
Open. Together.
OCP
SUMMIT

STORAGE

Learning of Designing Project Olympus JBOF with EDSFF SSD

Mark A Shaw
Principal Hardware Engineer Manager, Microsoft
Antson Chang
Assistant Technical Director, Wiwynn Corporation

OPEN PLATINUM™

OPEN Compute Project SOLUTION PROVIDER®

Agenda

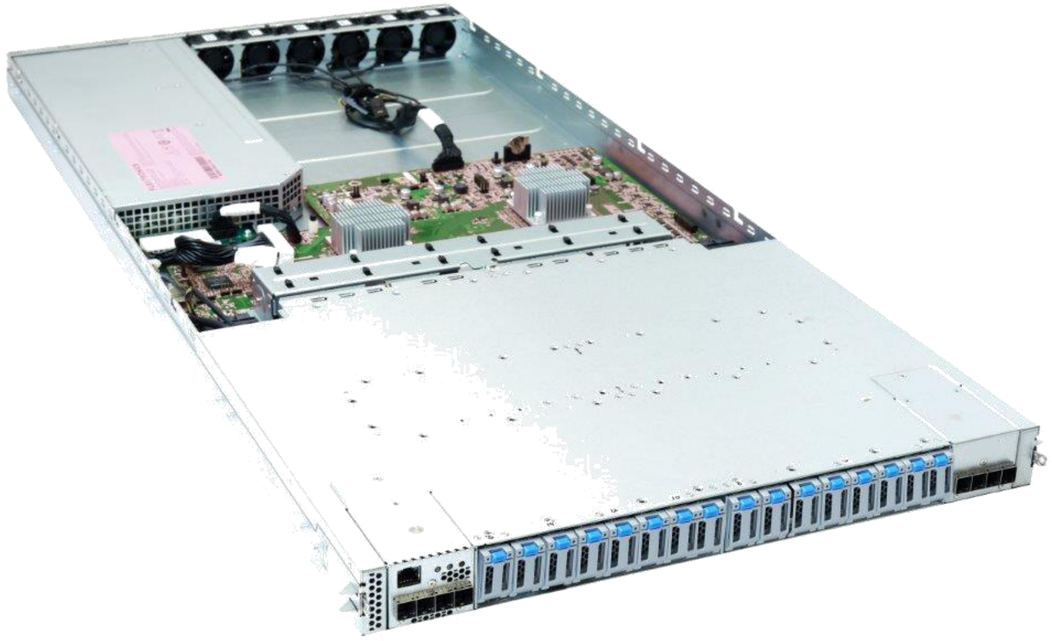
STORAGE

- Project Olympus FX-16 JBOF
- Design tips
- Power consumption and thermal efficiency
- Conclusion

Workshops
Summits

Summary of FX-16

Device	a. 1U JBOF b. Total capacity is 256TB (or higher)
I/O	a. 2 PCIe x16 uplinks in mini-SAS HD form factor running at PCIe gen3 b. 1 Ethernet port for management
SSD slot	a. 16 slots and per slot supports up to x8 b. EDSFF 1U Long (E1.L), 18mm width
SSD	a. Hot-pluggable and front loaded b. Multi-sourced
Power consumption	<400 watts



Storage Trends

- Performance

Fast

Type	Interface	Read (IOPs)	Write (IOPs)	Read (GB/s)	Write (GB/s)
HDD	SAS/SATA	200	200	0.15	0.15
SSD	SATA	95K	25K	0.5	0.5
	SAS	420K	50K	2.2	1.6
	PCIe	750K	180K	3.5	3.0

- Cost

Expensive

Media Price ($/GB)	2017	2018	2019	2020
Business critical HDD	$0.026	$0.023	$0.019	$0.017
Mission critical HDD	$0.117	$0.105	$0.093	$0.087
Data center SSD	$0.45	$0.26	$0.17	$0.13

Source: IBM Murali Lyer's presentation slide: NVMe Overview

- Density

Year	Density per 1U (Tera-Byte)	
	HDD	PCIe (NVMe) SSD
2018	210-350	64-128
2019	245-395	256-512
2020	280-450	512-1024
Note	2018: 12-16TB per HDD unit, 4U/70Bay: 840-1120TB 4U/90Bay: 1080-1400TB 2019: 14-18TB per HDD unit, 4U/70Bay: 980-1260TB 4U/90Bay: 1260-1580TB 2020: 16-20TB per HDD unit, 4U/70Bay: 1120-1400TB 4U/90Bay: 1440-1800TB	2018: 4TB per EDSFF SSD, 1U/16Bay: 64TB 1U/32Bay: 128TB 2019: 16TB per EDSFF SSD, 1U/16Bay: 256TB 1U/32Bay: 512TB 2020: 32TB per EDSFF SSD, 1U/16Bay: 512TB 1U/32Bay: 1024TB

Trade-off for what Data Center really wants!

- System and SSD features we want……
  - Hot-pluggable
  - Balanced airflow impedance and thermal efficiency
  - Total capacity
  - Flexibility for high performance/density SSDs

  ……etc.

- Put several M.2 together as SSD group?
- U.2 => M.2 carrier => EDSFF SSD

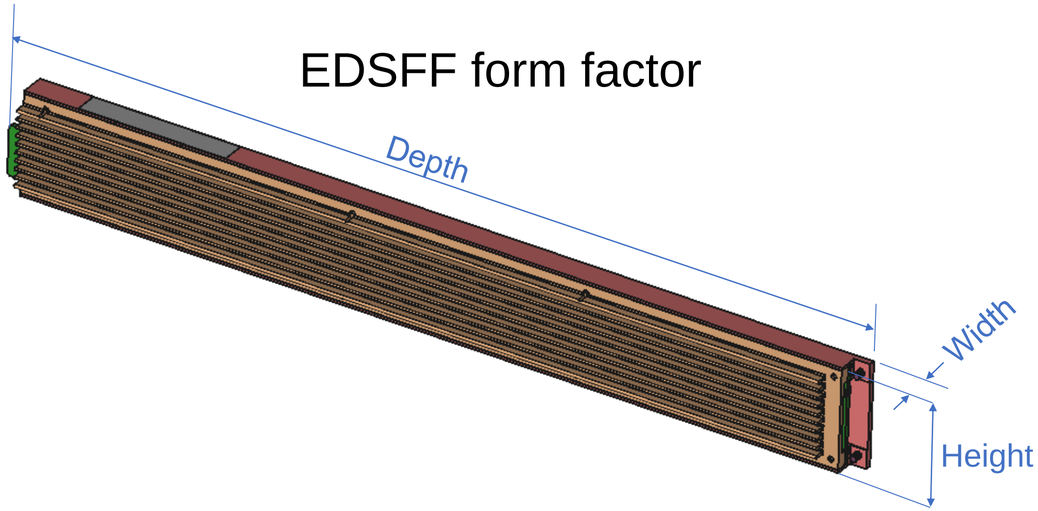
New SSD form factor: EDSFF

- *In the end of 2016*, Wiwynn worked with Microsoft on *M.2 Carrier*. In the same time, Samsung (M.3/NF1/NGSFF) and Intel (Ruler SSD) proposed similar form factors.
- *In the first half of 2017*, Intel invites Microsoft, Lenovo, Facebook, Dell/EMC, Samsung, HPE …etc. to join *EDSFF* (Enterprise & Datacenter SSD Form Factor)
- *In 2018 Q1*, Olympus FX-16 first proto-type supporting 16 E.1L 18mm SSDs was demonstrated at OCP Summit.
- *In 2019 Q2*, FX-16 is *ready for mass production*.

Dimension defined in EDSFF

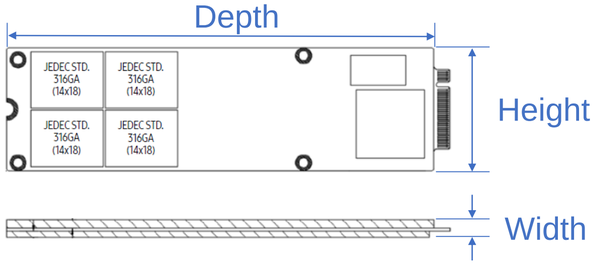
SSD Form factor		Dimension (mm)			Maximum Power (watt) (Recommended)	Specification
		Height	Width	Depth		
EDSFF, 1U (x4/x8)	Short	31.5	5.9	111.49	12	SFF-TA-1006
	Short	31.5	8.01	111.49	16	SFF-TA-1006
	Long	38.4	9.5	318.75	25	SFF-TA-1007
	Long	38.4	18	318.75	40	SFF-TA-1007
EDSFF, 3 inch (x4/x8/x16)	Short	76	7.5	104.9	20	SFF-TA-1008
	Short	76	16.8	104.9	40	SFF-TA-1008
	Long	76	7.5	142.2	35	SFF-TA-1008
	Long	76	16.8	142.2	70	SFF-TA-1008
Other existing SSD form factor						
U.2 (2.5”)		70	7/15	100	25	
M.2		22	4.3	110	8.25	PCI-sig
NGSFF (M.3)		30.5	4.3/4.8	110	11.5	Samsung (without cage)

EDSFF form factor



U.2 form factor

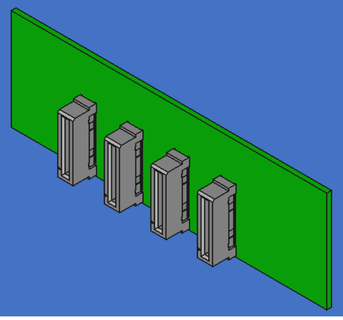
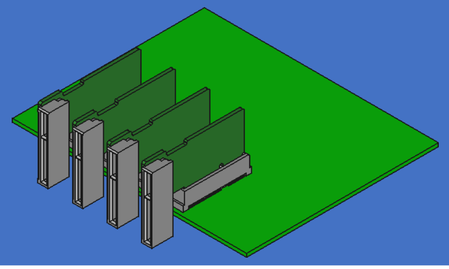
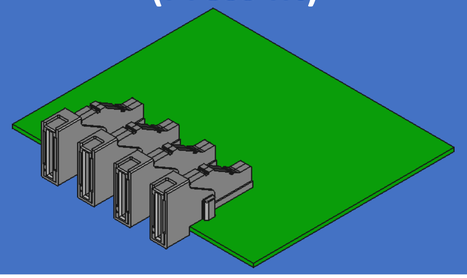
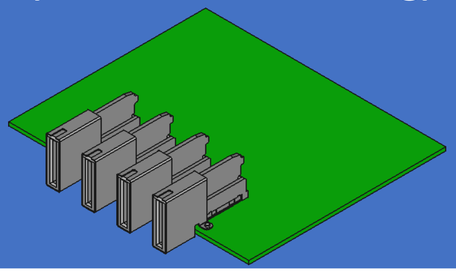
M.2 or NGSFF form factor



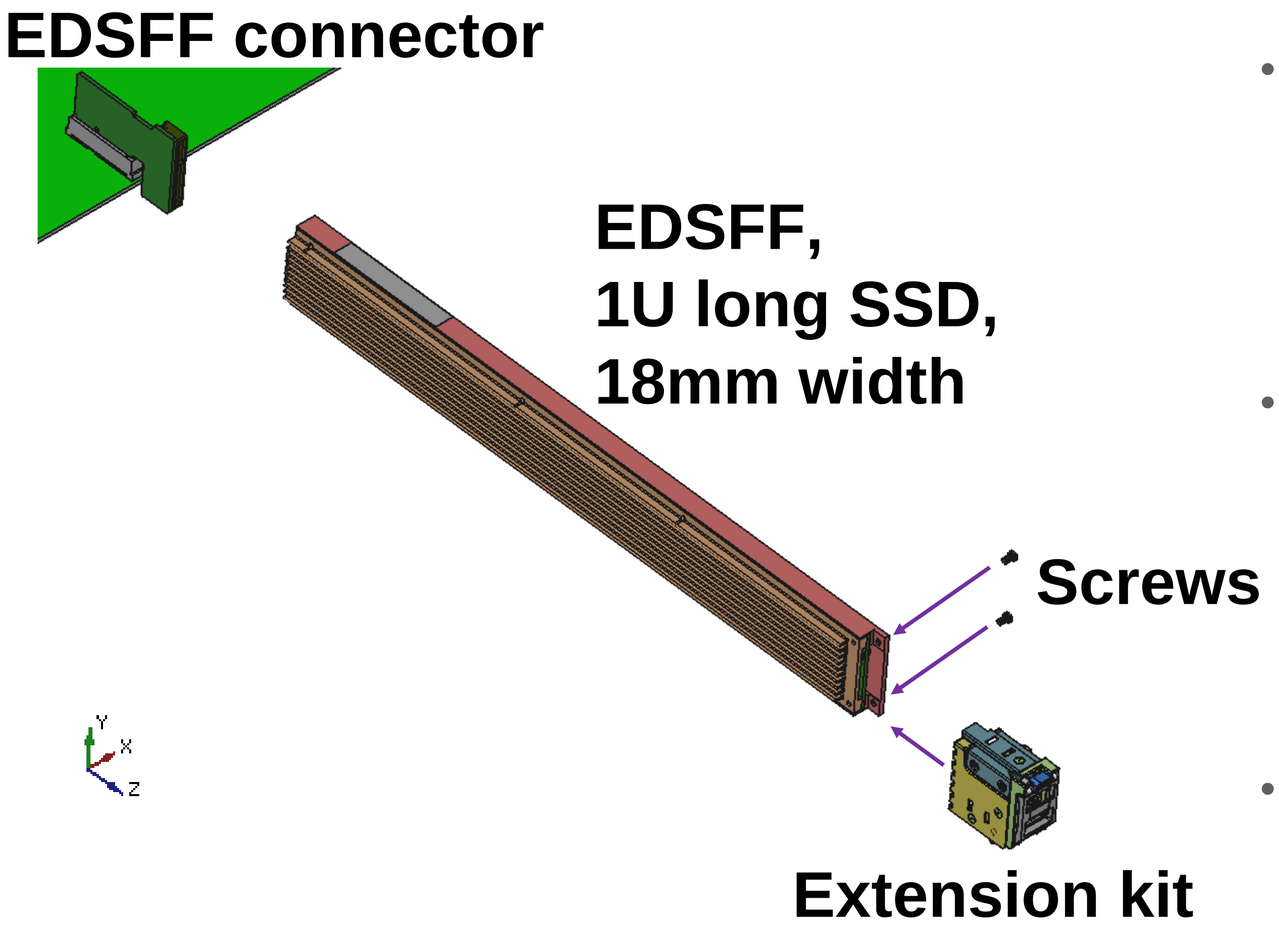
Agenda

- Project Olympus JBOF FX-16
- **Design tips**
- Power consumption and thermal efficiency
- Conclusion

System Implementation with EDSFF

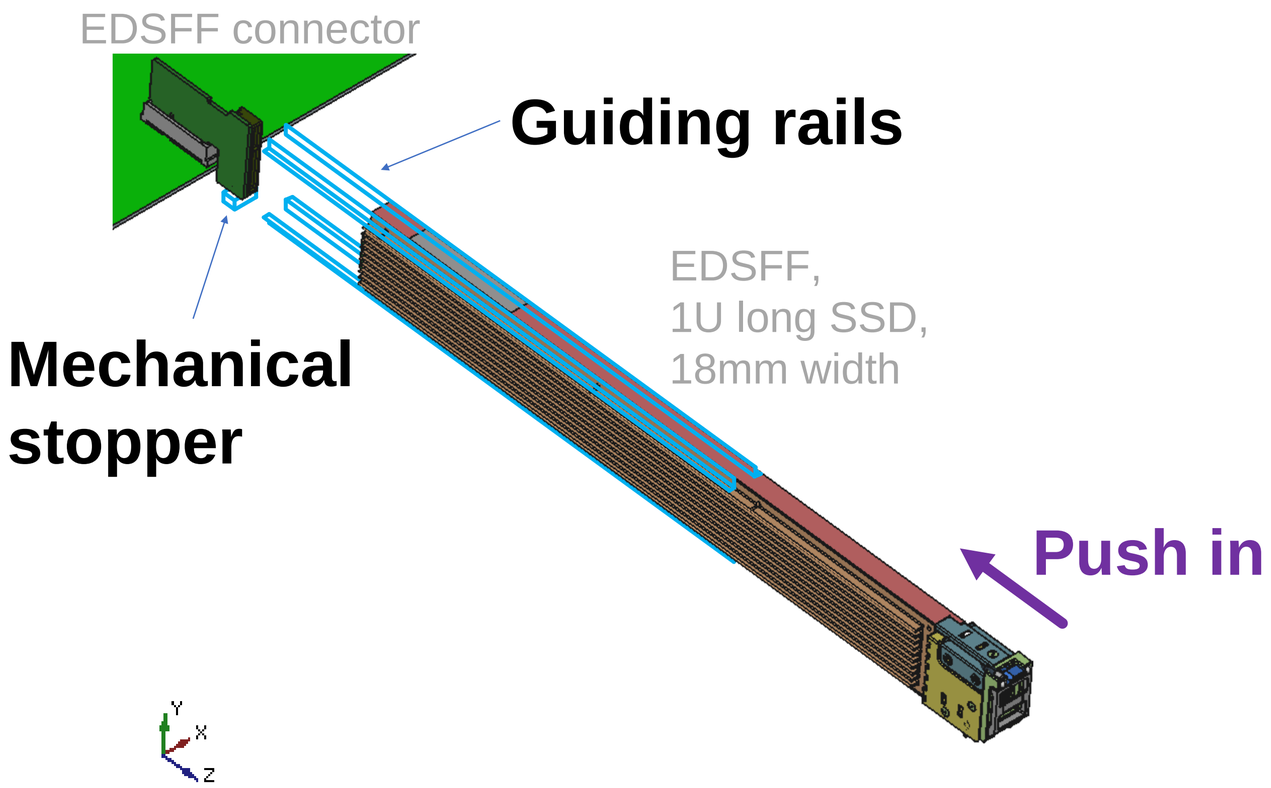
	Backplane	Paddle Card	Orthogonal conn. (Press fit)	Orthogonal conn. (SMT with assembling)
				
Pros.	• Support up to PCIe x16 in 2U chassis.	• Clear airway enhances thermal efficiency. • SMT connectors. Easier placement and SI routing • PCIe Gen4/5 support • Connector availability & sourcing	• Clear airway enhances thermal efficiency. • Less complicated mechanical design • Easier assembly process	• Clear airway enhances thermal efficiency. • SMT connectors. Easier placement and SI routing • PCIe Gen4/5 support • Screw fixing makes system mechanical design easier.
Cons.	• More restricted airflow in 1U. • To connect with MB, several high-speed connectors or cables are needed.	• More complicated mechanical design • More complicated assembly process	• New (2018 Q4). • Hard to place components next to connector. • Back-drill is needed to support up to PCIe Gen4/5. • Connector Sourcing	• New (2019 Q2).

Mechanical Design with EDSFF



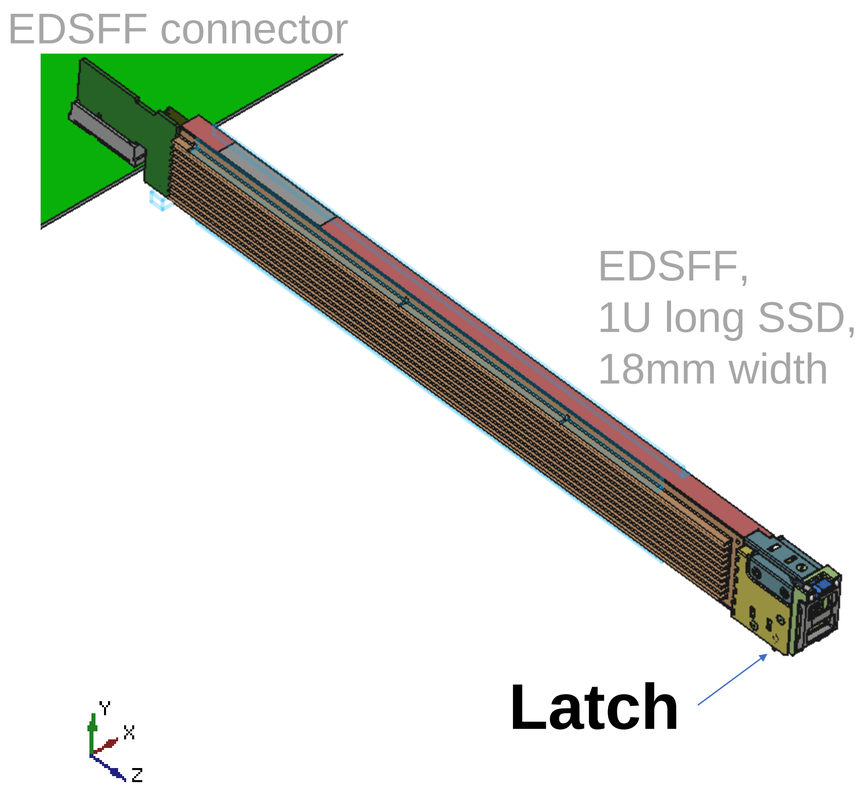
- **E1.L supports customized extension kit to aligned with existing panel looking.**
- For blind mating
  - Guiding rails
  - Mechanical stopper
  - Latch
- Shelter blocks airflow while SSD is not populated.

Mechanical Design with EDSFF



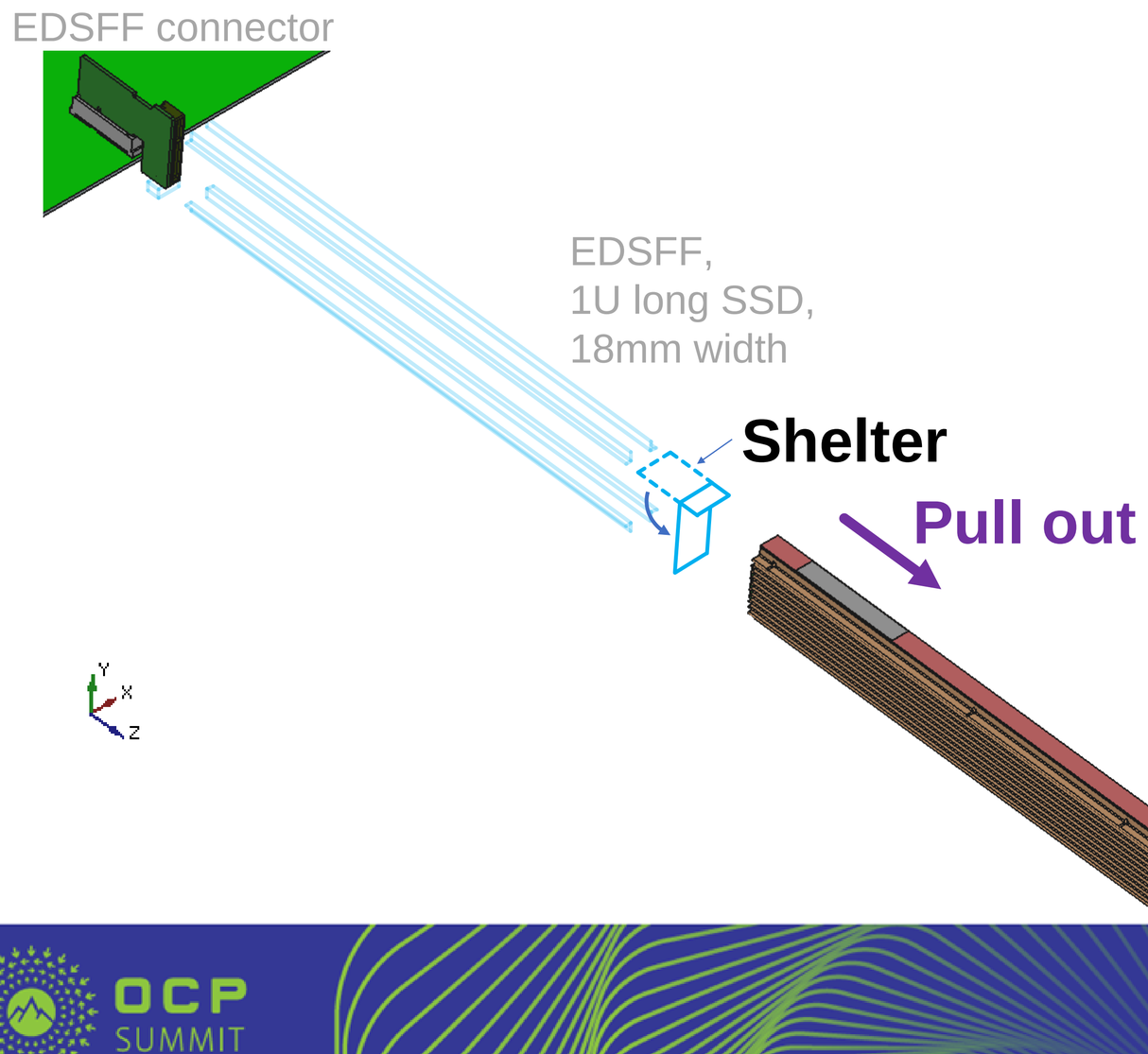
- E1.L supports customized extension kit to aligned with existing panel looking.
- **For blind mating**
  - **Guiding rails**
  - **Mechanical stopper**
  - Latch
- Shelter blocks airflow while SSD is not populated.

Mechanical Design with EDSFF



- E1.L supports customized extension kit to aligned with existing panel looking.
- **For blind mating**
  - Guiding rails
  - Mechanical stopper
  - **Latch**
- Shelter blocks airflow while SSD is not populated.

Mechanical Design with EDSFF



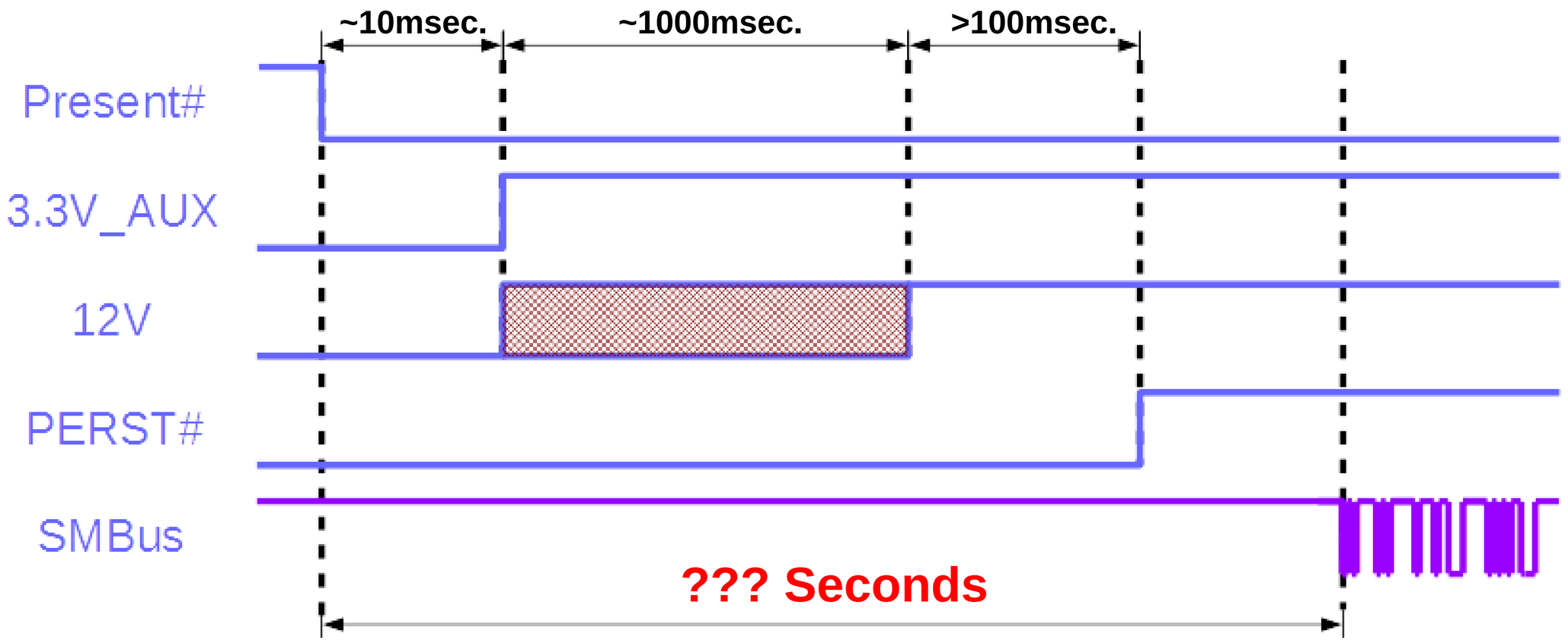
- E1.L supports customized extension kit to aligned with existing panel looking.
- For blind mating
  - Guiding rails
  - Mechanical stopper
  - Latch
- **Shelter blocks airflow while SSD is non-populated.**

Electrical Design for hot-plugging

Function	Circuitry
Isolation between slots	Suggest 100% power and signal isolation between slots to ensure that there is no power or signal interaction between slots. Insertion or removal of one slot cannot affect the functionality of other slots.
Power enabling	MOS + soft start (or hot-swap controller)
Inrush current protection	Extra logics for power enabling based on presence pin Note: There is no pre-charge pin in EDSFF pin definition.
Over Current Protection	Polyfuse (or hot-swap controller)

SMBus Access

- How much time is needed for SMbus access after plugging?



Dual port implementation

- *Clock and PERST# input* for Port B are needed.
- Pull-down DualPortEN# (B9 pin) by host *before enabling SSD 12V power.*
- Make sure *which lanes* connect to Port B

PCIe lanes	x4 Dual Port (2 ports x2)	x8 Dual Port (2 ports x4)	x16 Dual Port (2 ports x8)
PERp0, PERn0, PETp0, PETn0	Port A, lane 0	Port A, lane 0	Port A, lane 0
PERp1, PERn1, PETp1, PETn1	Port A, lane 1	Port A, lane 1	Port A, lane 1
PERp2, PERn2, PETp2, PETn2	Port B, lane 0	Port B, lane 0	Port B, lane 0
PERp3, PERn3, PETp3, PETn3	Port B, lane 1	Port B, lane 1	Port B, lane 1
PERp4, PERn4, PETp4, PETn4	No connect	Port A, lane 2	Port A, lane 2
PERp5, PERn5, PETp5, PETn5	No connect	Port A, lane 3	Port A, lane 3
PERp6, PERn6, PETp6, PETn6	No connect	Port B, lane 2	Port B, lane 2
PERp7, PERn7, PETp7, PETn7	No connect	Port B, lane 3	Port B, lane 3
PERp8, PERn8, PETp8, PETn8	No connect	No connect	Port A, lane 4
PERp9, PERn9, PETp9, PETn9	No connect	No connect	Port A, lane 5
PERp10, PERn10, PETp10, PETn10	No connect	No connect	Port B, lane 4
PERp11, PERn11, PETp11, PETn11	No connect	No connect	Port B, lane 5
PERp12, PERn12, PETp12, PETn12	No connect	No connect	Port A, lane 6
PERp13, PERn13, PETp13, PETn13	No connect	No connect	Port A, lane 7
PERp14, PERn14, PETp14, PETn14	No connect	No connect	Port B, lane 6
PERp15, PERn15, PETp15, PETn15	No connect	No connect	Port B, lane 7

Agenda

- Olympus JBOF FX-16
- Design tips
- **Power consumption and thermal efficiency**
- Conclusion

Power and Thermal

Stress tool (Diskspd)	Block size (Byte)	Ambient 25℃				Ambient 35℃			
		System fans (6pcs) PWM / Air flow / Power	System power (watts)	$\frac{CFM^{*}}{kW}$	$\frac{Fan\ power}{System\ Power}$	System fans (6pcs) PWM / Air flow / Power	System power (watts)	$\frac{CFM^{*}}{kW}$	$\frac{Fan\ power}{System\ Power}$
100% Sequential Write	512K	36% / 23.4 CFM / **6.50 W**	341	69	1.91%	54% / 37.2 CFM / **14.50 W**	343	108	4.23%
	4K	26% / 16.3 CFM / 4.00 W	249	65	1.61%	44% / 29.5 CFM / 9.75 W	255	115	3.82%
50% Sequential Write/Read	512K	34% / 22.0 CFM / 5.75 W	315	70	1.83%	51% / 34.9 CFM / 13.00 W	317	110	4.10%
	4K	24% / 14.9 CFM / 3.75 W	212	70	1.77%	41% / 27.1 CFM / 8.50 W	218	124	3.90%
100% Sequential Read	512K	22% / 13.6 CFM / 3.50 W	193	60	1.81%	35% / 22.7 CFM / 6.00 W	179	127	3.35%
	4K	22% / 13.6 CFM / 3.50 W	162	84	2.16%	32% / 20.5 CFM / 5.00 W	163	126	3.07%

*: Microsoft specification is less than 158 CFM/kW

Agenda

- Olympus JBOF FX-16
- Design tips
- Power consumption and thermal efficiency
- **Conclusion**

Conclusion

- EDSFF provides:

  - front access and hot pluggable capability,

  - high density system design

  - efficient thermal design

  - Scalable options for x4, x8 or x16

  - Single or dual port feature

  - Same connector definition supporting PCIe Gen3/4/5

Join us to use EDSFF for the next decade for SSDs!


Open. Together.
OCP Global Summit | March 14–15, 2019
OCP
SUMMIT

Agenda

- **Project Olympus FX-16 JBOF**
  - **Summary of FX-16,**
  - **Storage Trends,**
  - **Trade-off for what Data Center really wants!**
  - **New SSD form factor: EDSFF, and it's dimension definition**
- Design tips
- Power consumption and thermal efficiency
- Conclusion

Agenda

- Project Olympus JBOF FX-16
- **Design tips**
  - **System implementation with EDSFF,**
  - **Mechanical and Electrical design,**
  - **SMbus access,**
  - **Dual-port implementation**
- Power consumption and thermal efficiency
- Conclusion