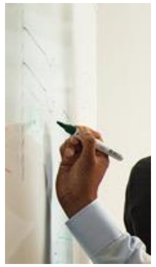
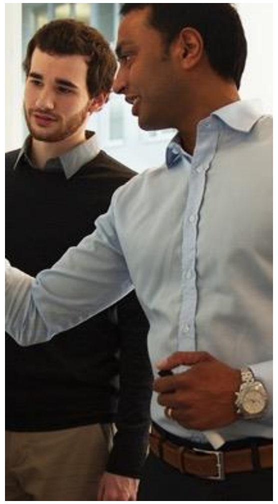
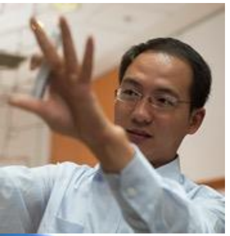






Examining System Challenges When Implementing Next Generation Data Center Input/Output (I/O) Connectivity

Nathan Tracy

1/25/18

EVERY CONNECTION COUNTS

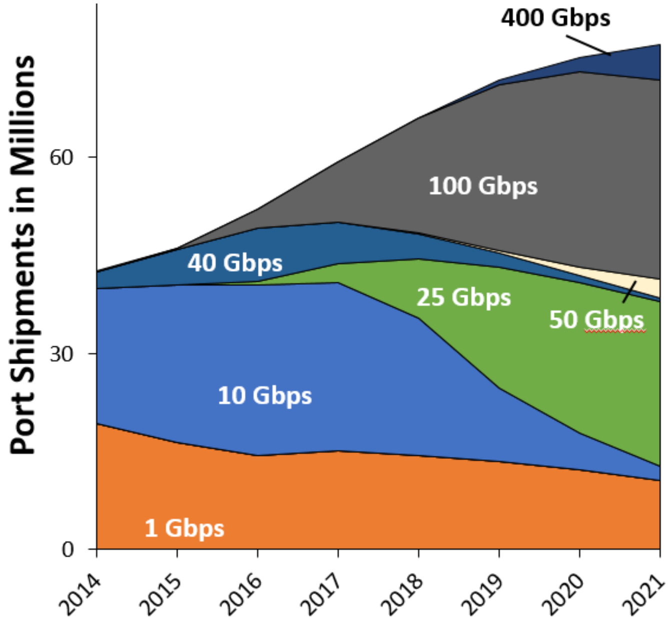
Agenda

- **Trends / Needs in Switching**
- **Challenges**
- **Next Gen I/O**
- **Equipment Impact**
    - Density
    - Electrical Performance
    - PCB Issues
    - Reach
    - Thermal Management
    - Air Flow
- **Summary**
- **Conclusions**

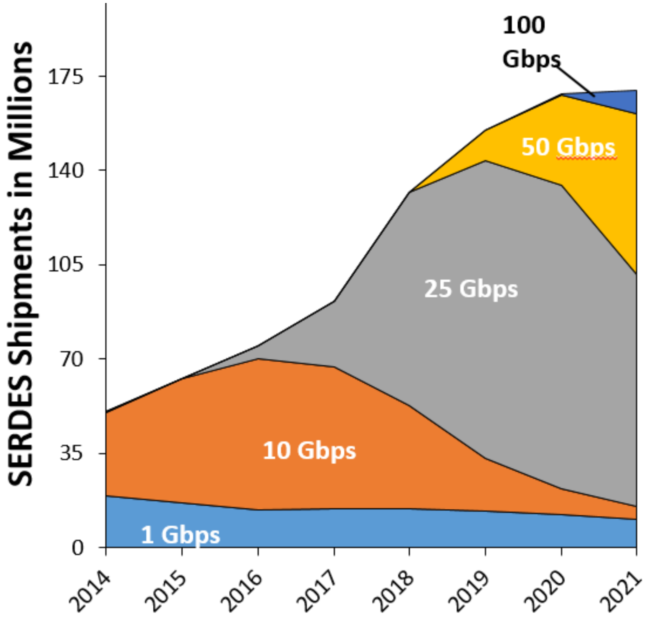
Industry Need/Trends - Bandwidth

- Datacom industry has a relentless thirst for more bandwidth. Many bottlenecks have to be overcome to quench that thirst

Ethernet Switch –
Data Center Total Port Shipments



Ethernet Switch –
Data Center Total SERDES Shipments



Ethernet switch port counts in data centers, more ports at higher speed = total bandwidth

SERDES shipments to data centers: rates have to increase to keep up with switch density and overall bandwidth

Data Courtesy of :

650 GROUP
MARKET INTELLIGENCE RESEARCH

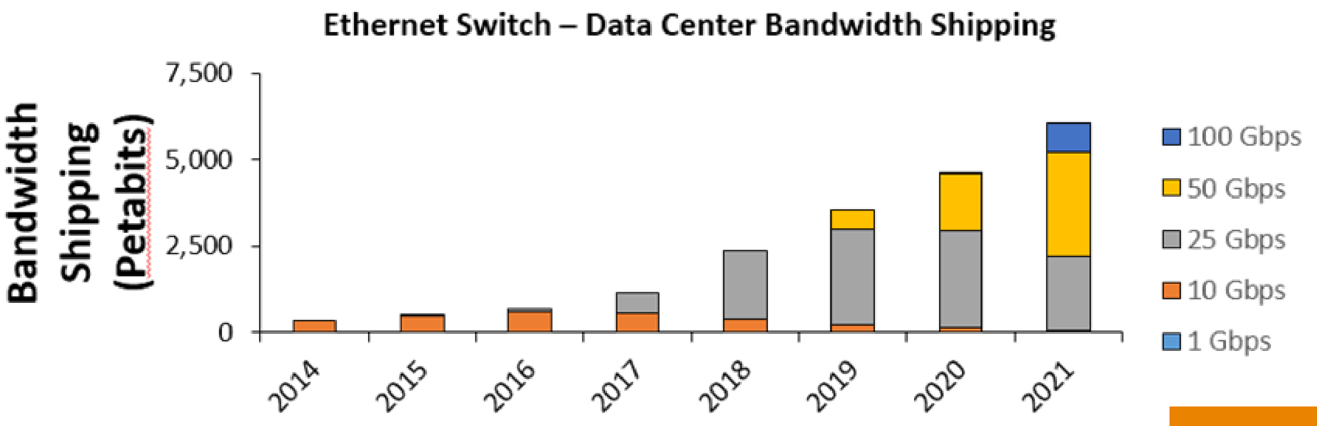
Merchant Silicon – Data Center Switching:

ASIC Usage in the Tier 1 Cloud

Merchant Silicon's product cycles accelerating in the Cloud

ASIC Size	SERDES Technology	Spine/Core Port Speed	2012	2013	2014	2015	2016	2017	2018	2019	2020	>2020
1.3 Tbps	10 Gbps	10/40 Gbps										
3.2 Tbps	25 Gbps	100 Gbps										
6.4 Tbps	25 Gbps	100/200 Gbps										
12.8 Tbps	50 Gbps	200/400 Gbps										
12.8 Tbps	100 Gbps	400 Gbps										

- Cloud is driving to 12.8 Tbps in a 1 RU box (32 ports of 400 Gbps)
- Cloud is looking past 400 Gbps today
  - Form Factors need to look beyond 400 Gbps now
- Cloud is looking for Ethernet Fabrics to replace Routing and Transport
  - Distance requirements for pluggables is increasing



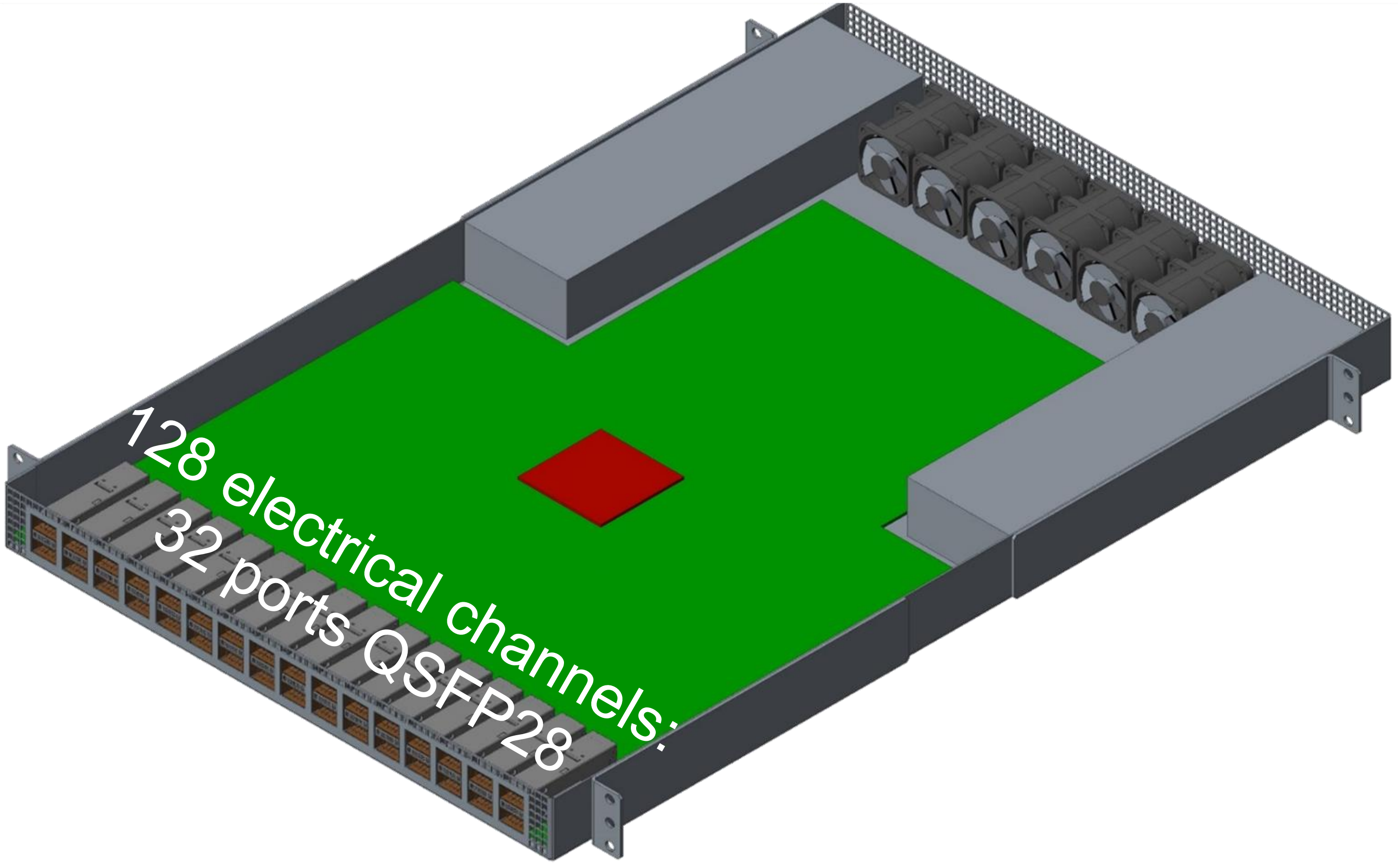
Data Courtesy of 650 GROUP MARKET INTELLIGENCE RESEARCH

Data Center Switches

- **Aggregate bandwidth:**
  - # of ports x bandwidth per port
- **Historically: 48 ports at 10 Gbps**
  - 480 Gbps per line card
  - 48 electrical channels at 10 Gbps
- **Today: 32 ports at 100 Gbps**
  - 3.2 Tbps per line card
  - 128 electrical channels at 25 Gbps
- **Next Generation: 32 ports at 400 Gbps**
  - 12.8 Tbps per line card
  - 512 electrical channels at 25 Gbps
  - 256 electrical channels at 50 Gbps

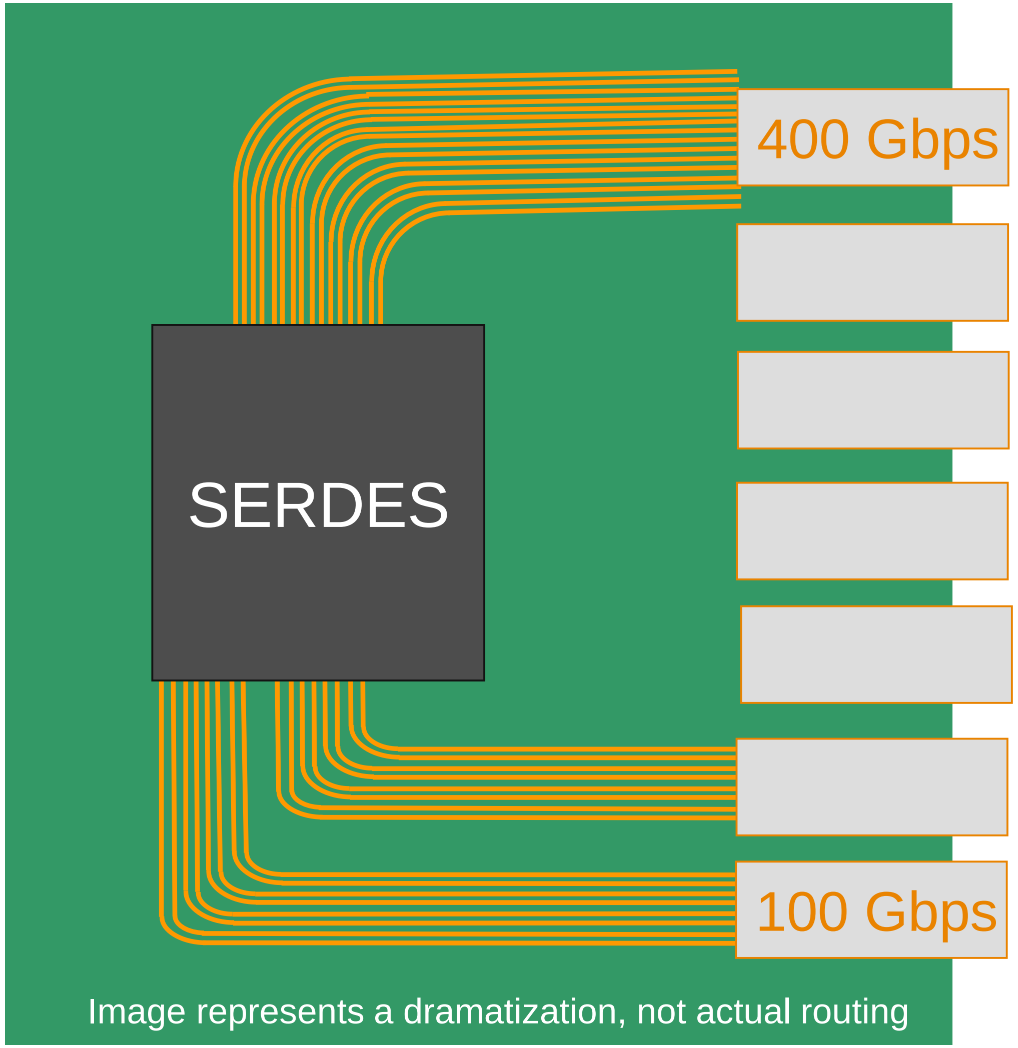
Next Generation Electrical Channels

- **512 channels at 25 Gbps is impractical**
  - Limited by SERDES package solder balls
  - Limited by PCB routing density
  - Limited by connector / module interconnect
- **256 channels at 50 Gbps is what we will focus on:**
  - 50 Gbps PAM4 signaling has recently been defined
- **256 channels represents a doubling of today's current practice of 128 electrical channels**



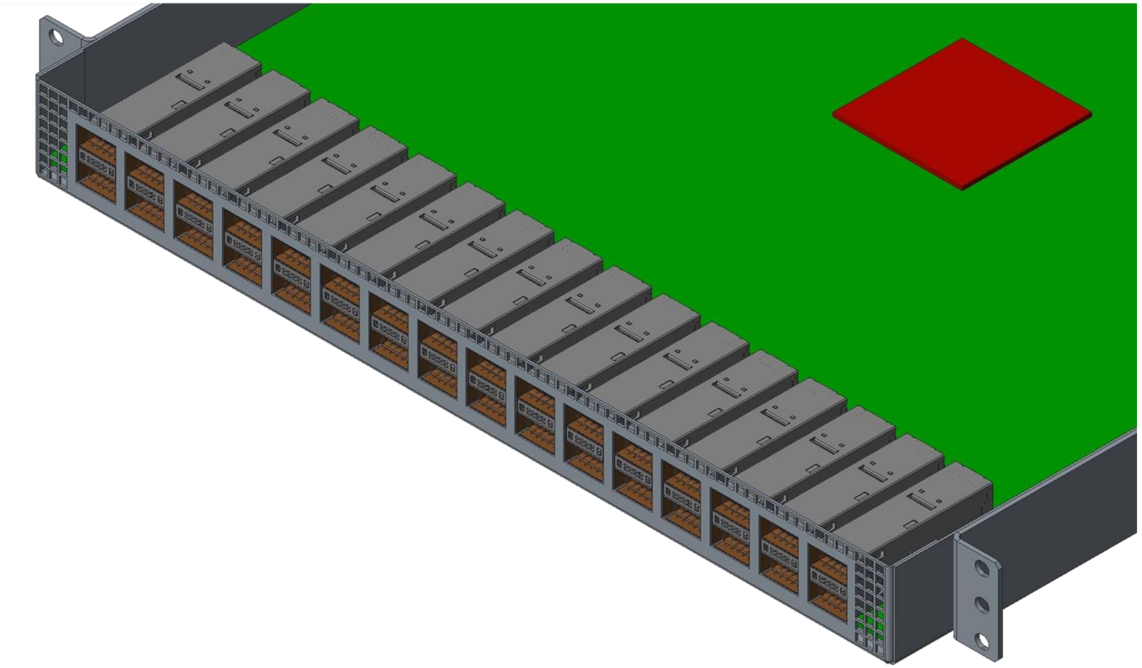
Electrical Channel Density Challenges

- Moving from 128 channels to 256 channels creates cross-talk concerns due to increased density
- Channel quality such as return loss, impedance, etc. due to routing implementations
- Reach or insertion loss is critical. For pluggable optic modules it is dominated by PCB and connector performance. In the case of direct attach copper cables, cable size (wire gauge) is a critical factor and this is determined by the module form factor cross sectional area
- Higher bandwidth-density creates thermal management challenges as next generation rates dissipate more power while density constraints are putting them closer and closer together



What's a Port? Key Equipment Considerations

- **I/O ports are valued for their flexibility**
- **Consist of connectors and cages that accept pluggable modules**
    - Passive direct attach copper cable
    - Short reach optical modules
    - Medium reach optical modules
    - Long reach optical modules
- **Allows end users to flexibly choose the appropriate reach and cost solution**
- **Provide good signal integrity**

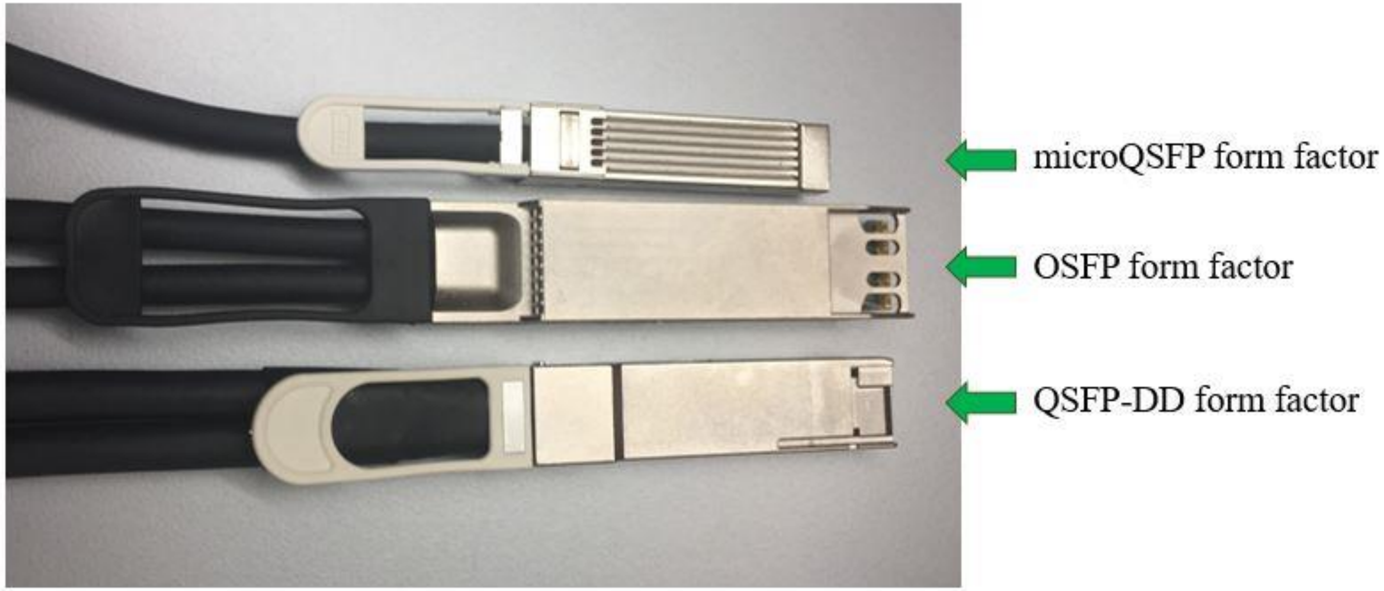


- **Optimize thermal dissipation from the optics**
- **Different channel counts**
- **Port selection determines aggregate bandwidth and granular bandwidth**

The Candidate Form Factors

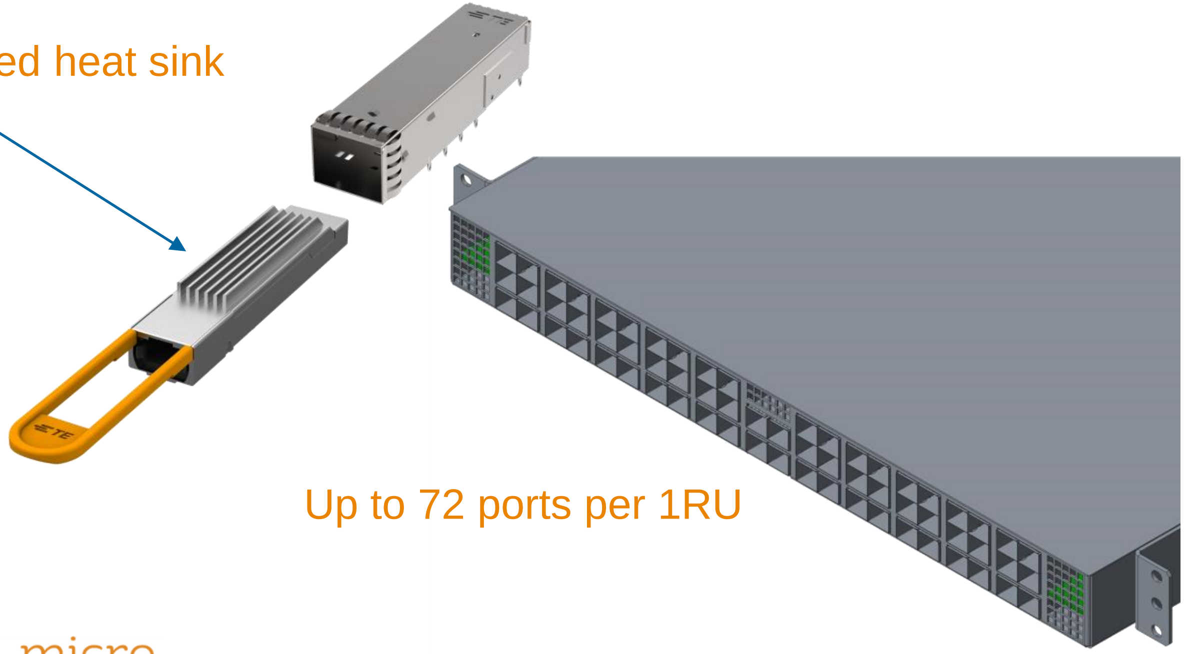
- microQSFP
- OSFP
- QSFP-DD

- All three solutions can accommodate more than 256 channels in 1RU (up to 288 channels)
- Different implementations bring different strengths and weaknesses
- TE is a founding member of all three MSAs and offering product to market, i.e. first hand experience/data



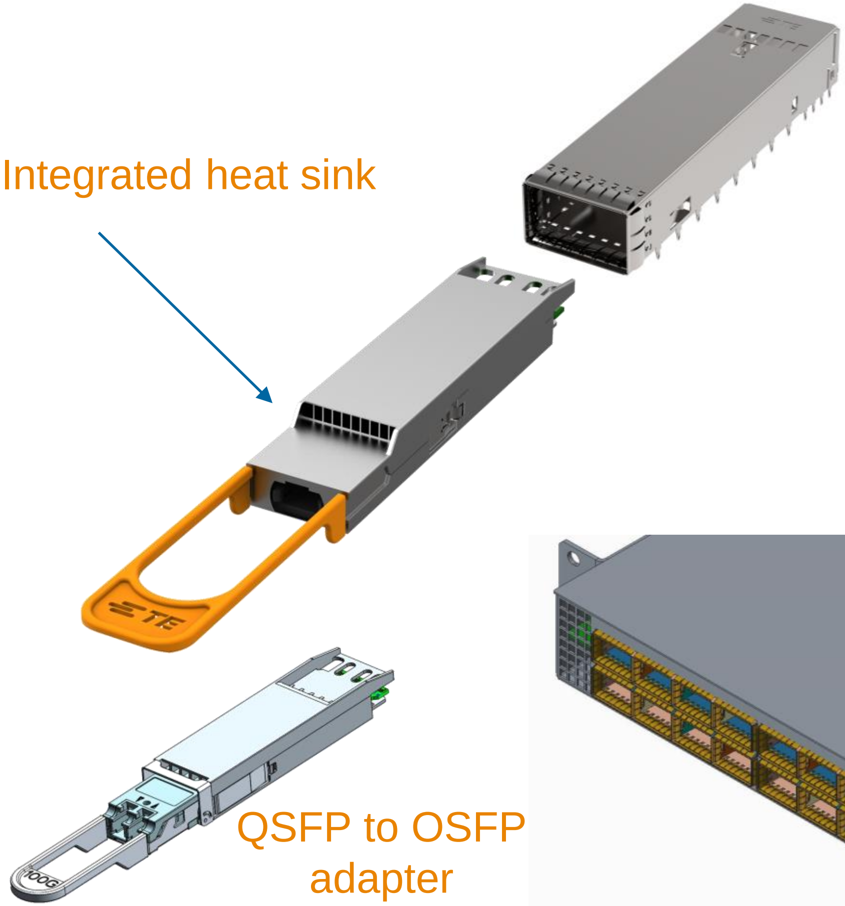
microQSFP Form Factor

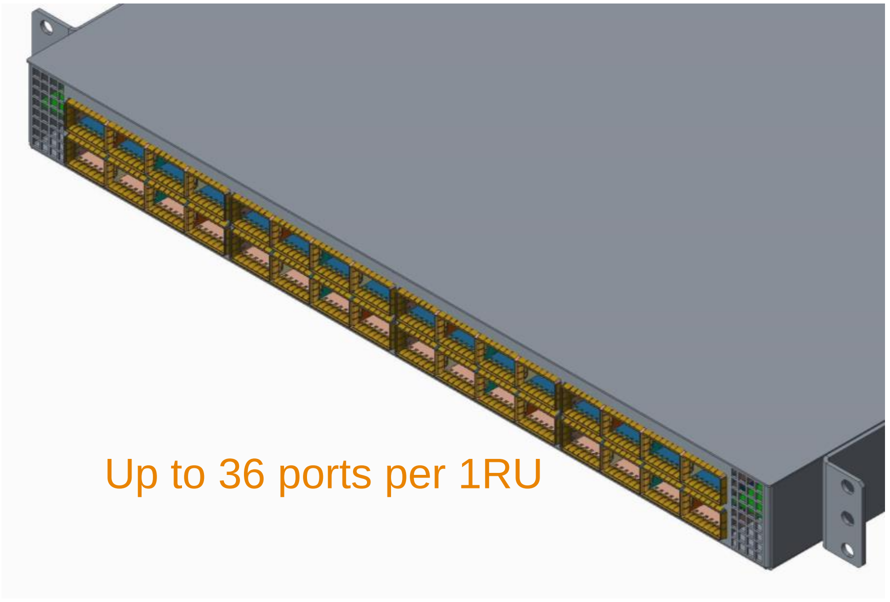
- A four channel port that fits 256 channels in 1RU with 64 microQSFP ports (up to 72 ports can fit but we will consider 64 ports since it equates to 256 channels)
- Able to support stacking of 3 ports to achieve density
- Achieves increase in density by going to 0.6mm contact pitch (vs. today's 0.8mm contact pitch)
- Uses a new module integrated thermal management solution to achieve higher power dissipation capability
- Can provide backward compatibility to SFP modules with the use of an adapter



OSFP Form Factor

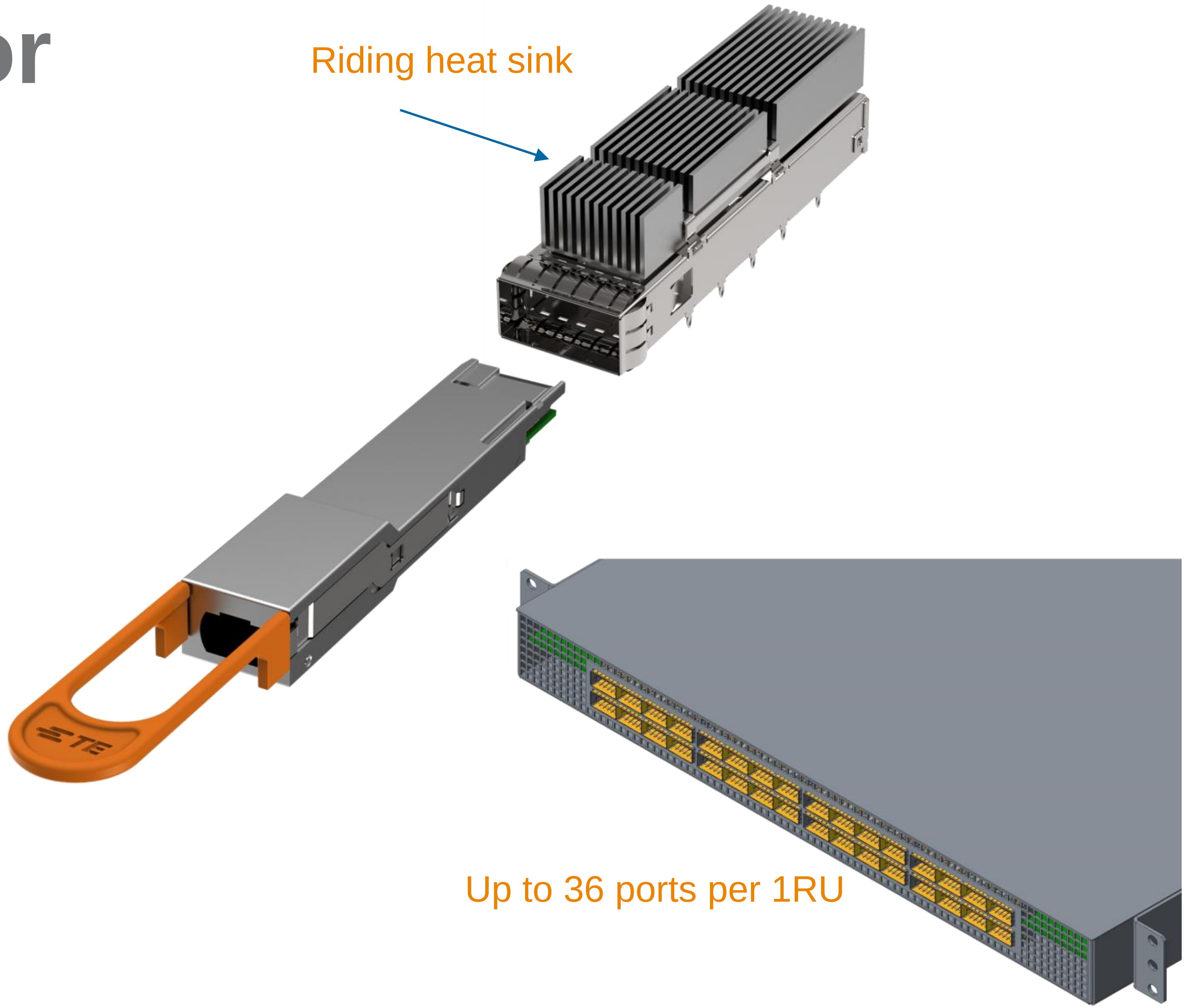
- OSFP is an eight channel port that accommodates 256 channels in 1RU via 32 modules (up to 36 modules can fit in in 1RU but we will focus on 32 modules since it equates to 256 channels)
- It achieves density by using a 0.6mm connector contact pitch (vs. today's 0.8mm contact pitch)
- Like microQSFP, it implements a module integrated heat sink to achieve higher levels of power dissipation
- Can provide backward compatibility to QSFP modules with the use of an adapter





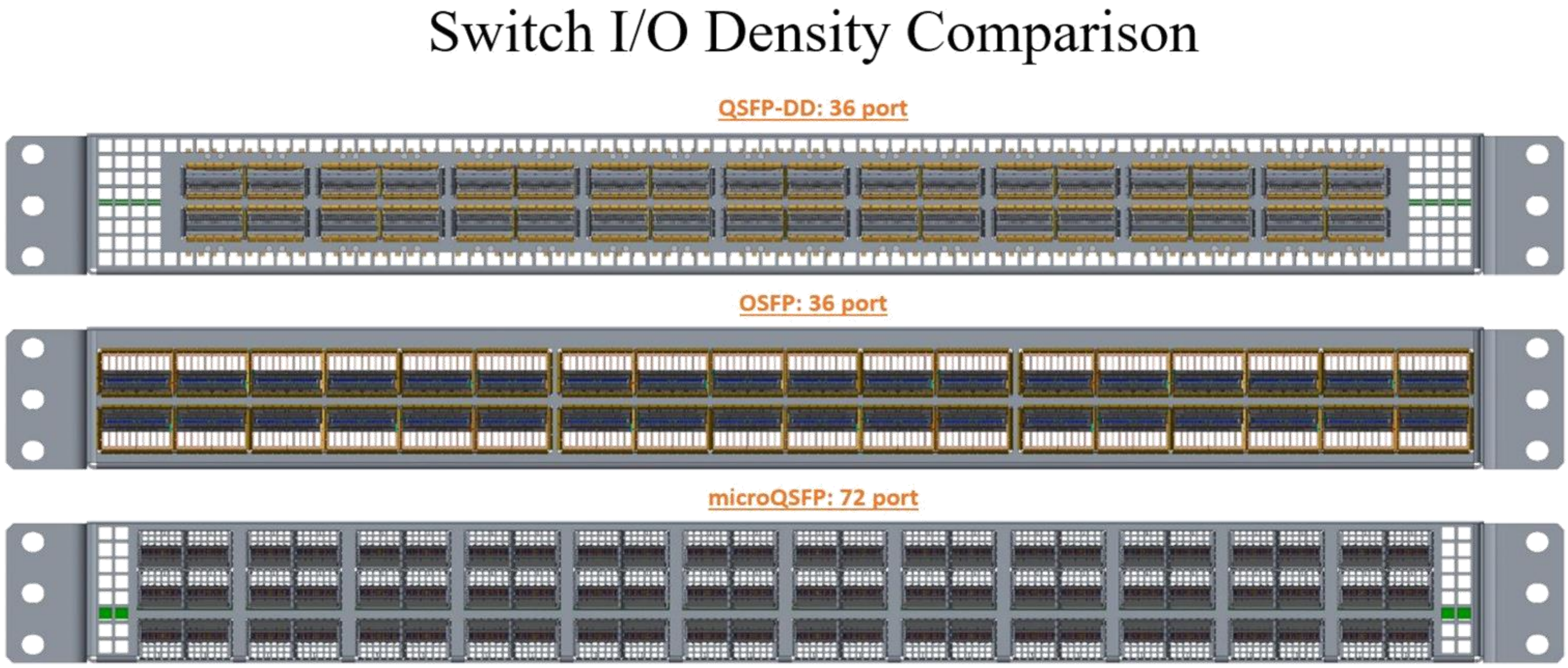
QSFP-DD Form Factor

- QSFP-DD is a new form factor port that enables backwards compatibility with existing QSFP modules
- Because of the backwards compatibility, it keeps the connector contacts on 0.8 mm pitch and adds additional rows of recessed contacts
- It uses the traditional riding heat sink thermal management methodology
- QSFP-DD allows an extra 15mm of module length outside the faceplate
- QSFP-DD can support 256 channels in 1RU with 32 modules in 1RU (36 modules can be supported but we will focus on 32 modules since this equates to 256 channels)



Switch Density Comparison

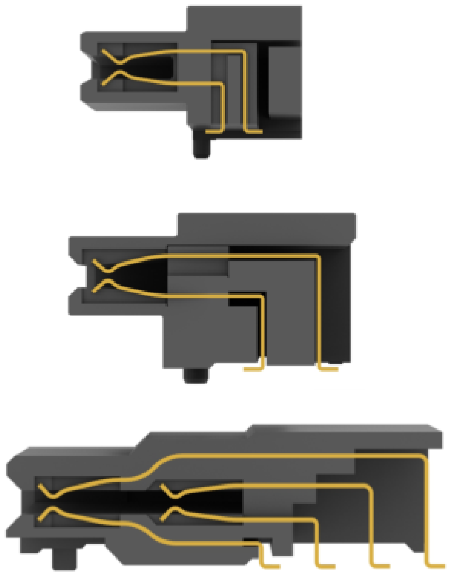
- All three form factors can more than meet the 256 electrical channel objective
- 288 electrical channels shown in the image



Differences in Connector Design to Achieve Density

- **microQSFP and OSFP achieve density by reducing connector contact pitch from 0.8 to 0.6mm**
- **QSFP-DD achieves density by adding additional recessed rows of contacts on 0.8mm pitch**
- **The additional rows of contacts on QSFP-DD have more impact on connector cross talk than the tighter pitch on microQSFP and OSFP**

Cross section views of connectors



Cage front views

microQSFP connector

OSFP connector

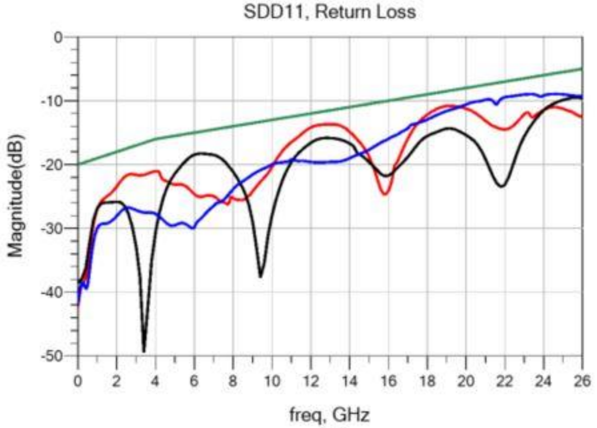
QSFP-DD connector

Front views not to scale with cross-section views

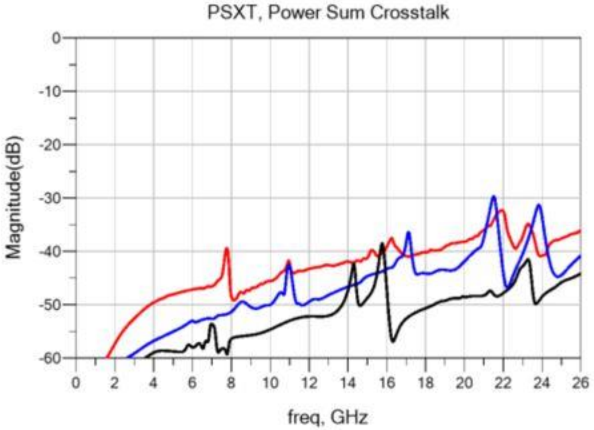
Signal Integrity - Simulation

Simulated Electrical Performance

(a)



(b)



(c)

	ICN FEXT (mV)	ICN NEXT (mV)	ICN Total (mV)
802.3bs Limits	4.2	1.5	4.4
OSFP	0.989	0.355	1.049
microQSFP	1.629	0.251	1.648
QSFP-DD	2.686	0.517	2.736

(d)

a.) Insertion Loss b.) Return Loss c.) PowerSum Crosstalk d.) ICN Table
OSFP (Black), microQSFP(Blue), QSFP-DD (Red)

Victim and Aggressors for simulated cross-talk

- Victim
- FEXT Aggressor
- NEXT Aggressor

60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
G	Tx1+	Tx1-	G	Tx3+	Tx3-	G	Tx5+	Tx5-	G	Tx7+	Tx7-	G	SB	SB	SB	SB	G	Rx8-	Rx8+	G	Rx6-	Rx6+	G	Rx4-	Rx4+	G	Rx2-	Rx2+	G
G	Tx2+	Tx2-	G	Tx4+	Tx4-	G	Tx6+	Tx6-	G	Tx8+	Tx8-	G	SB	SB	SB	SB	G	Rx7-	Rxy+	G	Rx5-	Rx5+	G	Rx3-	Rx3+	G	Rx1-	Rx1+	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

OSFP Pinout

38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
G	Tx1-	Tx1+	G	Tx3-	Tx3+	G	SB	SB	SB	SB	SB	G	Rx4+	Rx4-	G	Rx2+	Rx2-	G
G	Tx2-	Tx2+	G	Tx4-	Tx4+	G	SB	SB	SB	SB	SB	G	Rx3+	Rx3-	G	Rx1+	Rx1-	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

microQSFP Pinout

38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
G	Tx1-	Tx1+	G	Tx3-	Tx3+	G	SB	SB	SB	SB	SB	G	Rx4+	Rx4-	G	Rx2+	Rx2-	G
G	Tx5-	Tx5+	G	Tx7-	Tx7+	G	SB	SB	SB	SB	SB	G	Rx8+	Rx8-	G	Rx6+	Rx6-	G
76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58

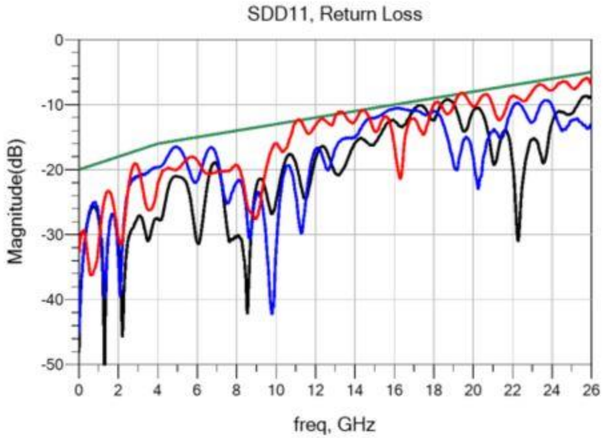
39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
G	Tx6-	Tx6+	G	Tx8-	Tx8+	G	SB	SB	SB	SB	SB	G	Rx7+	Rx7-	G	Rx5+	Rx5-	G
G	Tx2-	Tx2+	G	Tx4-	Tx4+	G	SB	SB	SB	SB	SB	G	Rx3+	Rx3-	G	Rx1+	Rx1-	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

QSFP-DD Pinout

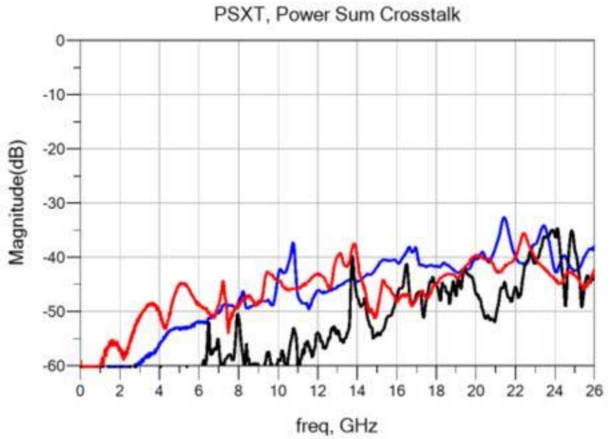
Signal Integrity- Measurement

Measured Electrical Performance

(a)



(b)



(c)

	ICN FEXT (mV)	ICN NEXT (mV)	ICN Total (mV)
802.3bs Limits	4.2	1.5	4.4
OSFP	0.574	0.696	0.902
microQSFP	2.144	0.612	2.229
QSFP-DD	2.416	xx	xx

(d)

a.) Insertion Loss b.) Return Loss c.) PowerSum Crosstalk d.) ICN Table
OSFP (Black), microQSFP(Blue), QSFP-DD (Red)

Victim and Aggressors for measured cross-talk

- Victim
- FEXT Aggressor
- NEXT Aggressor

60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
G	Tx1+	Tx1-	G	Tx3+	Tx3-	G	Tx5+	Tx5-	G	Tx7+	Tx7-	G	SB	SB	SB	SB	G	Rx8-	Rx8+	G	Rx6-	Rx6+	G	Rx4-	Rx4+	G	Rx2-	Rx2+	G
G	Tx2+	Tx2-	G	Tx4+	Tx4-	G	Tx6+	Tx6-	G	Tx8+	Tx8-	G	SB	SB	SB	SB	G	Rx7-	Rxy+	G	Rx5-	Rx5+	G	Rx3-	Rx3+	G	Rx1-	Rx1+	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

OSFP Pinout

38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
G	Tx1-	Tx1+	G	Tx3-	Tx3+	G	SB	SB	SB	SB	SB	G	Rx4+	Rx4-	G	Rx2+	Rx2-	G
G	Tx2-	Tx2+	G	Tx4-	Tx4+	G	SB	SB	SB	SB	SB	G	Rx3+	Rx3-	G	Rx1+	Rx1-	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

microQSFP Pinout

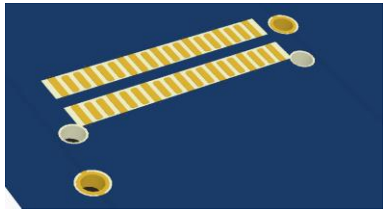
38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20
G	Tx1-	Tx1+	G	Tx3-	Tx3+	G	SB	SB	SB	SB	SB	G	Rx4+	Rx4-	G	Rx2+	Rx2-	G
G	Tx5-	Tx5+	G	Tx7-	Tx7+	G	SB	SB	SB	SB	SB	G	Rx8+	Rx8-	G	Rx6+	Rx6-	G
76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58

39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
G	Tx6-	Tx6+	G	Tx8-	Tx8+	G	SB	SB	SB	SB	SB	G	Rx7+	Rx7-	G	Rx5+	Rx5-	G
G	Tx2-	Tx2+	G	Tx4-	Tx4+	G	SB	SB	SB	SB	SB	G	Rx3+	Rx3-	G	Rx1+	Rx1-	G
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

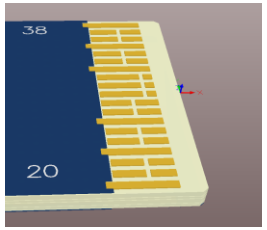
QSFP-DD Pinout

PCB Implications

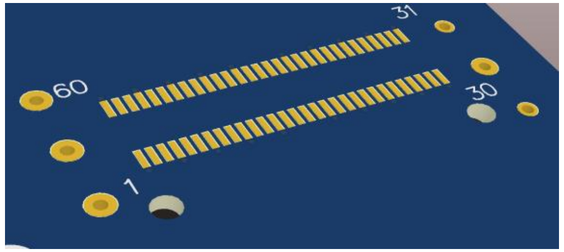
- The microQSFP and OSFP two-row connectors are easier to route both at the host board and at the module card edge PCB
- The QSFP-DD four-row connector adds complexity to both the host and the module PCB which impact cost and signal integrity
- The electrical effects of these routing differences are included in the measured data



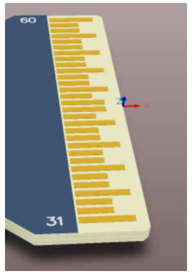
microQSFP host footprint



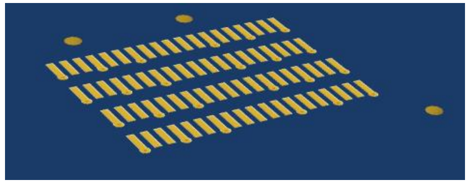
microQSFP card edge PCB



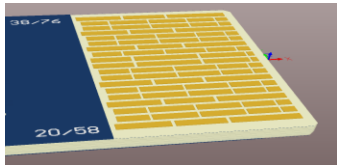
OSFP host footprint



OSFP card edge PCB



QSFP-DD host footprint



QSFP-DD card edge PCB

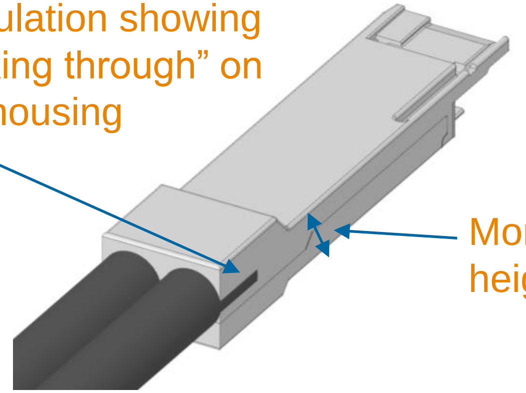
Direct Attach Cable Considerations

- **Industry standards typically specify minimum reach based on 26AWG cable**

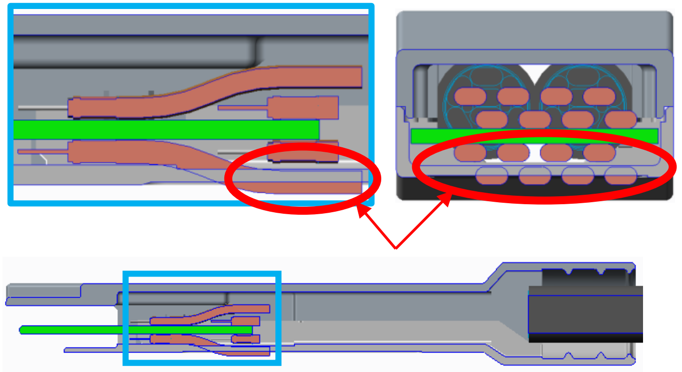
OSFP and microQSFP cable assemblies have been delivered with 26 AWG cable

QSFP-DD with 26 AWG cable has challenges with fitting into the exposed area of the backshell as well as the reduced height section of the module

26AWG simulation showing cable “breaking through” on the diecast housing



More challenging reduced height section

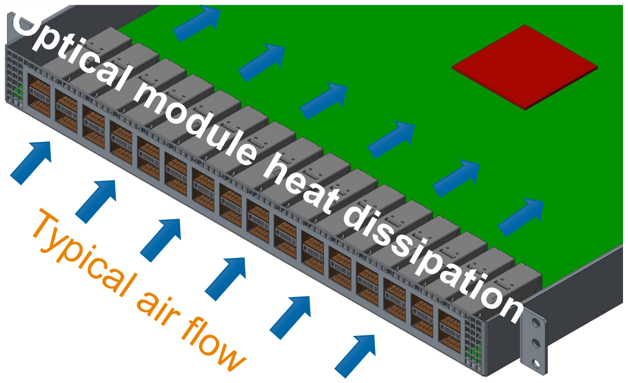


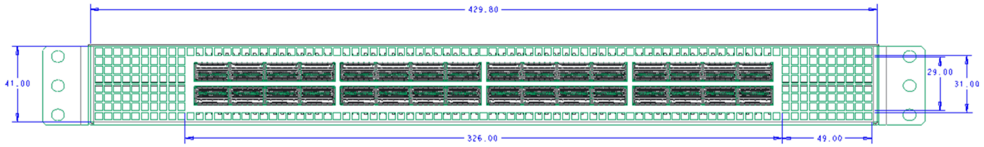
- **microQSFP and OSFP will always have a reach advantage due to internal packaging volume**

Thermal Management Factors

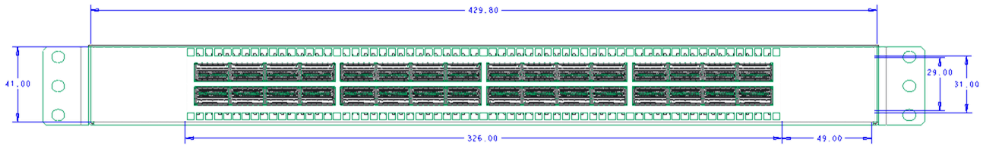
- **Pluggable I/O's concentrate the heat dissipation of the optical conversion at the faceplate of the equipment where the airflow for cooling the full equipment originates**
- **With 400 Gbps, optics modules are expected to be as high as 15W vs. 5W at 100 Gbps!**
- **Ports need the lowest possible thermal resistance with the best possible volume of air flow**
- **Significant air needs to be focused on the modules, otherwise the thermal management of the modules degrades**

QSFP example





Best airflow, worst module cooling



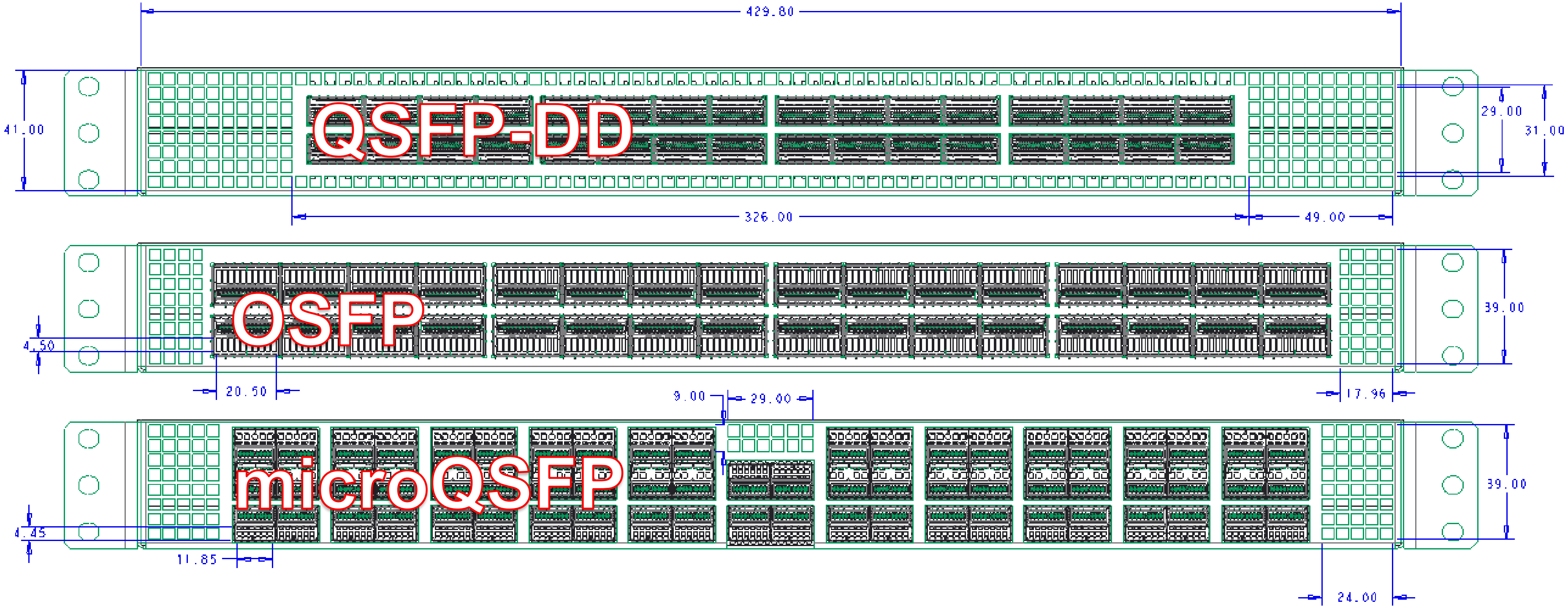
Best module cooling, restricted airflow

Airflow Trade-Offs

- Desire to maximize perforation area for equipment cooling
- Excess perforations “starve” the port cooling, resulting in high module temperatures

Airflow Perforation Comparison

Max air volume condition



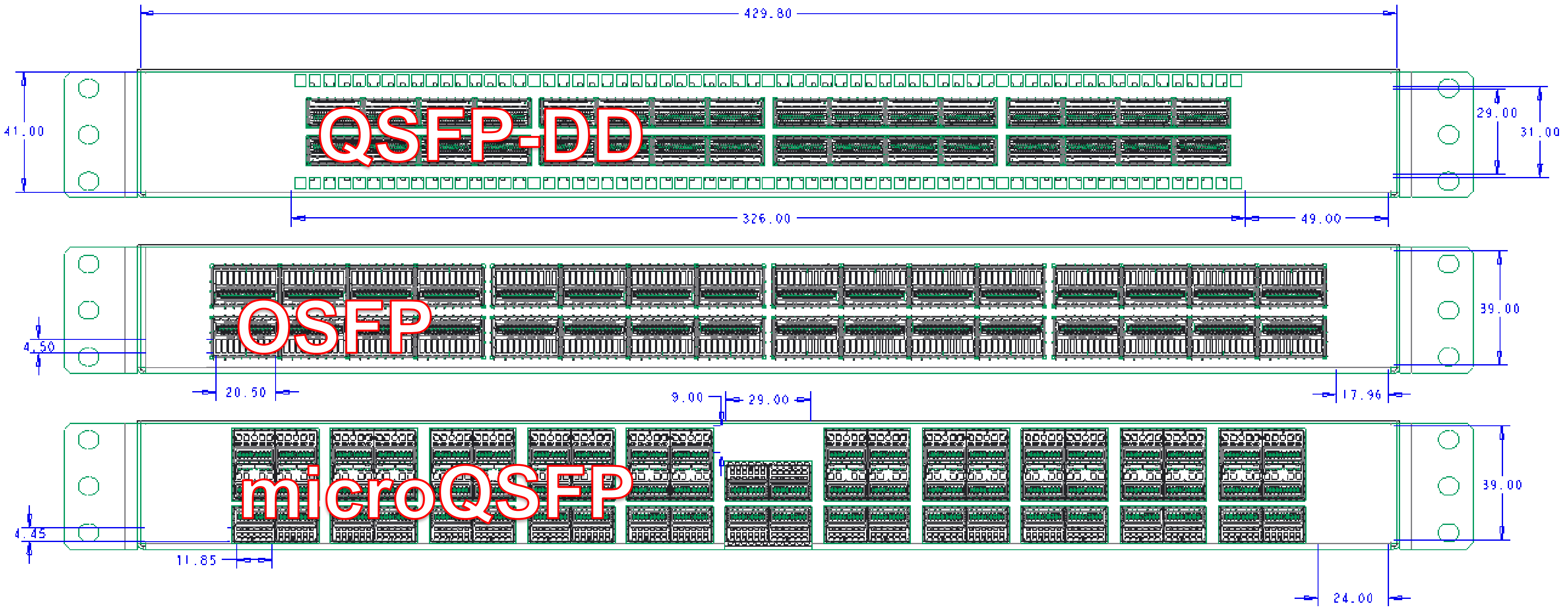
Switch I/O	I/O Port Qty	Available Faceplate Area	Perf Area in Faceplate	Perf Area in Cage	Total Perf Area	Percentage Perf
QSFP-DD	32	17,621.8	6,266.0	0.0	6,266.0	35.6%
OSFP	32		1,400.9	2,952.0	4,352.9	24.7%
microQSFP	64		2,133.0	3,374.9	5,507.9	31.3%

Airflow Trade-Offs, continued

- **Restricting airflow to cool the potentially 15W modules**
- **Ports that allow airflow have a significant benefit to also cooling the equipment**

Airflow Perforation Comparison

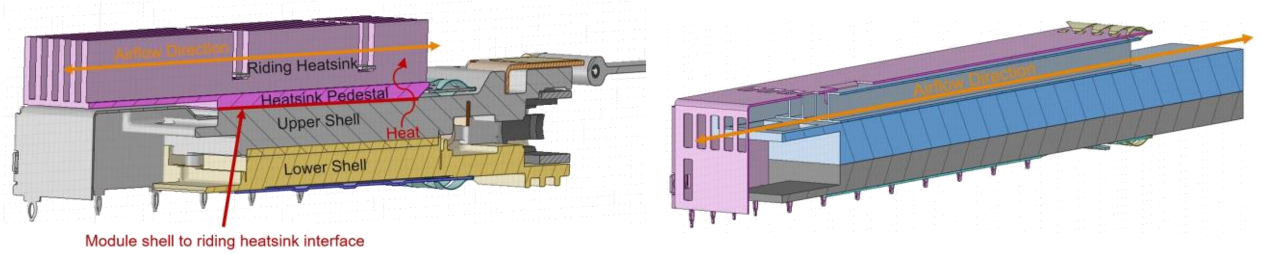
Optimized module cooling condition



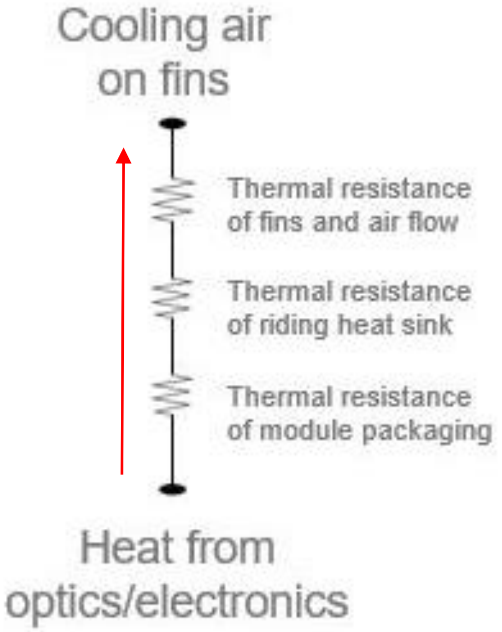
Switch I/O	I/O Port Qty	Available Faceplate Area	Perf Area in Faceplate	Perf Area in Cage	Total Perf Area	Percentage Perf
QSFP-DD	32		2,608.0	0.0	2,608.0	14.8%
OSFP	32	17,621.8	0.0	2,952.0	2,952.0	16.8%
microQSFP	64		0.0	3,374.9	3,374.9	19.2%

Thermal Mgmt – Airflow and Thermal Resistance

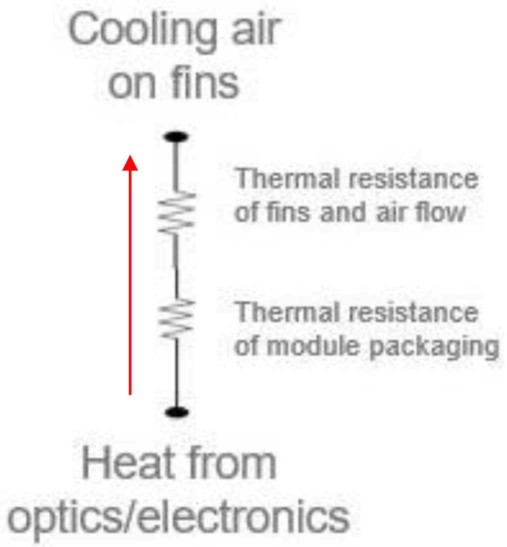
Cross section views:
Riding heat sink module vs. Integrated heat sink module

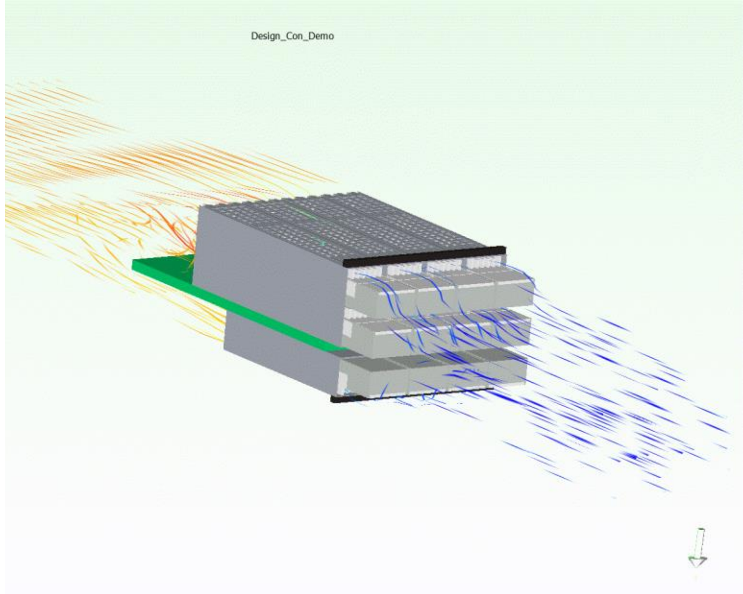


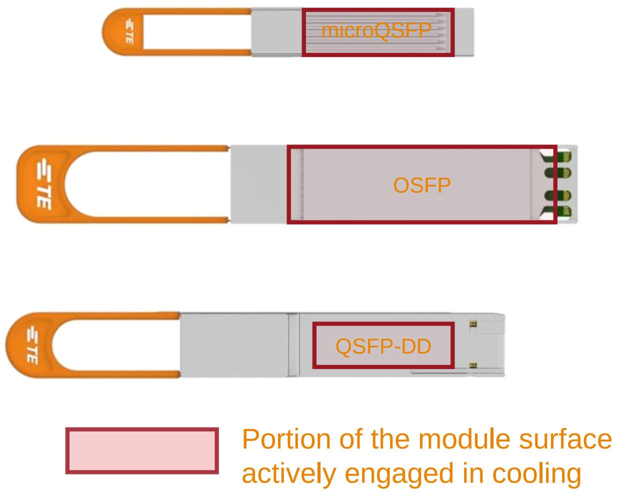
Riding Heat Sink



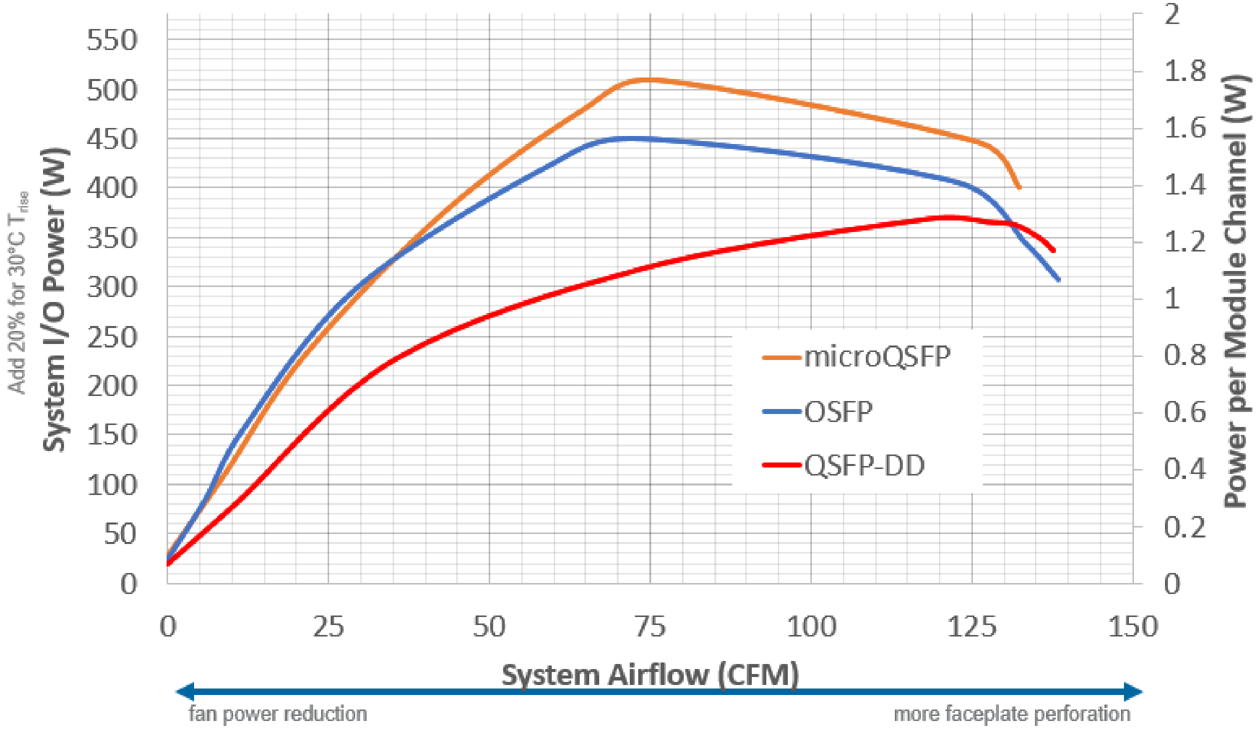
Integrated Heat Sink







Thermal Management – Comparative Simulation



QSFP-DD Belly/Belly
36 ports

OSFP Belly/Belly
36 ports

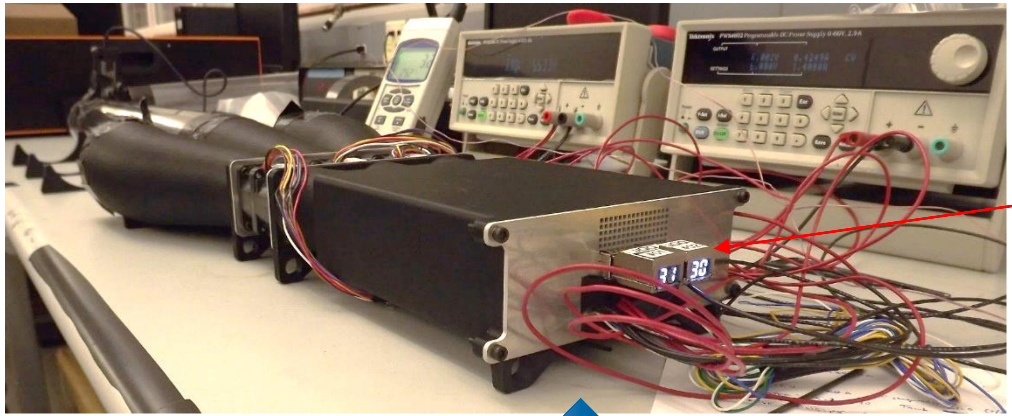
microQSFP 3-High
72 ports

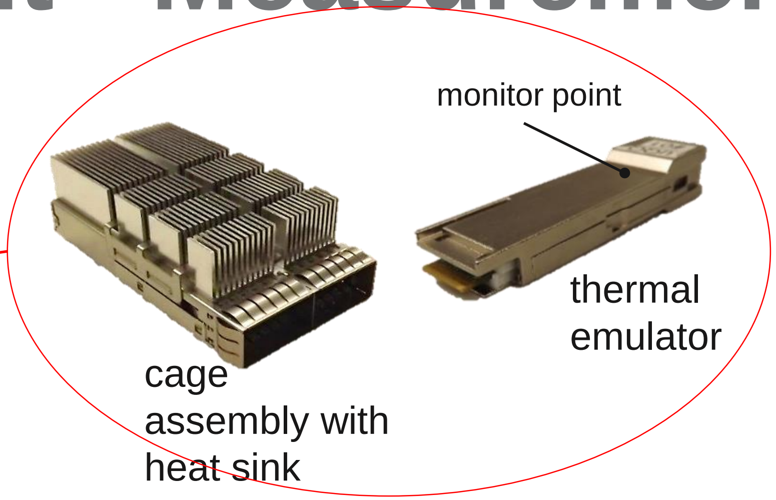
Comparative side by side by side simulations:

- Same 1RU enclosure
- Same fans
- Face plate perforations are optimized for each form factor
- Monitoring module hot spot at 70ºC over range of module powers and airflows

Results for total equipment IO power and per electrical channel power

Thermal Management - Measurements





- Per port airflow control, 2-15 CFM (64-480 CFM for 32 ports)
- Cage & heat sink characterization platform
- Module power settings from 1-15W
- Multiple temperature monitor points
- Thermal test modules
- Airflow bypass control

Mini thermal airflow test beds

Full 1RU test enclosures

Results:

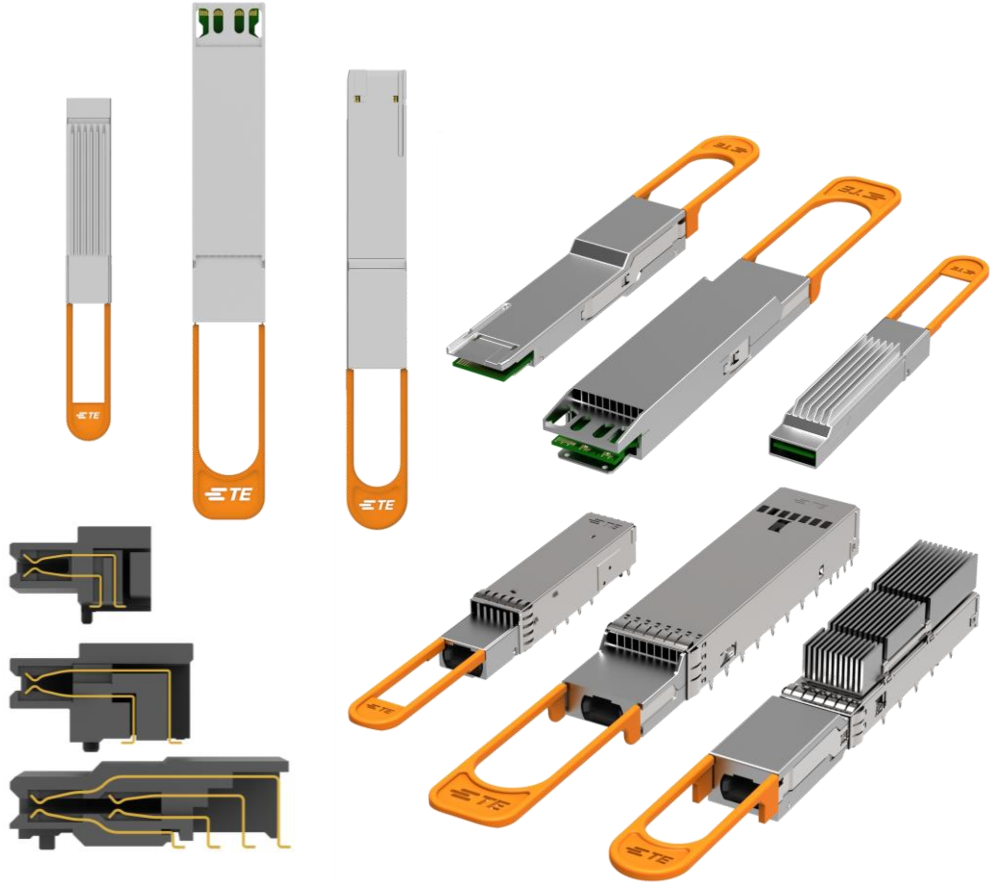
microQSFP: 1.9W per channel (7.5W for 4 channel module)

OSFP: 1.9W per channel (15W for 8 channel module)

QSFP-DD: 1.5W per channel (12W for 8 channel module)

Summary

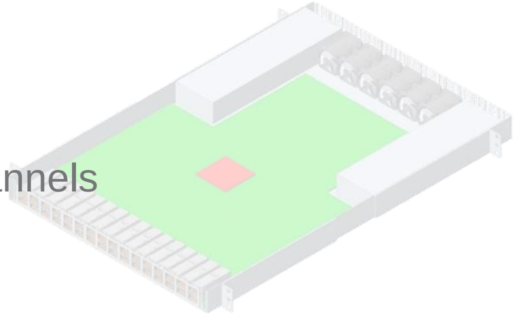
	Signal Integrity	Thermal mgmt	Larger Wire AWG	Channel Density	Backwards Compatibility
microQSFP					
Result	Modeled ICN of 1.6mV	1.9W per channel, 7.5W per module	26AWG fits	288 channels	SFP with adapter
OSFP					
Result	Modeled ICN of 1.0mV	1.9W per channel, 15W per module	26AWG fits	288 channels	QSFP with adapter
QSFP-DD					
Result	Modeled ICN of 2.7mV	1.5W per channel, 12W per module	26AWG is difficult	288 channels	Directly accepts legacy QSFP



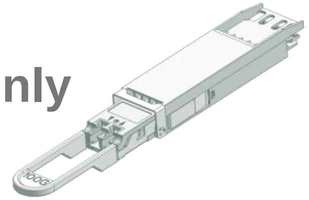
Conclusions

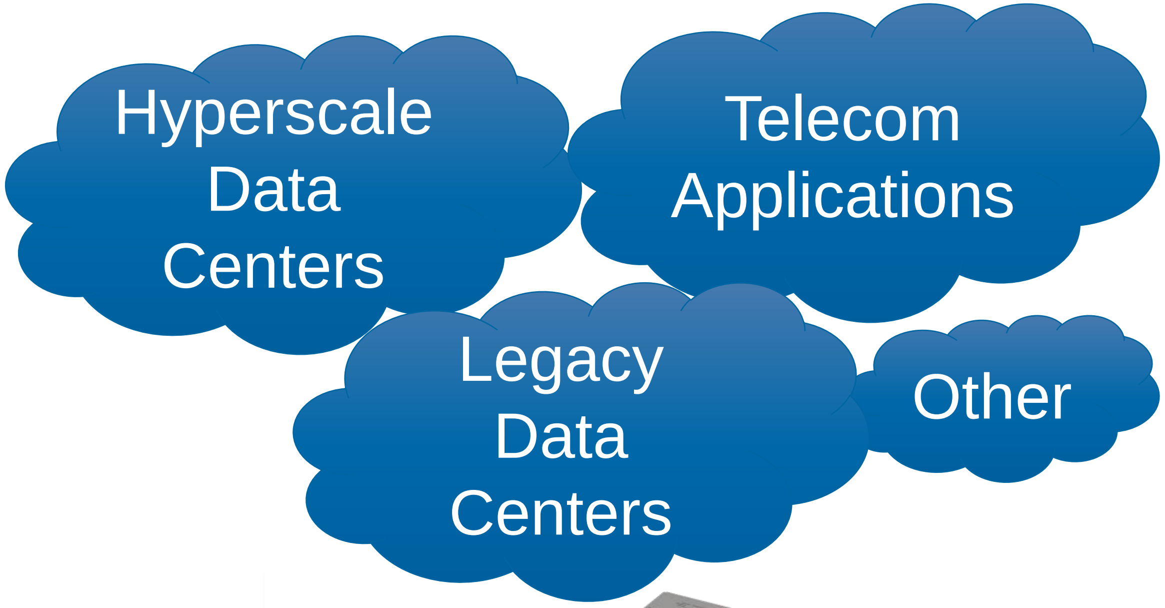
- **All three candidates solutions have been shown to be capable of enabling the new 400 Gbps generation of I/O, but with trade-offs**

- **Backwards compatibility is an important consideration for equipment, but at what cost (margin)?**
  - Thermal limitations
  - Use of retimers to extend channels
  - Higher performing fans
  - Etc.



- **Adapters (to enable backwards compatibility) are an extra part, but only burden the port for legacy cases, preserve margin for new cases**





- **What's your use case?**
- **What's the equipment lifecycle?**
- **What equipment performance attributes are most important to your customer?**

Thank you!

QUESTIONS?