (8.01mm)	Device with Symmetric Enclosure (9.5mm)	Device with Asymmetric Enclosure (25mm)
Recommended sustained power (W)	12	16	20	25
Touch point Temperature limit (° C)	80	80	80	80
Enclosure Max Inlet air temperature, < 950 m (° C)	35	35	35	35
Enclosure Max Inlet air temperature, 950 m to 3050 m (° C)	35 - (1° C for 175 m of elevation gain)	35 - (1° C for 175 m of elevation gain)	35 - (1° C for 175 m of elevation gain)	35 - (1° C for 175 m of elevation gain)
Add in card to add in card pitch (mm)	9	11	13	26
Recommended Fan Pressure loss across device (Pascal)	83	52	64	21
Airflow, average min per device (CFM). 1 CFM = 1.7 m^3/h)	1.41 – (0.01 CFM for every 1° C below 35° C inlet temp)	1.71 – (0.06 CFM for every 1° C below 35° C inlet temp)	2.02 - (0.02 CFM for every 1° C below 35° C inlet temp)	4.10 - (0.04 CFM for every 1° C below 35° C inlet temp)

1 **Table 7-1. Thermal guidelines for a 1U short system implementation**