Beyond 100 Gb/s High-speed Optical (and Copper) Interconnects

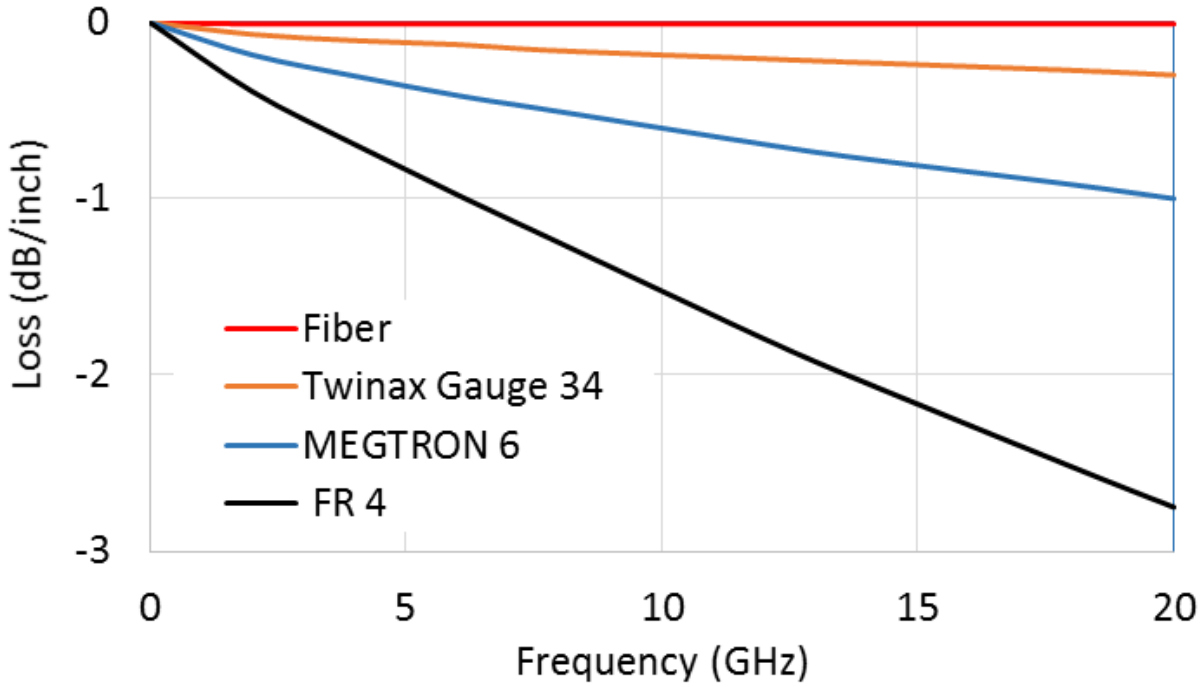
Marc Verdiell, CTO, Samtec Optical Group

TWEPP-17 - Topical Workshop on Electronics for Particle Physics, Santa Cruz, CA, Sept. 11-14th 2017

Interconnect challenges for data acquisition

- Raw speed
  - Modern data acquisition techniques and FPGAs can generate and process many Tb/s of data
  - Today at 16G and 28G per lane, soon going to 56G and 100G.
- Distance
  - Data needs to be transported out of detector, between FPGAs, or to compute resources over significant distances
  - Depending on rate, optics might already be needed beyond a few 10's of cm.
- Size
  - Space is limited on detectors and FPGA cards
  - Copper cabling, even if short, might be too bulky
- Cooling and thermals
  - Many designs are limited by power dissipation, low power is key
  - Active interconnects need to support many cooling methods: air flow, cold plate, immersion
- Various environmental constraints
  - Extended temperature range, shock and vibe, radiation hardness

High-speed Interconnect Challenges

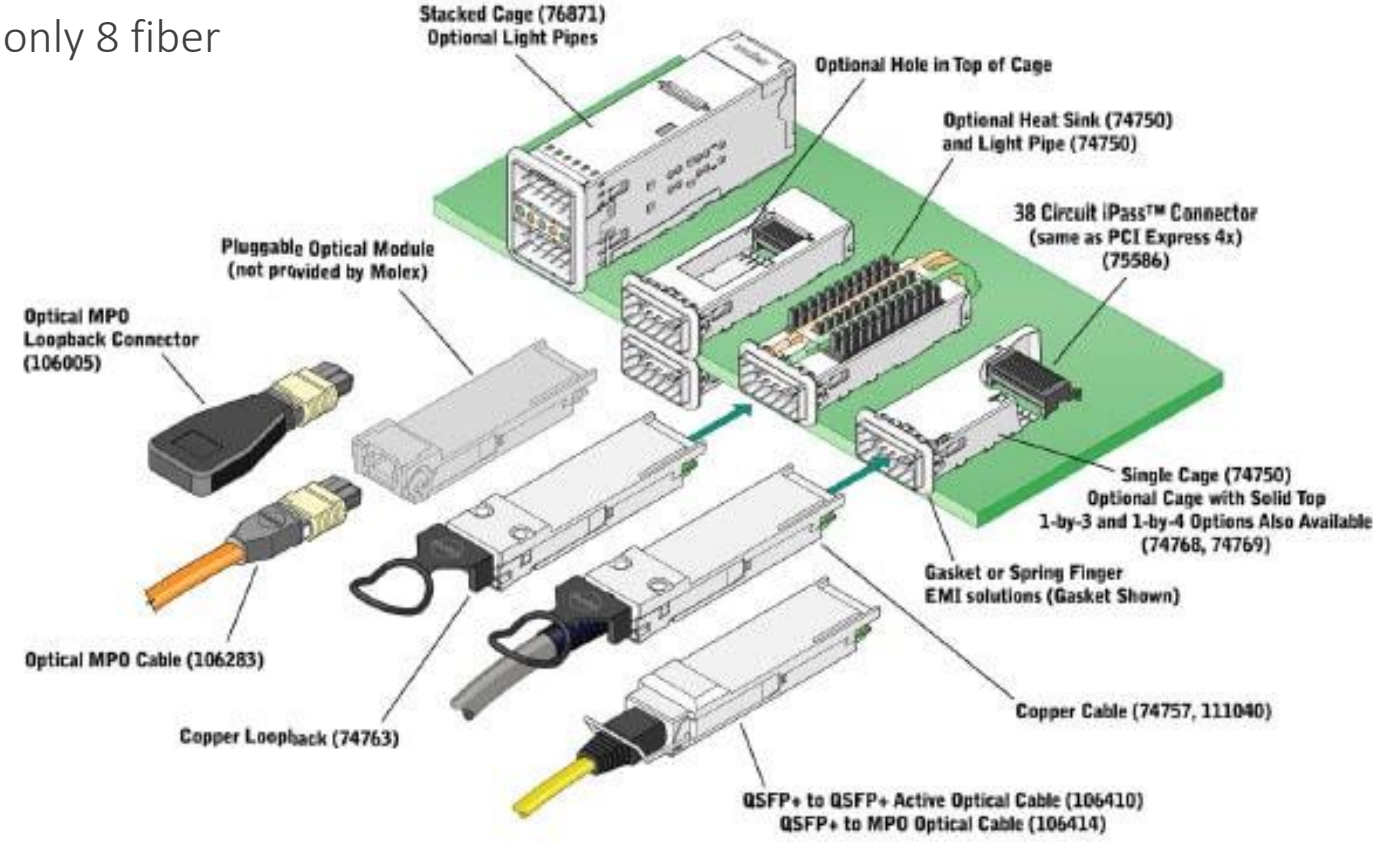


Bandwidth\ Materials\Reach	FR408	MEGTRON 6	Micro-Twinax	Optics
10G	<10”	10”+	10”+	10”+
14G	<5”	<10”	10”+	10”+
28G	<2”	<5”	10”+	10”+
56G	0.0”	<2”	<10”	10”+
112G	0.0”	0.0”	<5”	10”+

Traditional optical interconnect solutions: QSFP

QSFP = Quad Small Form-factor Pluggable

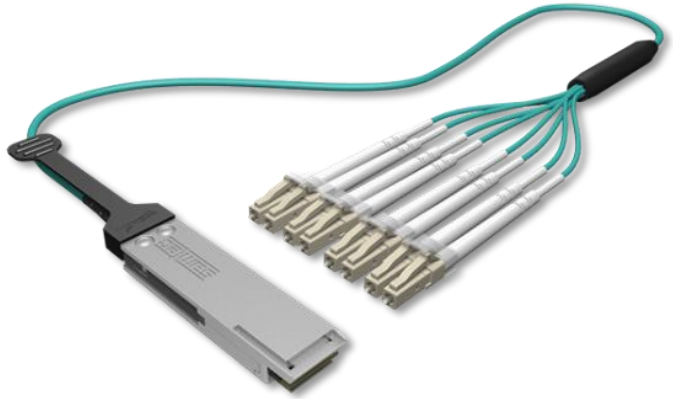
- It's not small at all!
- 4x bidirectional connections at 4x 10, 14 ("QSFP+"), 25 and 28G (QSFP28)
- Uses standard 12 fiber ribbon jumper cables (only 8 fiber used, 4 dark)
- Pluggable electrically and optically
- Commodity products in datacom

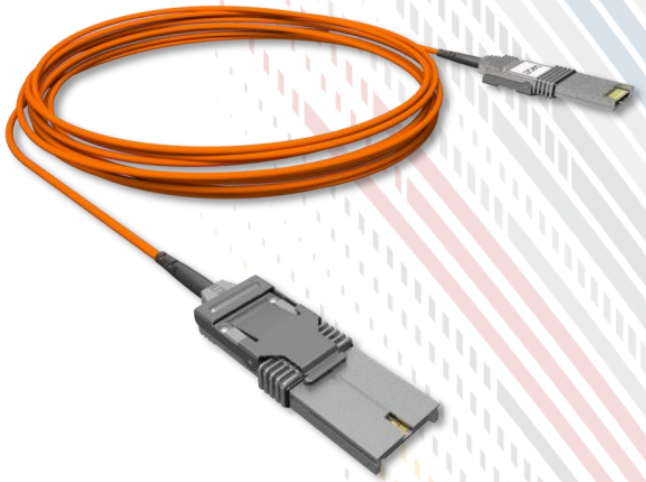


Traditional Optical Interconnect Solutions: AOC's

Active Optical Cables

- Great for point to point connections, when length is known in advance
- Handled (almost) like a copper cable, no optical connector or cleanliness related problems
- Available in "Octopus" breakouts
- Available in PCIe versions at Gen2 and Gen3: quite popular in data acquisition.





LIGO Gravitational Wave Observatory

Uses PCIe over optics to reduce noise interference.

200G and 400G: QSFP-DD and OSFP

Both are 8 lanes “double QSFP” devices, 8x 25G or 8x 56G, 16 fiber ribbon.

OFSP is larger width and length than QSFP28.

- Helps with power dissipation and manufacturability

QSFP-DD is same width but longer than QSFP

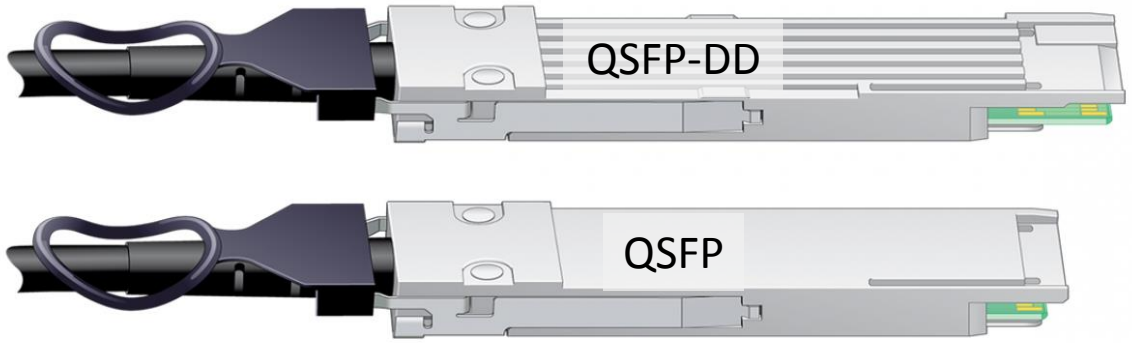
- Clever dual connector accepts both legacy standard and double density DD-QSFP
- Harder SI for the connector

Both start to be very challenging thermally

- 6W at 200G, 10-12W at 400G!

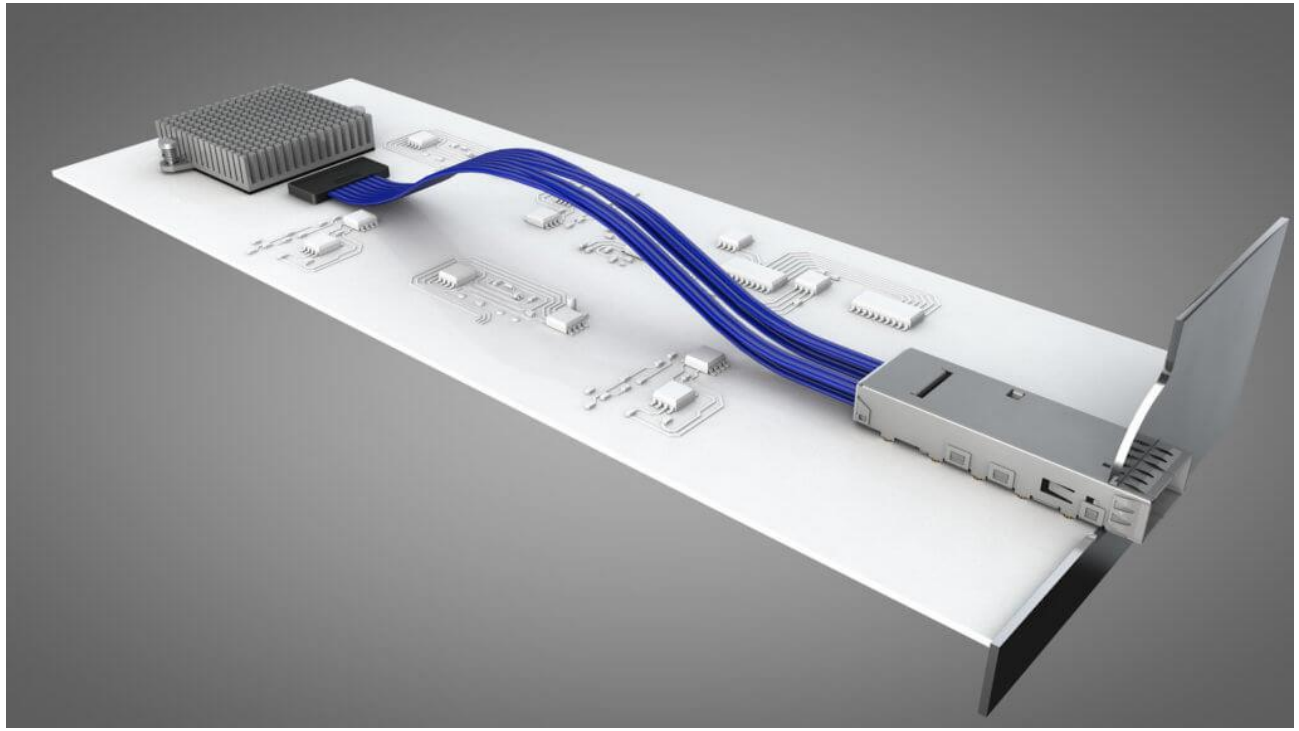
Becomes challenging to reach the module electrically

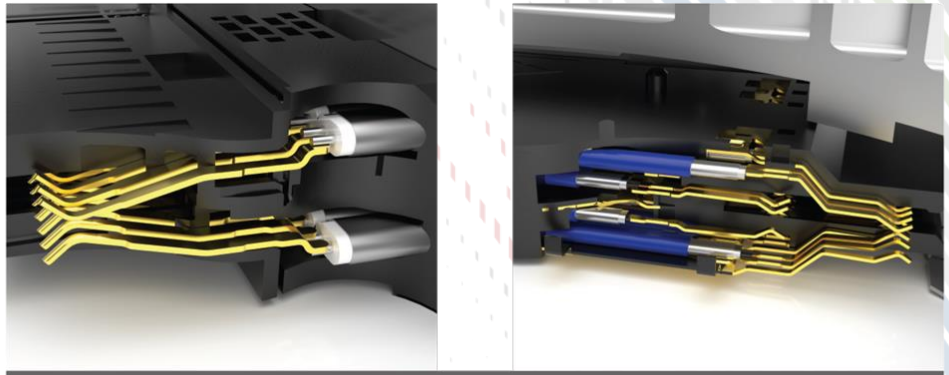
- input rate is 8x 56G (PAM4 signaling), module on faceplate



Flyover Technology

Low skew, low loss micro-twinax “flies over” the PCB, and connects directly to the back of the QSFP cage





Direct Solder Attach to Connector Contacts

On-board Miniature Optics

Why Miniature On-board Optics?

Bring optics on board, closer to the chipset

- No more electrical reach issues

Optical “flies over” to backplane or front panel

- Much denser than copper flyover

Highest density

- Miniature x12 channel engines in a 11.5x16mm form factor
- Can make arrange in multi-row arrays

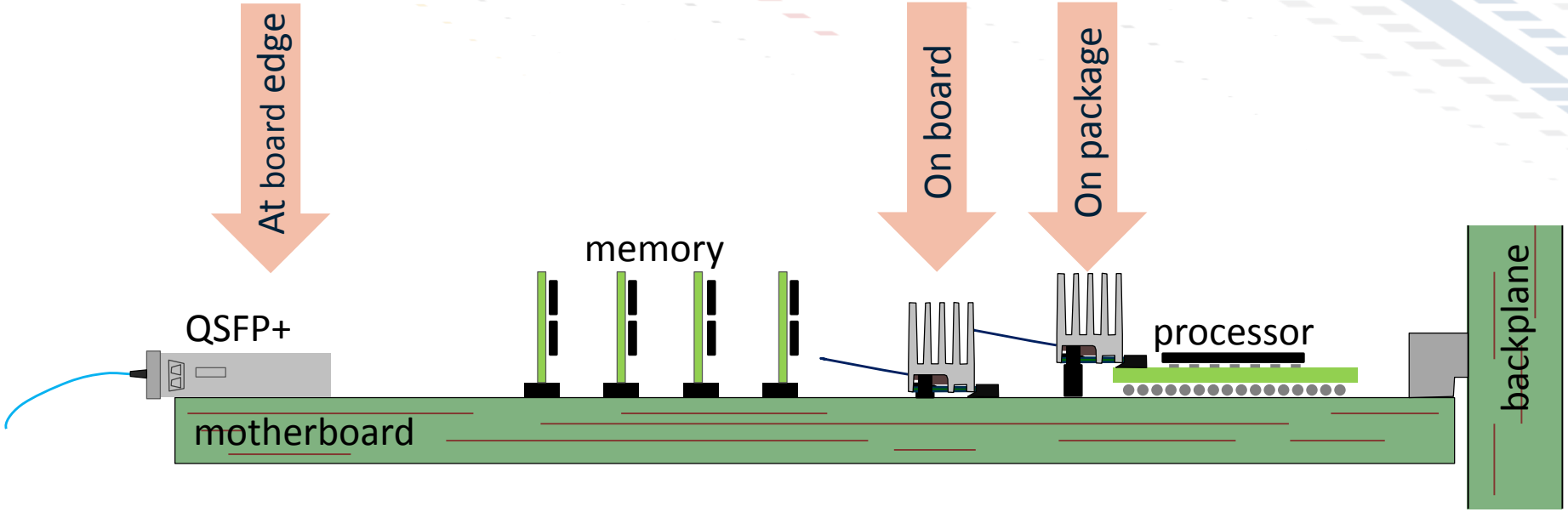
Many cooling options

- Traditional pin-fin heatsink
- Cold plate
- Immersion

Qualified to Telcordia, sometimes MIL standards

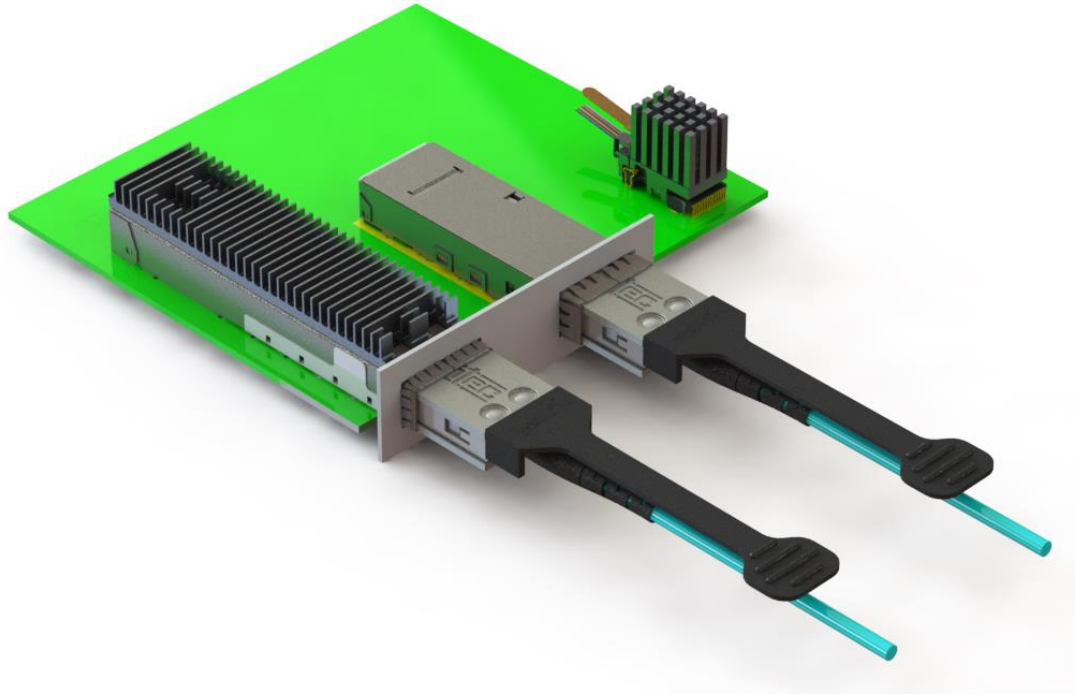
- Lots of reliability data

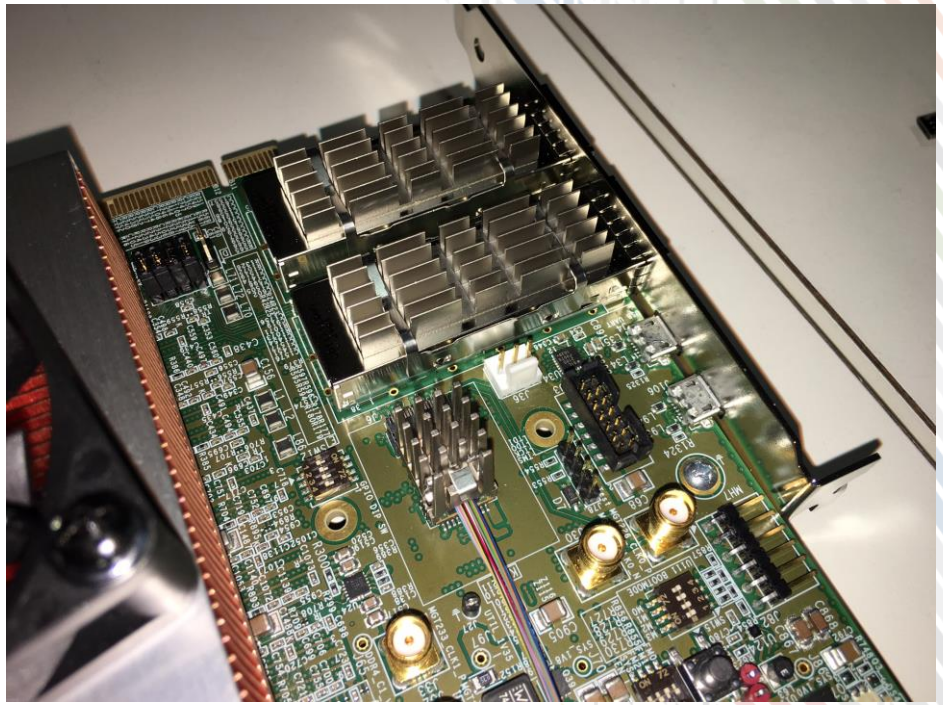
Onboard Optics Advantages



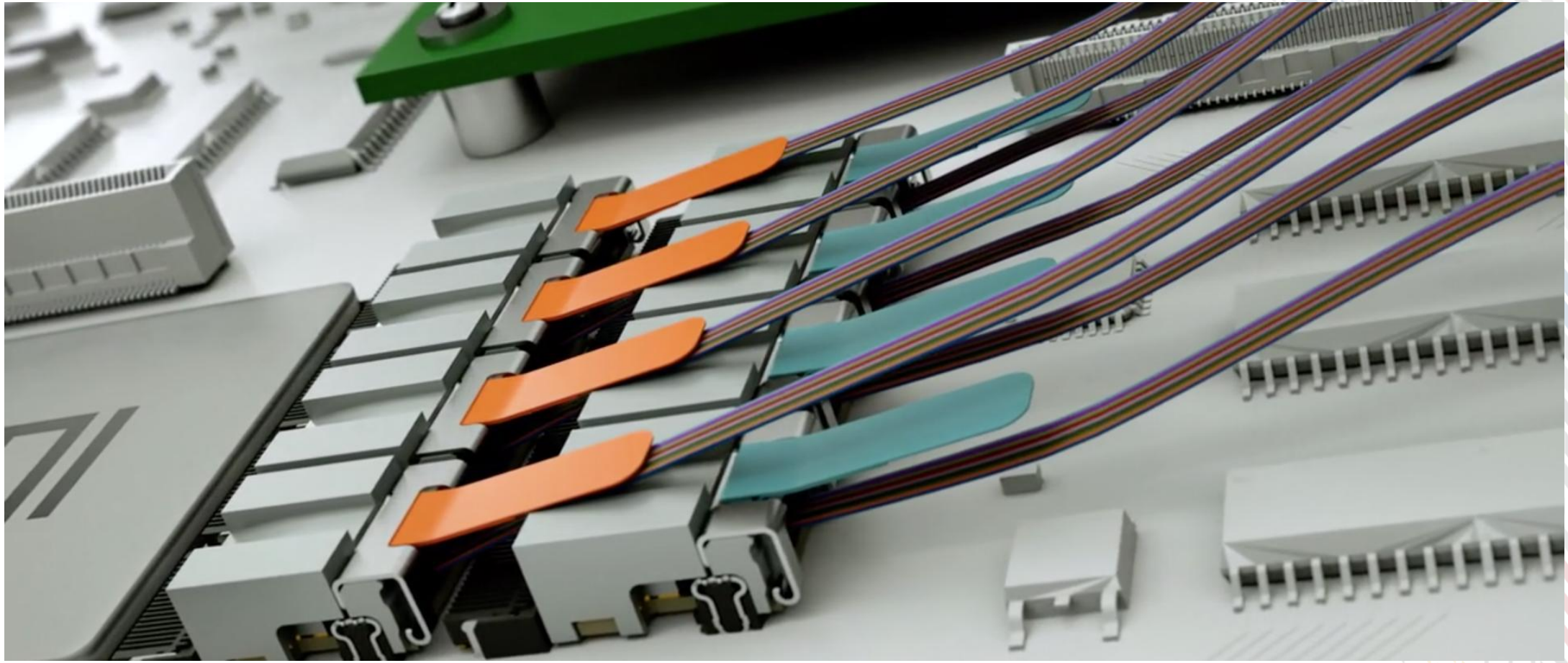
Easier Layout, Higher Density Better Signal Integrity, Lower Power

Size comparison with QSFP and QSFP-DD





On-board optics in array configuration



Higher Patch Panel / Face Plate Density

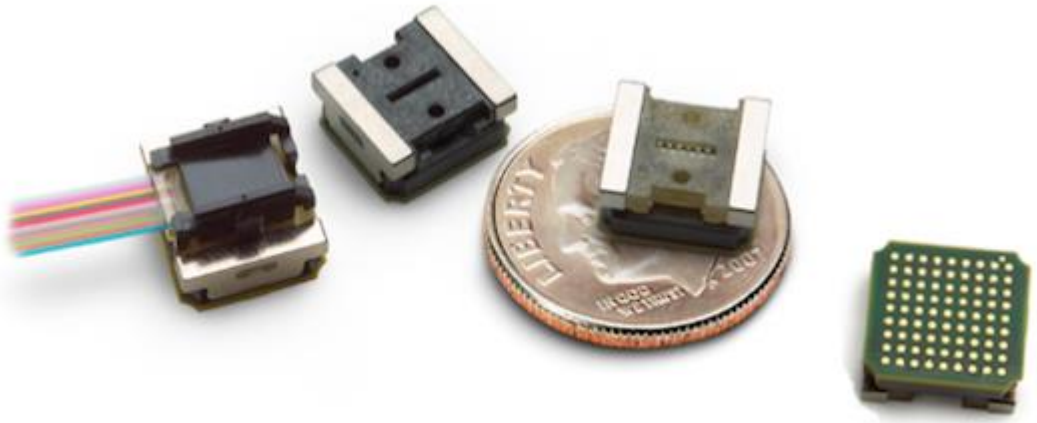
32 QSFP+

16 MT ferrules

Onboard Optics: Avago MicroPOD

Pioneering miniature onboard engine

- Very small optical module, 12x 10Gb/s
- Used in an fully optically interconnected implementation of IBM's Blue Gene
- Size advantage somewhat negated by compression hardware and pressure plate
- LGA connector makes it difficult to implement and install



FireFly Onboard Optics

Connector system specifically designed for high-density On-board optics

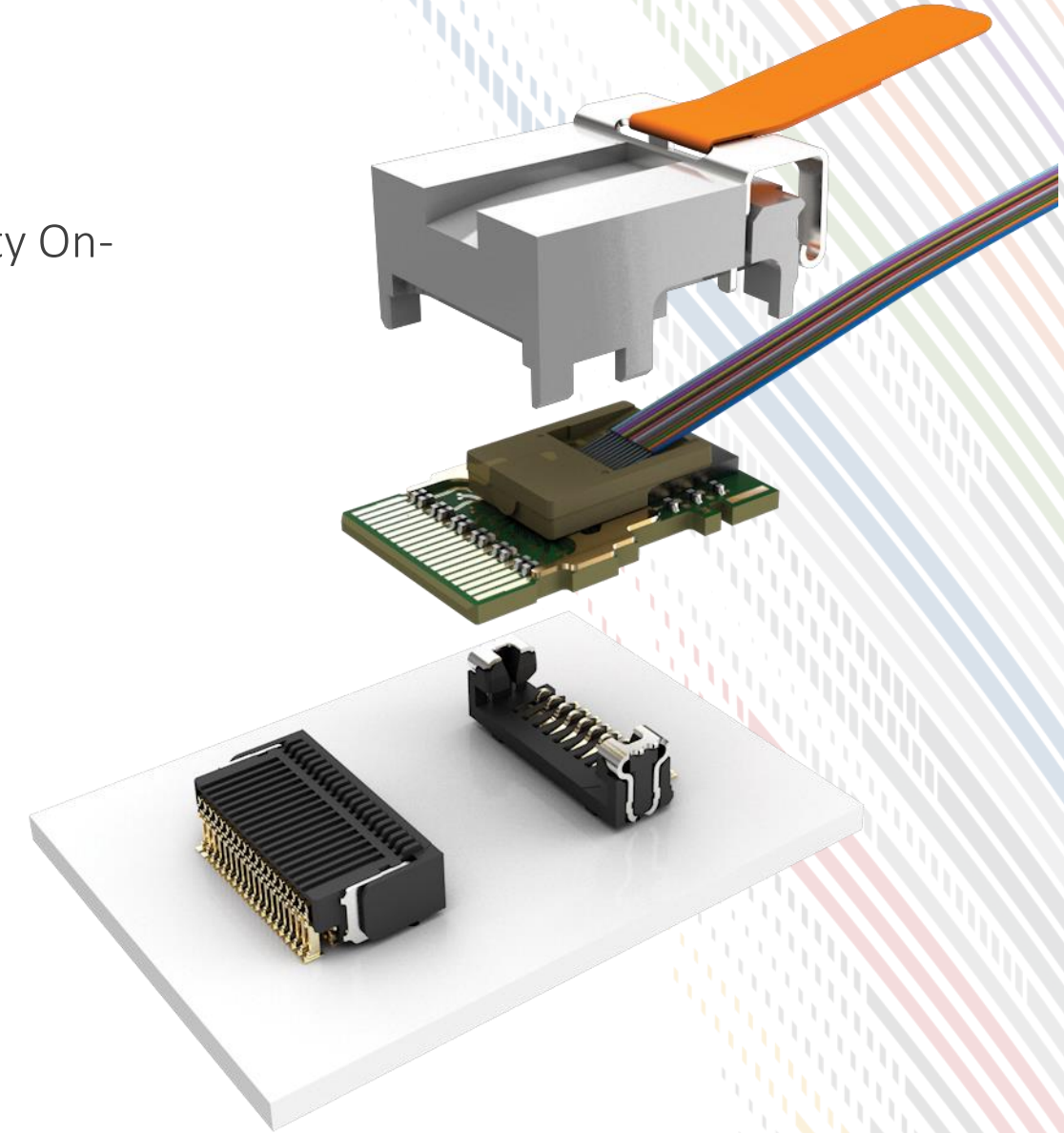
Back connector with high-speed lanes

- 12 differential pairs (GSSG)
- Designed for 56 Gbps
- Only 11.5 mm wide

Front connector

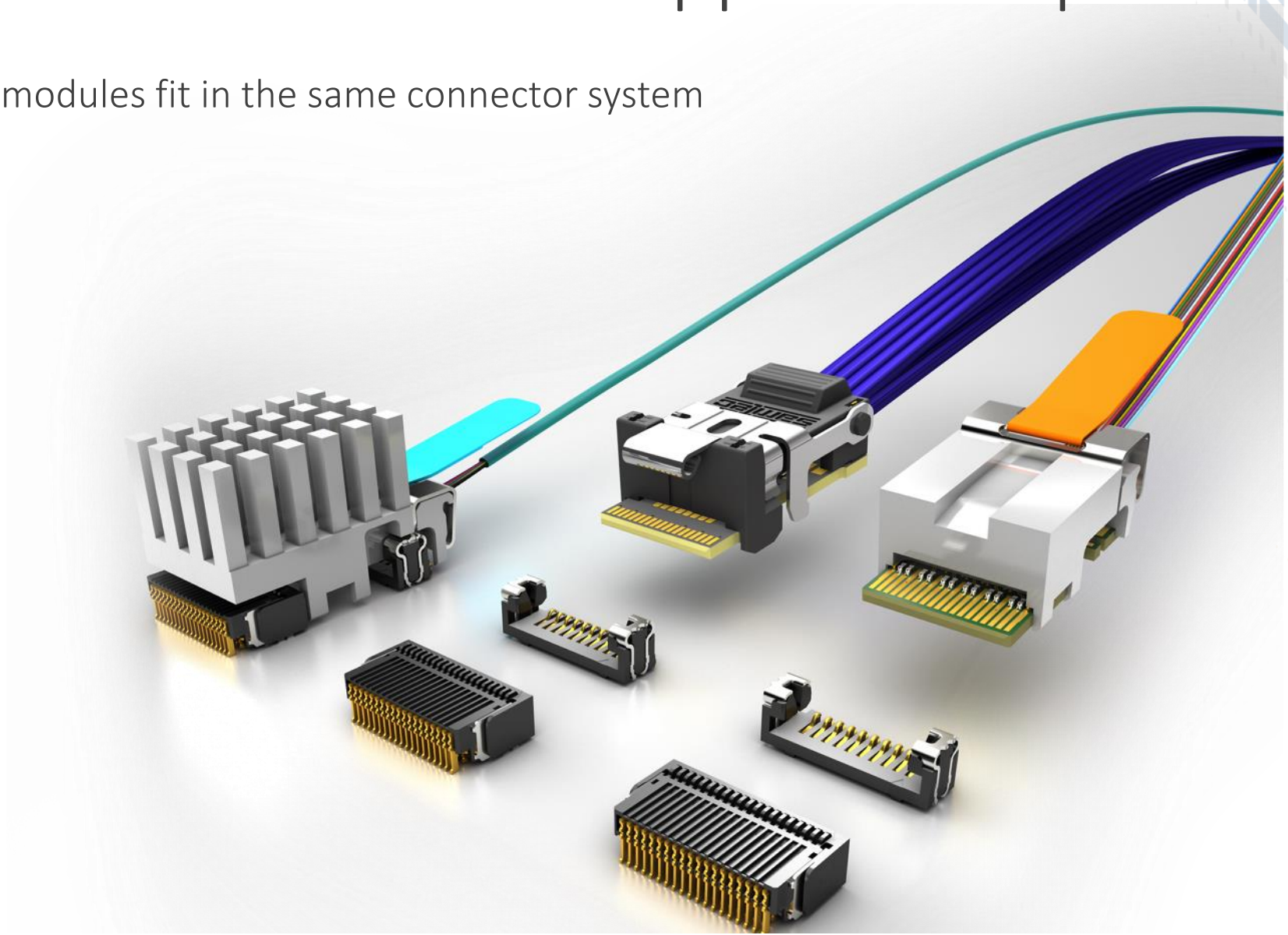
- Power and control pins
- Holds module
- Latching mechanism
- Tested to MIL standards

Now the basis for upcoming “COBO” standard



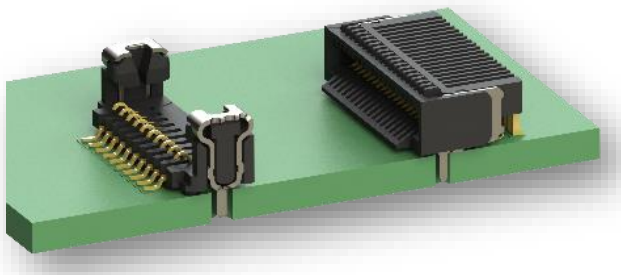
Available with both Copper and Optical Modules

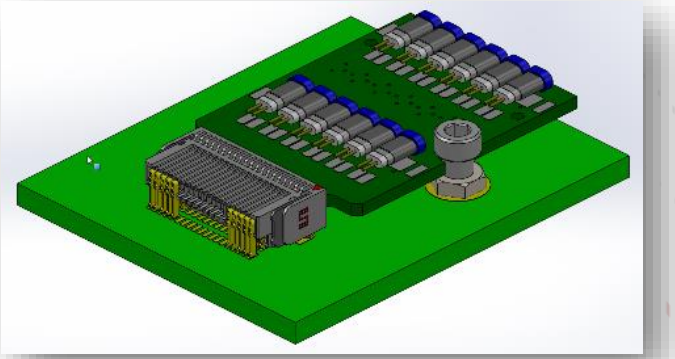
Both modules fit in the same connector system

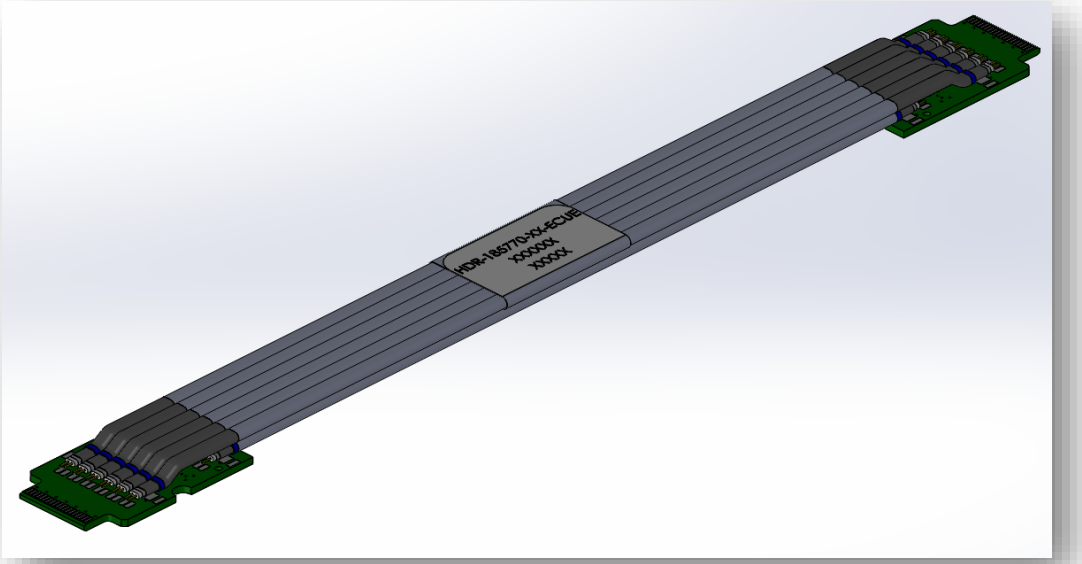


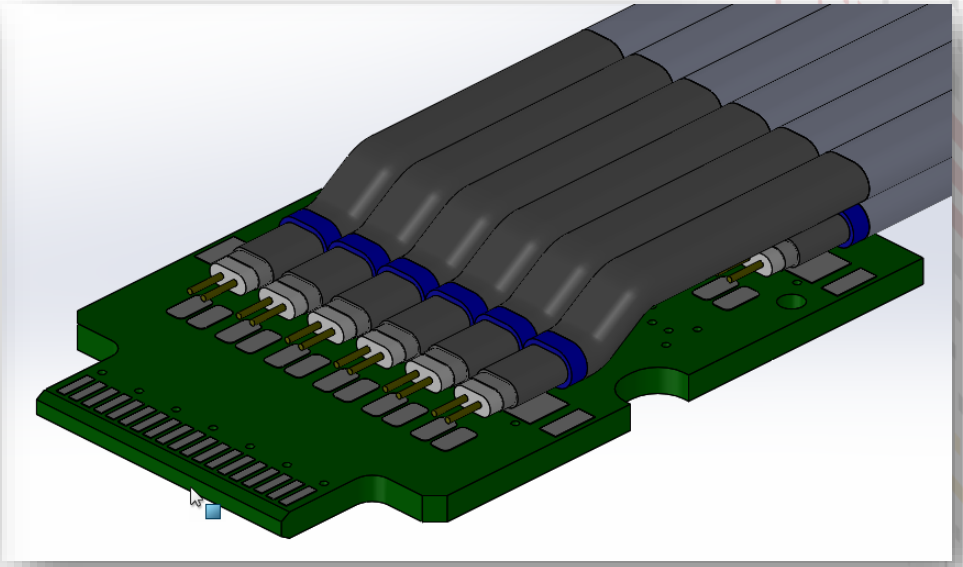
ALICE custom UEC5 cable for CERN

Customized signal mapping, gauge, dielectric material…

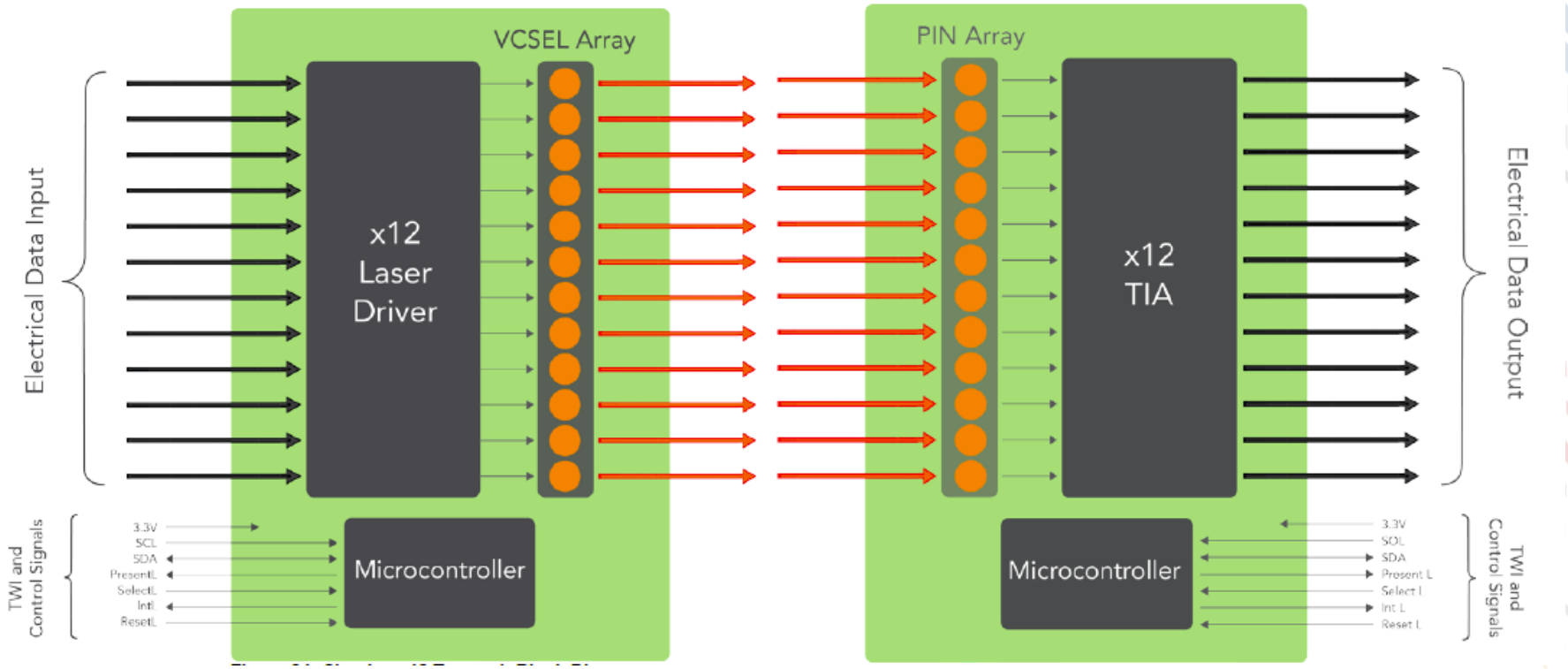








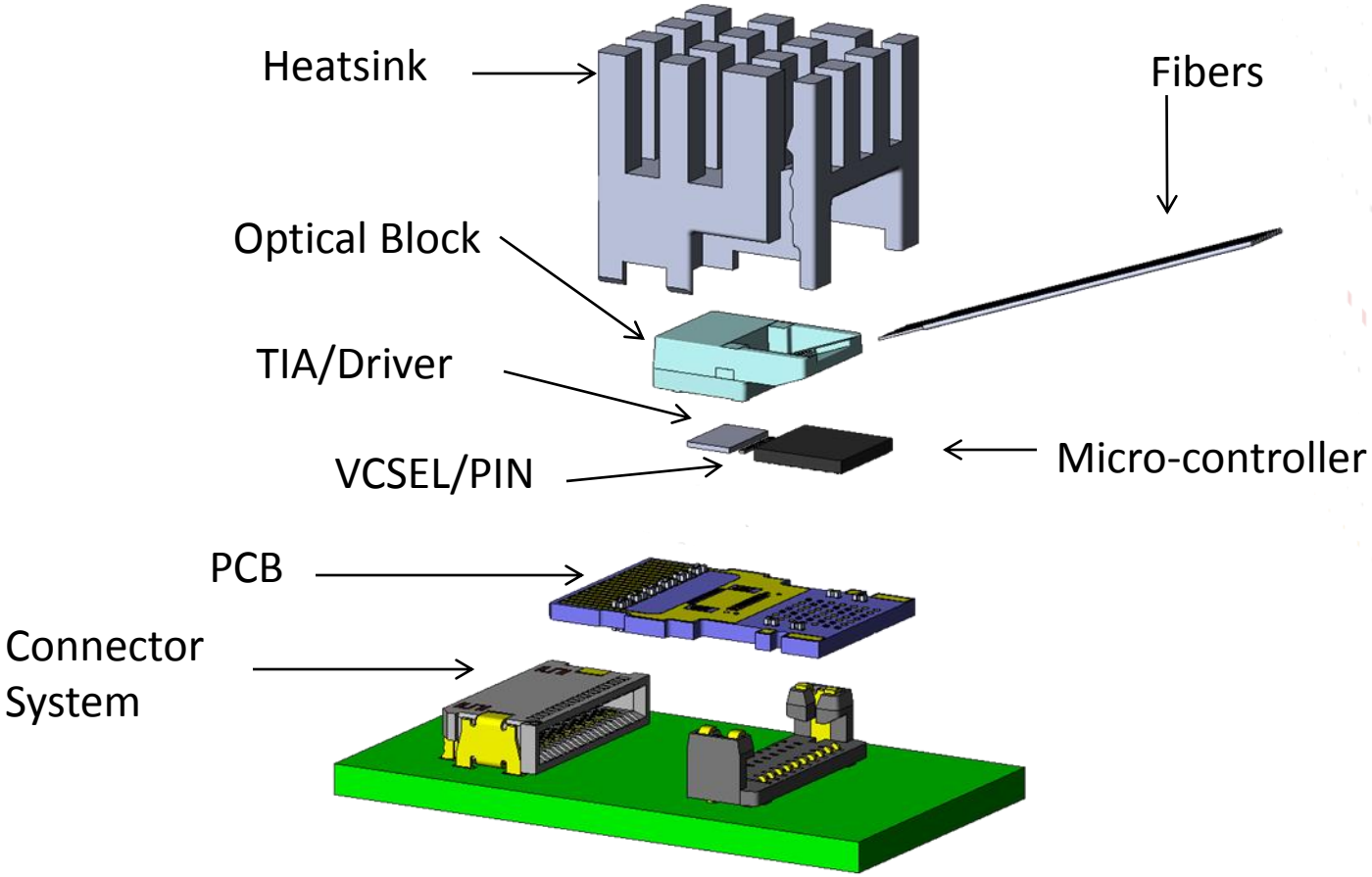
Traditional VCSEL Based Transceiver (12x 14G)



Transmitter

Receiver

VCSEL Based Optical Engine Assembly

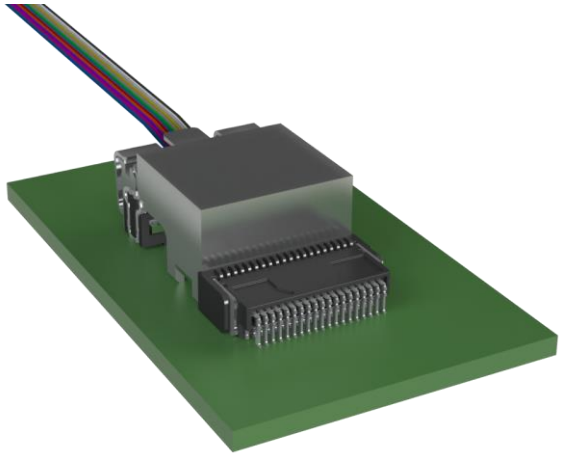


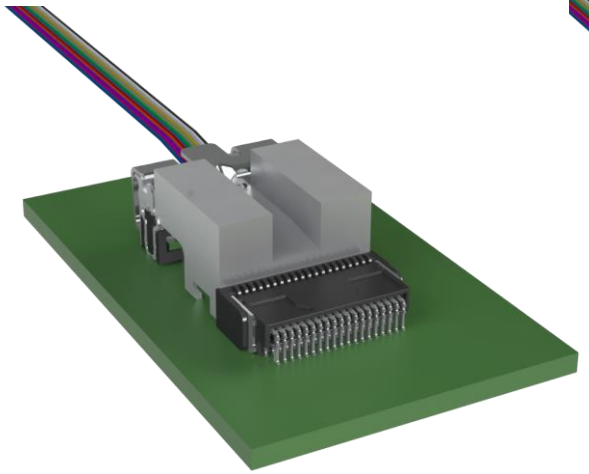
Many off-the shelf variants

			Alpha	Beta	Volume
ECUO x12	14 Gbps	Ethernet / Infiniband compatible			✓
ECUO x12	16 Gbps	Engineered link for FPGA i/o			
ECUO x4	14 Gbps	Ethernet / Infiniband compatible			✓
ECUO x4	28 Gbps	Ethernet / Infiniband compatible	✓	Q3	Q1 ‘18
ETUO x12	10 Gbps	Extended (-40/+85) Temperature Range			✓
ETUO x4	10 Gbps	Extended (-40/+85) Temperature Range			✓
PCUO Gen3	Gen3 8 GTps	X4 / x8 / x16 Gen1, Gen2, Gen3 PCIe	✓	Q2	Q3

Many Heatsinking Options

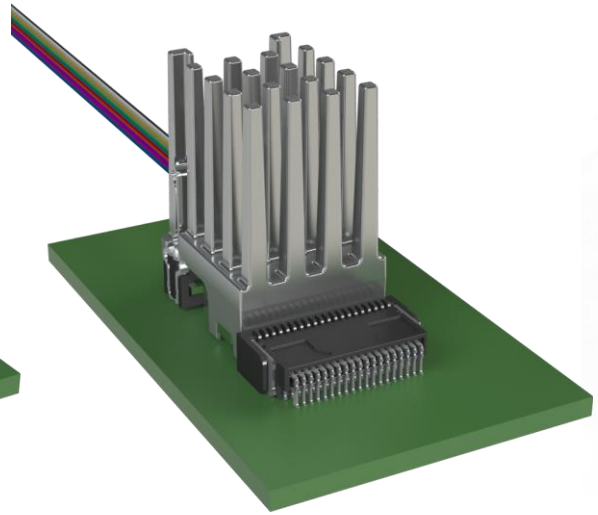
Conduction Cooling

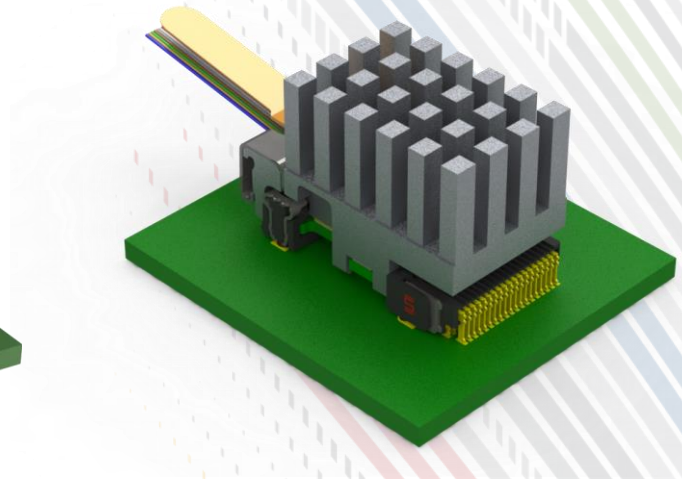




Groove allows ribbon cables to pass through which allows FireFly to be placed closer together

Convection Cooling



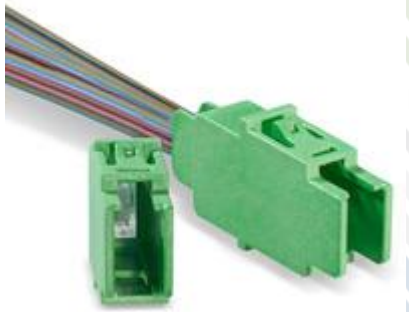


Lower profile is PCIe® CEM compliant

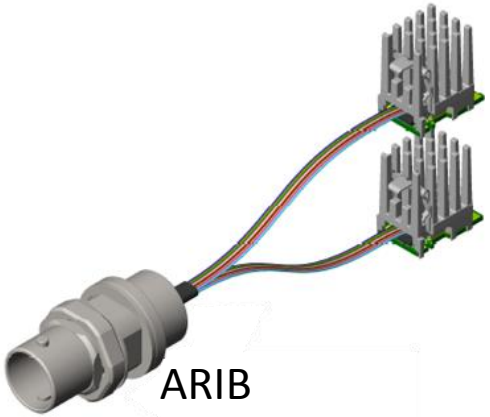
Many Optical Connectivity Options

MT

MPO

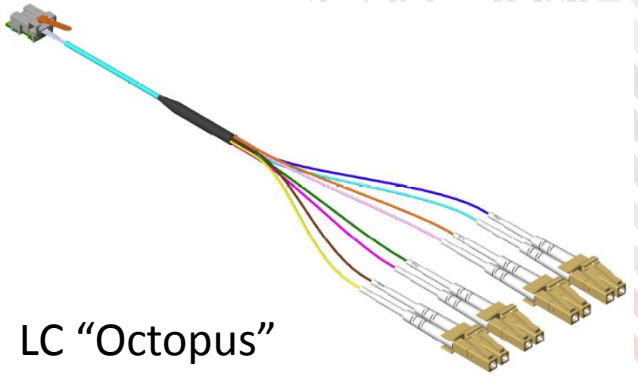


MXC



ARIB

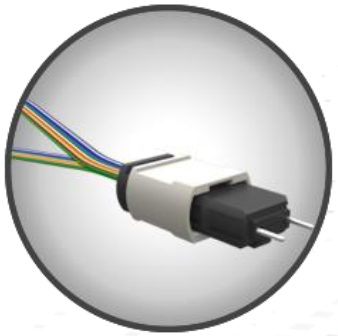
Amphenol MT38999



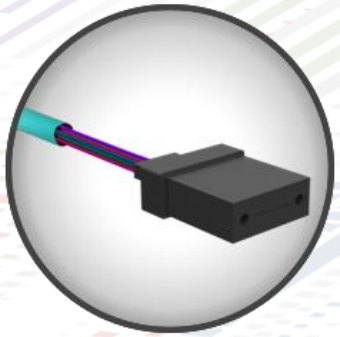
LC “Octopus”

Flexible optical connectivity solutions

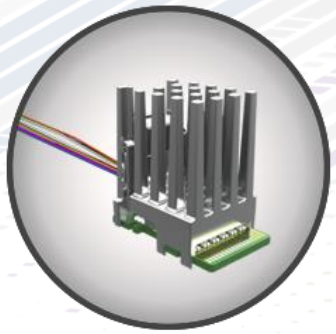
End two connector options:



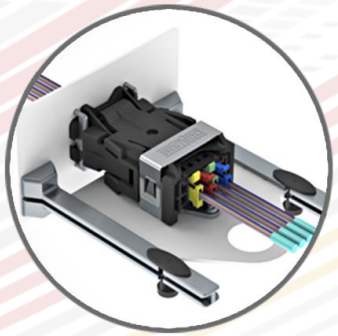
MTP® / MPO



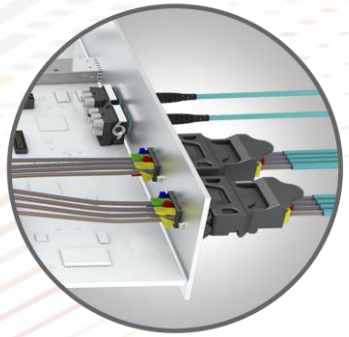
MT / MXC™



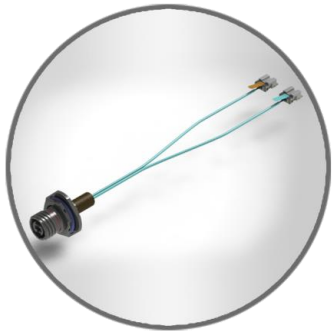
ECUO
Optical FireFly™



Ganged Backplane Optical Connector



Ganged Passive I/O



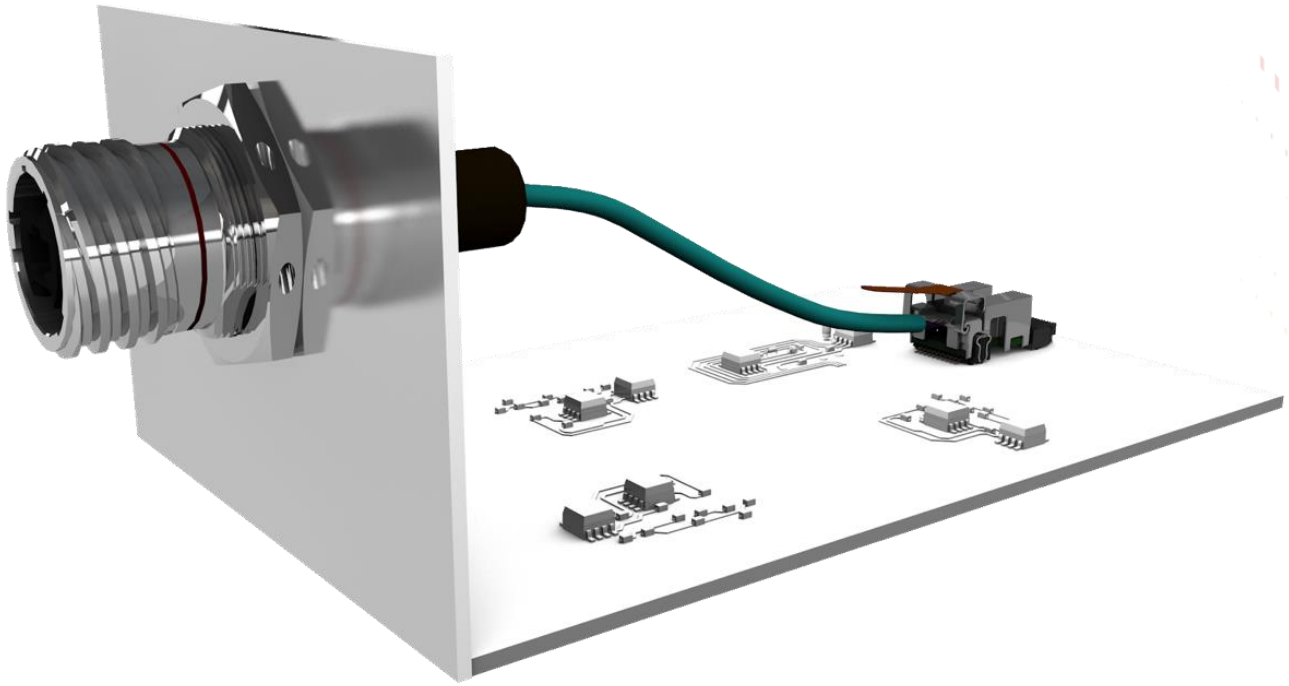
Samtec Rugged
Ext Temp FF to MT38999

*MTP is a registered trademark of US Conec Ltd. | MXC is a trademark of US Conec Ltd.

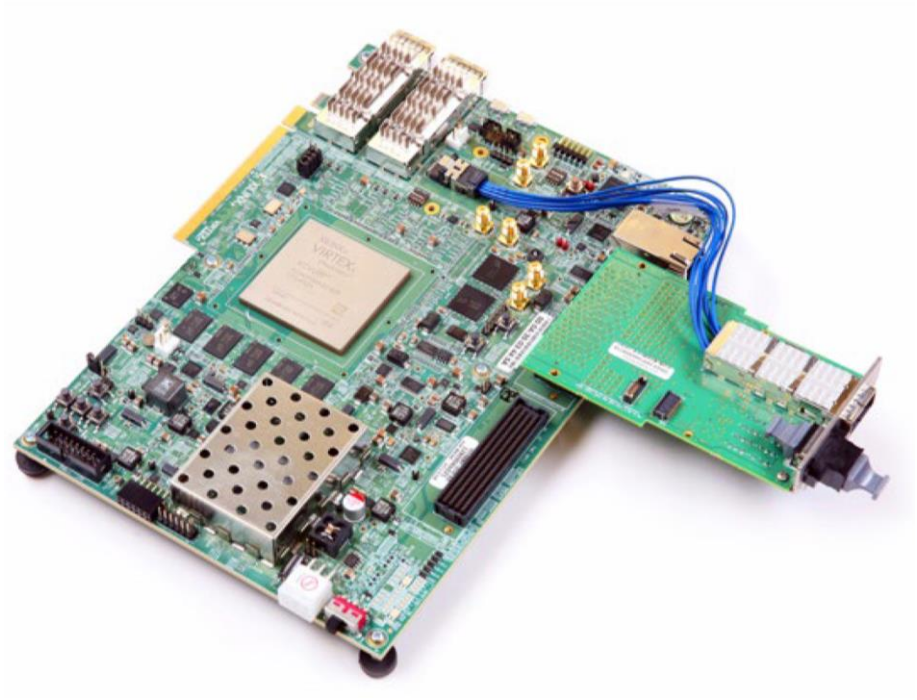
Industrial options

Extended temperature

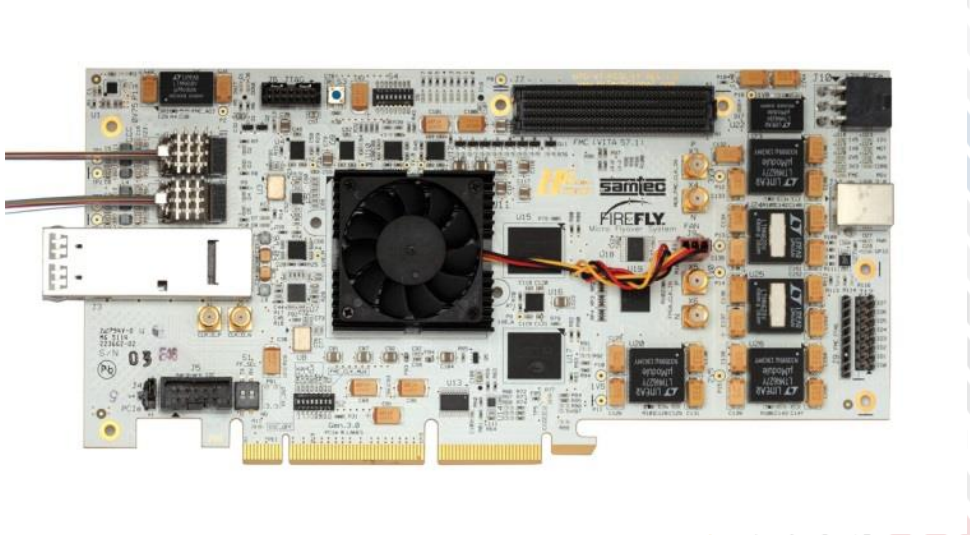
- -40/+85C (standard part is 0/75C)
- Rugged optical termination (38999, ARIB)
- MIL shake and vibe



Examples of designs: FPGA eval boards



Xilinx VCU 118
(shown with FQSFP connected to Firefly socket)

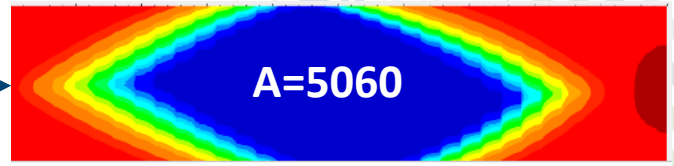


Hitech Global

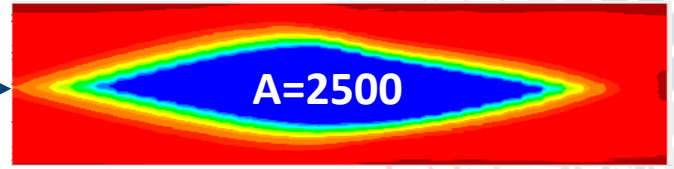
Xilinx VCU118: comparative study between solutions

25.7 Gbps

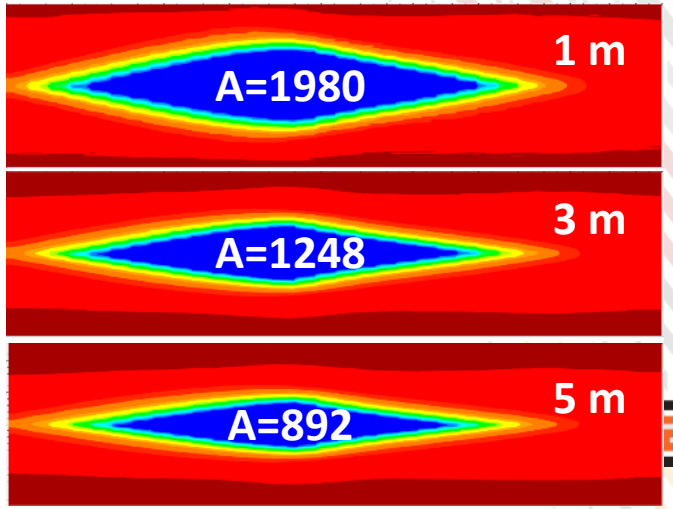
Firefly mid-board optics with 20m of fiber loopback, no CDR, DFE Off



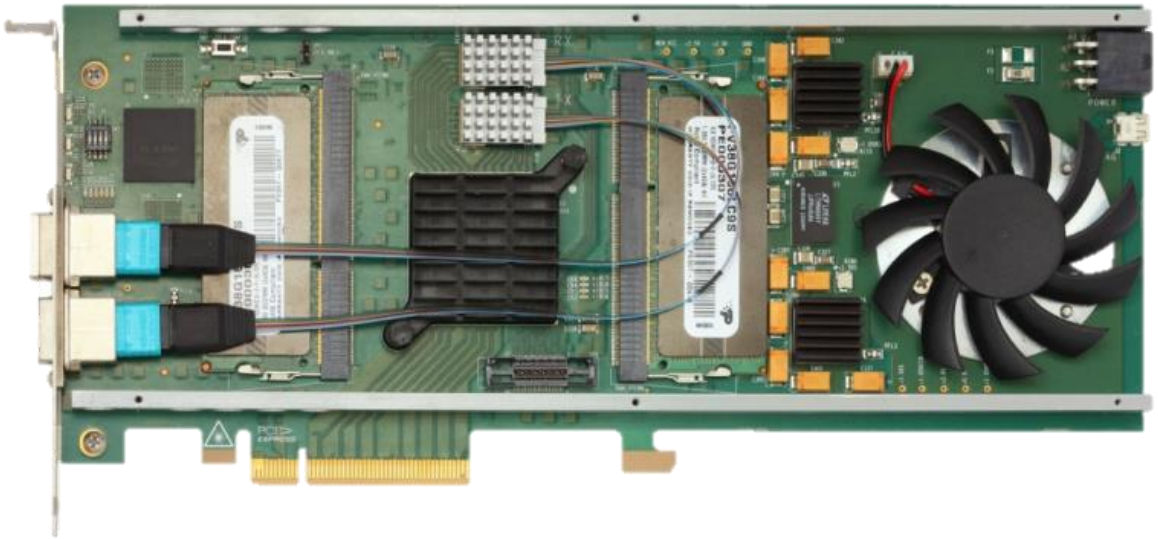
Flyover twinax to QSFP cage and electrical loopback, DFE on



Direct Attached Copper (DAC)
DFE on



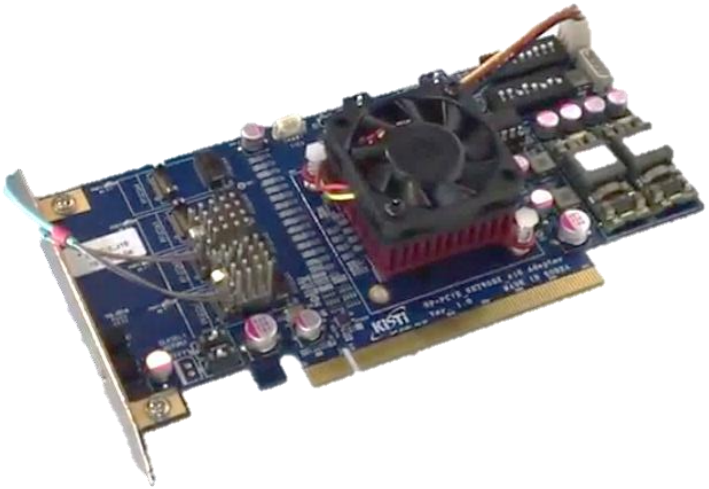
Examples of designs: Data Acquisition



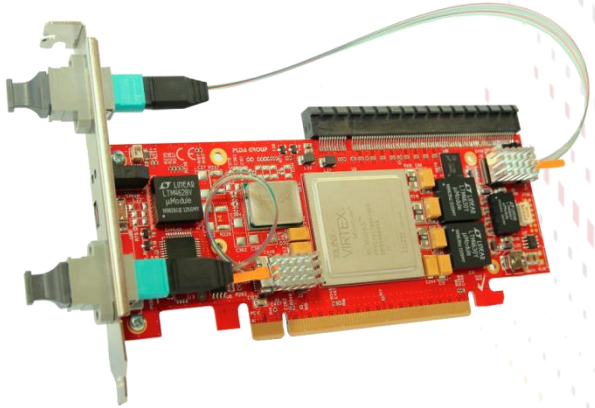
Guzik

Examples of designs: PCIe over optics

PCIe Gen3 x16
(KISTI)

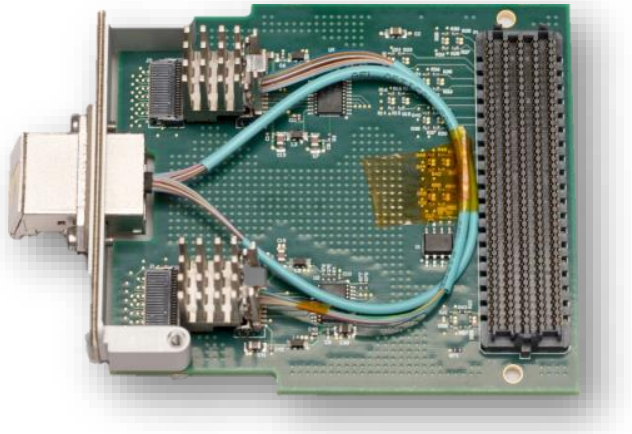


PCIe Gen4 x8
(PLDA)

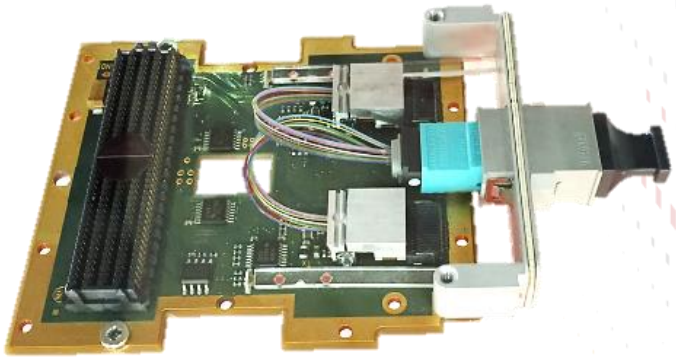


Examples of designs: FMC Cards

x12 10G FMC
(Samtec)

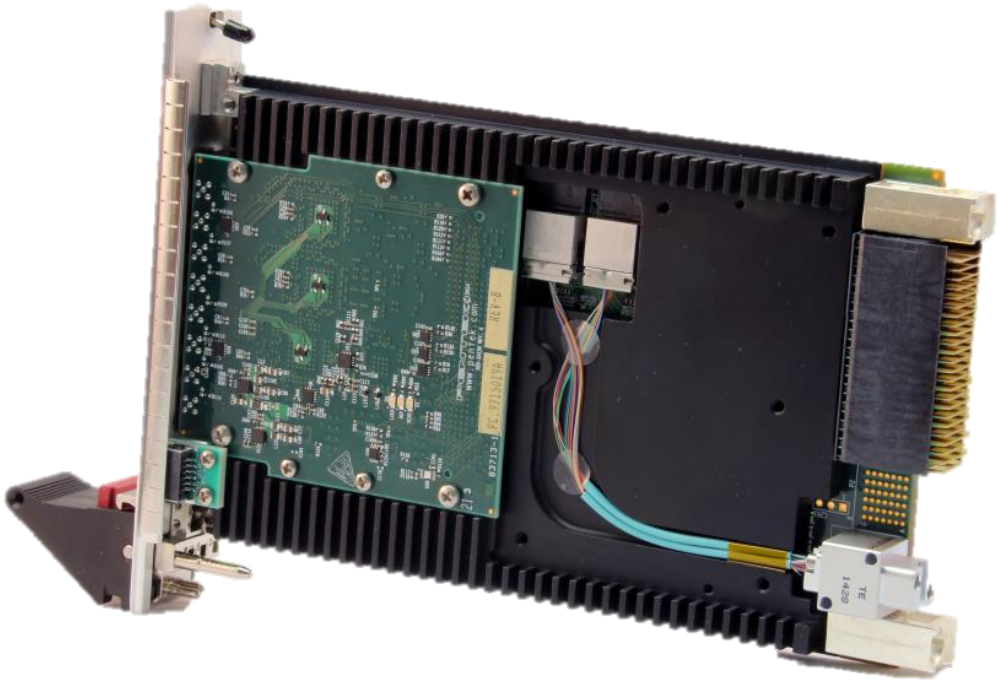


Rugged FMC
(Techway)



Examples of designs: VITA 66

Flying data acquisition hardware
(Pentek)



Firefly Onboard Optics Extended Capabilities

Lower BER rates

- All of our parts are specified 10-12 over lifetime at 28Gb per lane

-40/+85C extended temperature operation

- Available at 10G rates only at this point
- Can potentially be extended to 14 or 16G using Silicon Photonics

Vacuum operation

- Standard parts used by our customers in high-vacuum conditions

Immersion cooling

- Modified design tested successfully in 3M Fluorinert cooling liquid

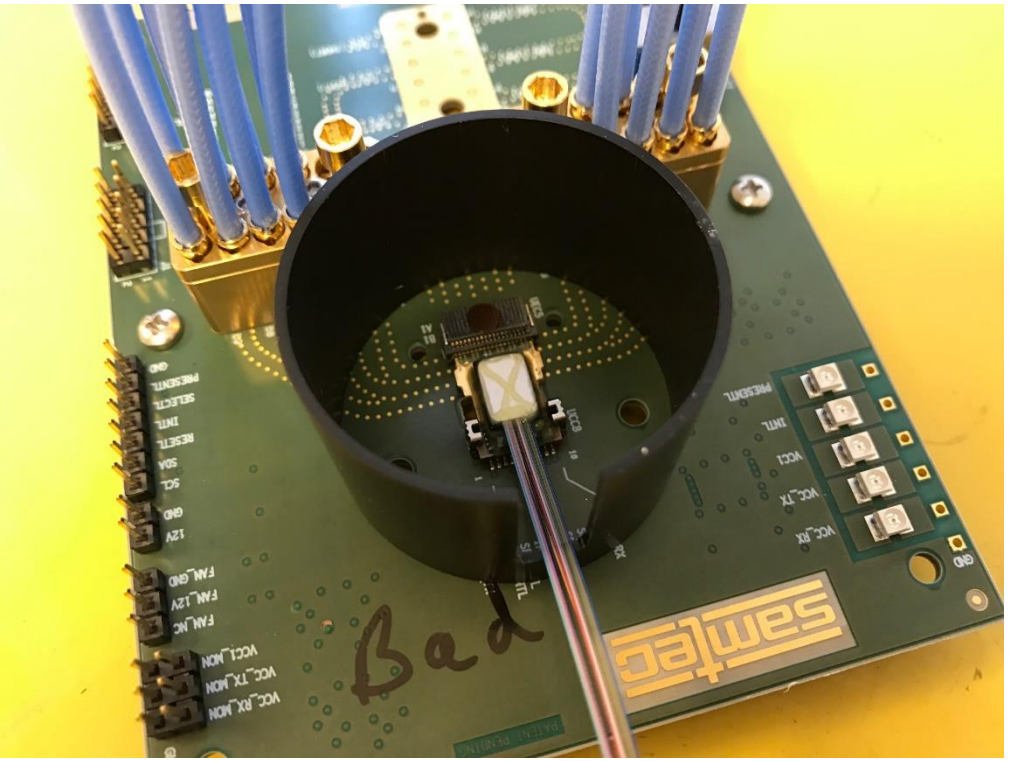
Radiation resistance

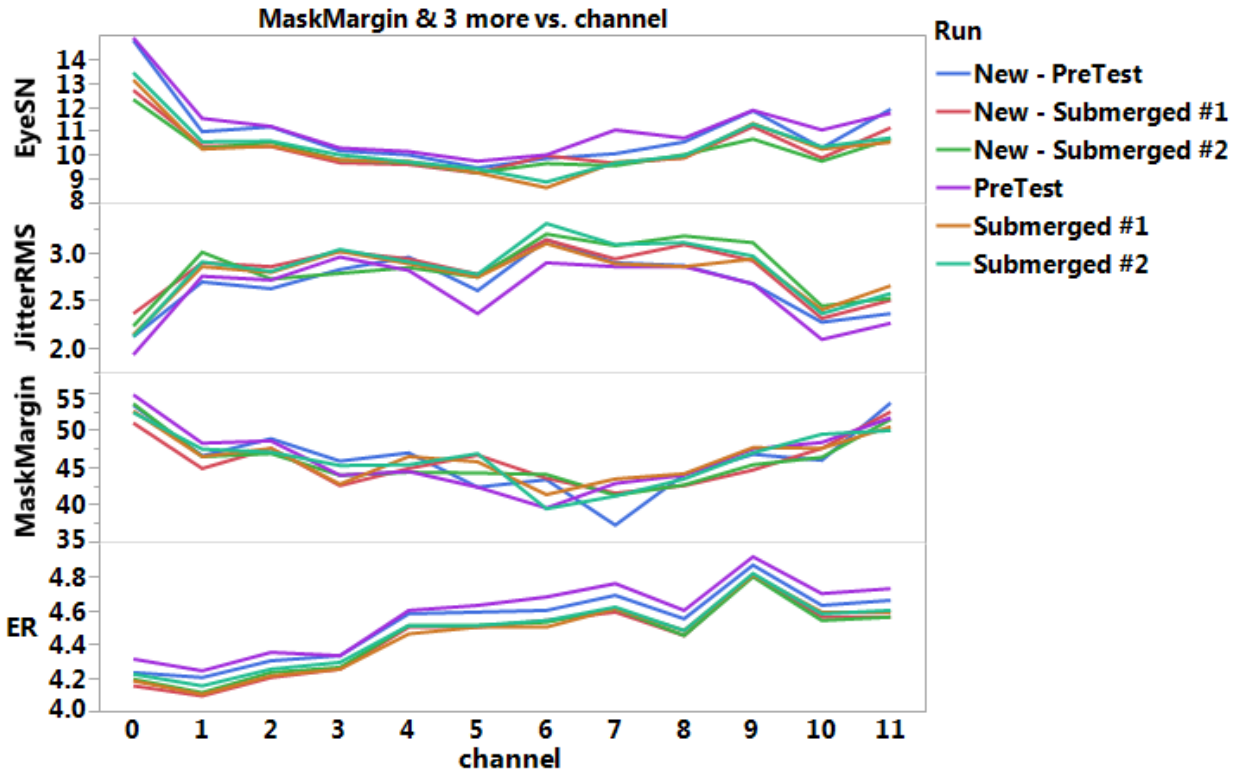
- Adapt our platform to receive customer designed Rad-hard TIA/Driver

Ultra-thin form factors

- Exploring less than 3mm thick optical engine variants

Immersion Cooled Firefly





Trends For Next Generation

Higher Speed/lane Using Advanced modulation formats

- PAM4 for 56G electrical and 110G optical
- Matches FPGA, Ethernet switches, and CPU evolution

Power consumption goes up

- More equalization electronics, CDR, PAM4 circuitry
- Effort under way to bring it back down to around a few pJ/W for close to chip on-board

BER goes up

- Standards parts have very loose BER (10-5 at 28G), requires strong FEC
- However we keep BER <10-12 at 28G, BER at still 110G unknown

Silicon Photonics Integration

- Higher speeds, single mode, but higher power consumption

Standardization

- “COBO” form Factor
- Upcoming ODI standard for instrumentation

COBO (Consortium for On-Board Optics)

Effort to standardize on-board optics

- Led by large cloud vendor (Microsoft) and switch vendor (Cisco)
- Strong Ethernet/cloud bend, not so much HPC/data acquisition
- Not that friendly to “engineered links”

Has adopted the Firefly connector mounting concept

- Mechanical outline and pinouts of modules standardized
- Optical module specs not part of the standard

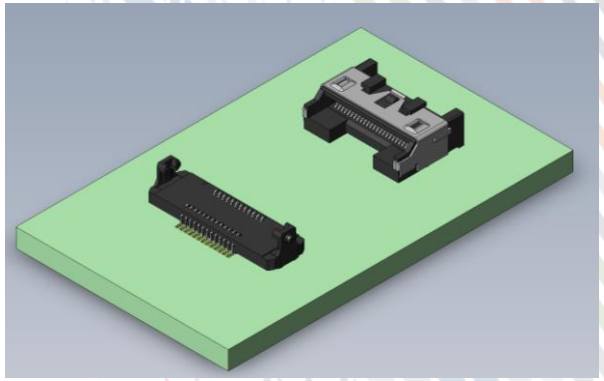
Two lane width proposed:

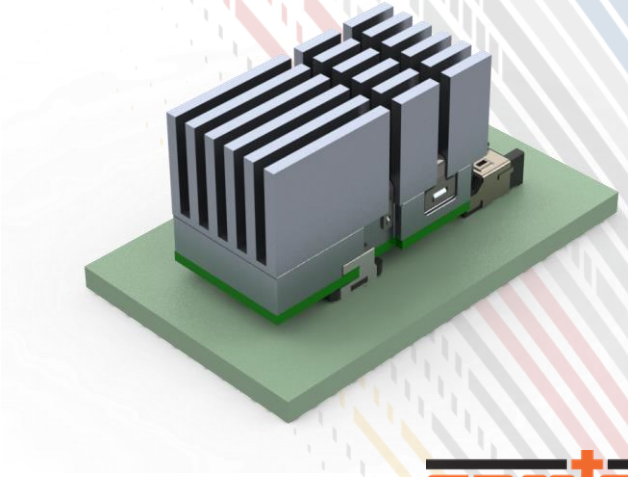
- x8, x16

3 module length proposed:

- 30mm, 40mm and 60 mm
- Rather large module height (to accommodate coherent optics)

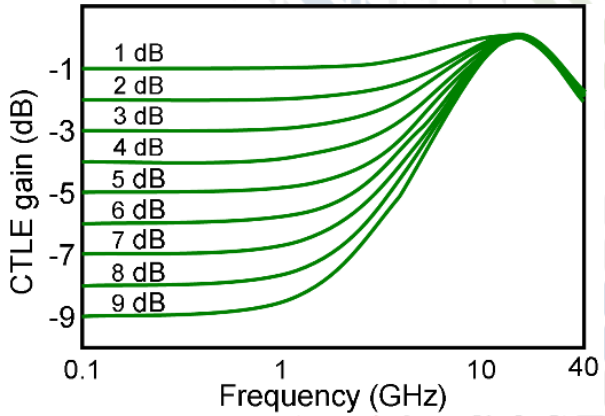
Full specification expected by mid-year with first sample at the end of 2017

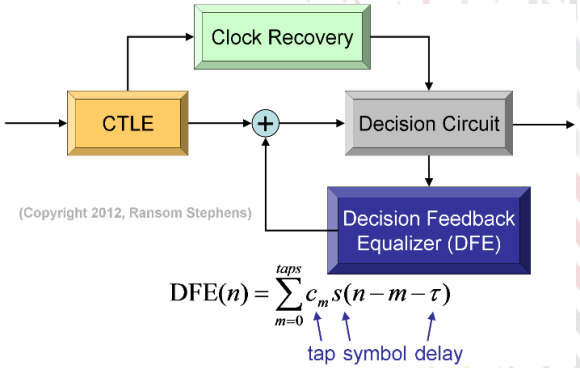




Signal Processing Techniques

- Premphasis (already used at 10G)
- Tx FFE (Feed Forward Equalization)
  - Emphasis of high-speed content to combat frequency dependent attenuation in channel
- Rx CTLE (Continuous Time Linear Equalization)
  - De-emphasizes low frequency, peak at Nyquist, filters off high frequency
- DFE (Decision Feedback Equalization)
  - Non-linear, feedback from decision circuit
  - Power hungry…
- CDR (Clock and Data recovery)
  - Used at 28G and up, in combo with CTLE or other equalization
- PAM4 Modulation (56G and up)





Clock and Data Recovery (28G and up)

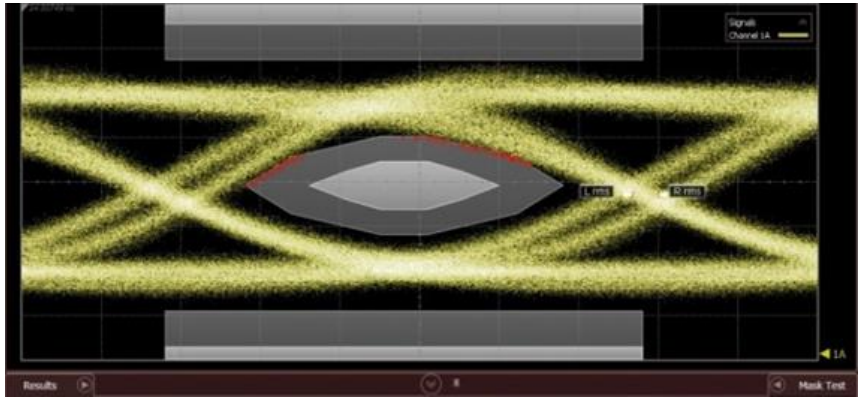
Clock and Data recovery circuit both at Tx and Rx in the optical module

- Decision circuit, retimes the signal

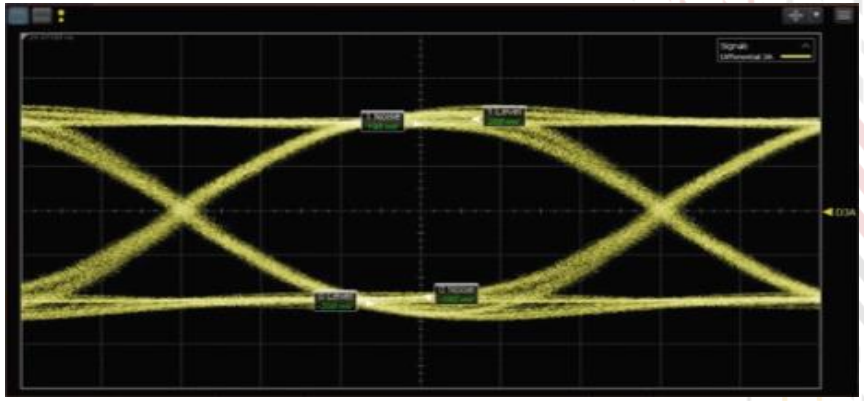
Significant improvement in BER (down 10^{-12} in our parts), but does not come for free:

- Not rate independent: PLL only locks within narrow bit rate (around 25.7G and 28.1G)
- Some increase in power consumption

28G NRZ, no CDR



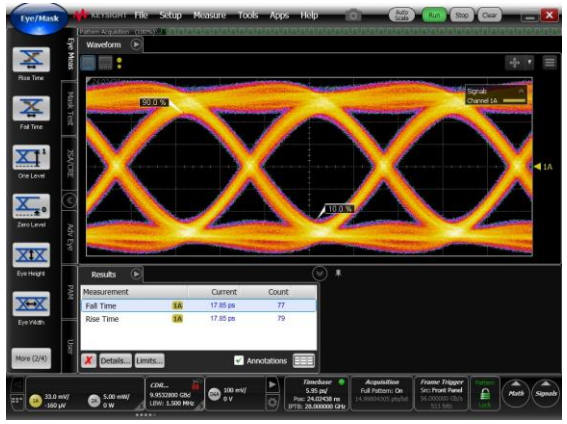
28G NRZ, with CDR



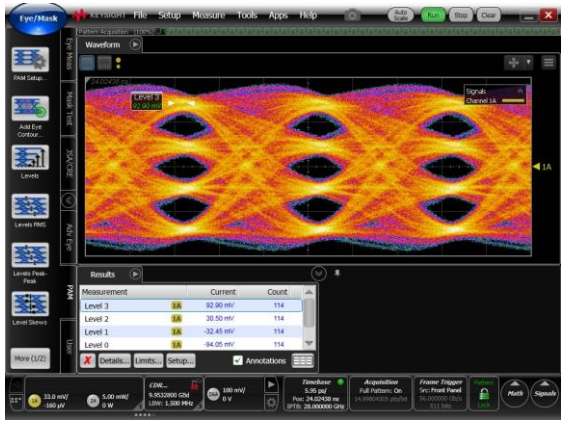
PAM4 Modulation: 56G and up

- Use 4-level signaling (PAM4) on both electrical and optical side
  - 28GBaud, PAM4 electrical in: 8 lanes at 56 Gb/s
  - 56GBaud, PAM4: 4 fibers at 112Gb/s optical out
- Goes faster, but doesn't come for free either
  - Increased electronic power consumption
  - Lower SNR: BER increases

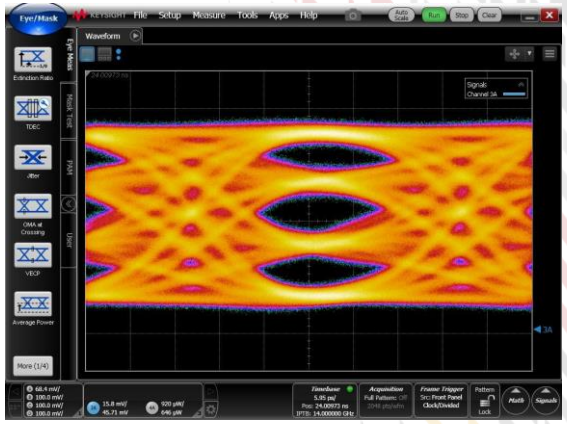
56G NRZ



56GBaud PAM4 (112 G)



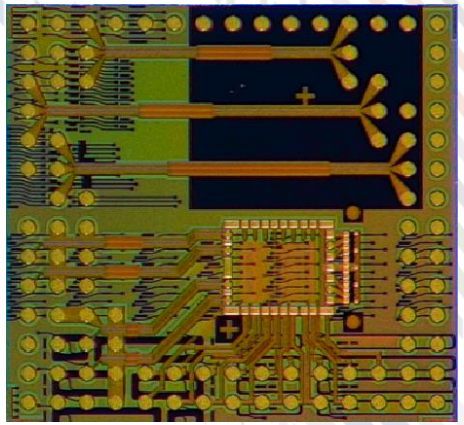
56GBaud Optical Out



SILICON PHOTONICS

Traditional VCSEL (Vertical Cavity Surface Emitting Lasers)

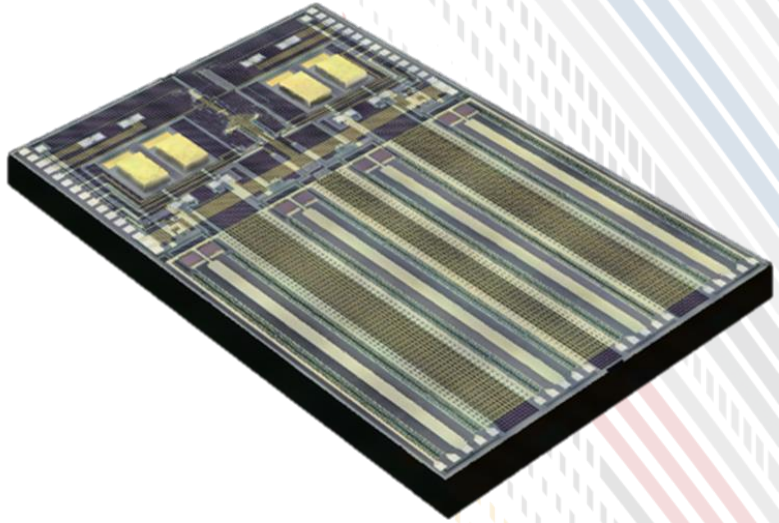
- Maximum NRZ rate is about 28Gb/s
- can achieve 56Gb/s using PAM4 (28 GBaud)
- Limited distance (<100m, multimode fiber) and BER
- But: lowest power consumption, lowest packaging cost



Courtesy LETI

Silicon Photonics

- Integrated optical components on a Silicon Wafer, using silicon manufacturing technology
- Faster modulators: 56Gb/s and 112Gb/s (56 Gbaud)
- Longer distance (500m to 2 km), due to single mode fiber
- Higher power consumption, higher packaging cost



Courtesy Macom

SiPho integration gives a path to 100Gb/s and much beyond

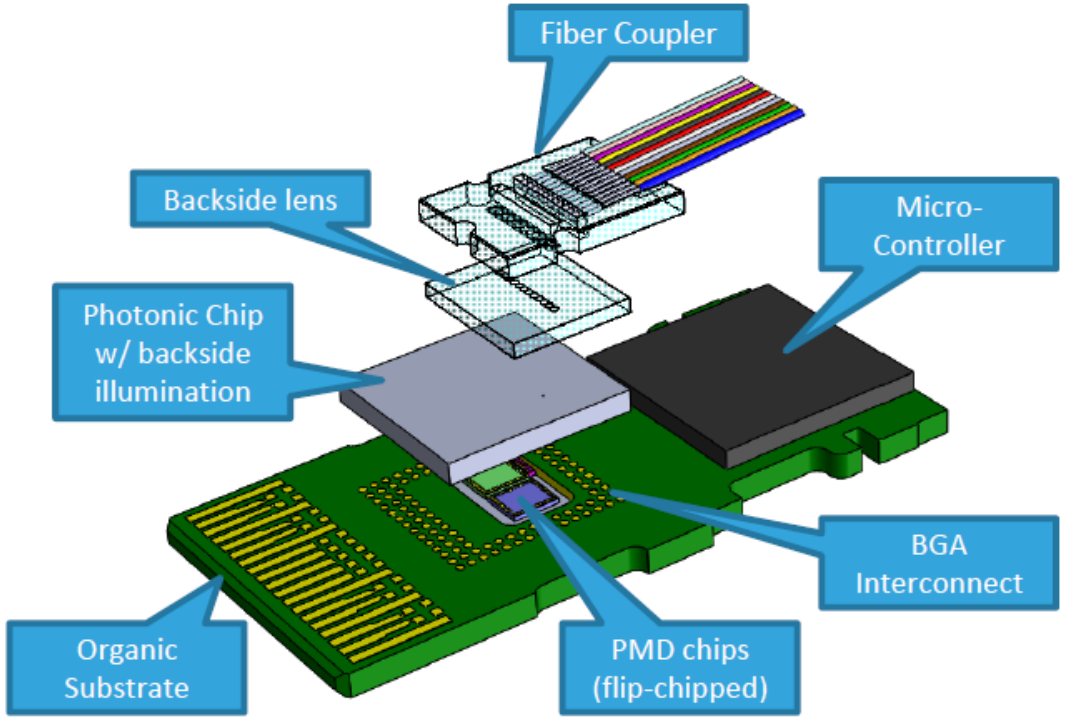
- Use the integration and WDM (Wavelength Division Multiplexing)
- We are working to reduce power consumption in line (~2.5W at 800 Gb/s)
  - Assumes on-board optics and close proximity to the FPGA

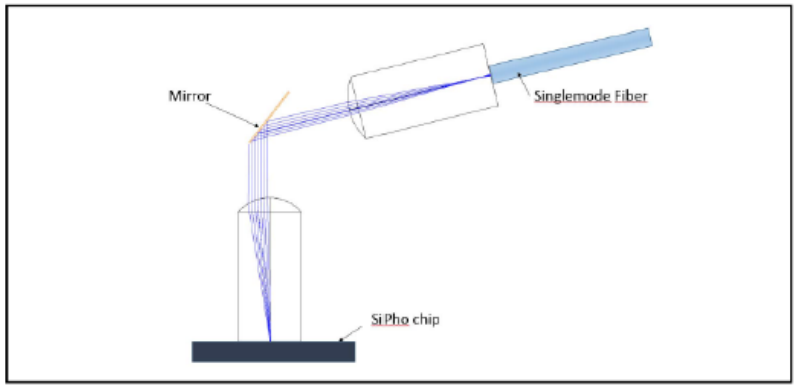
You still need to package the SiPho Chip!

Multidimensional submicron alignment

- Single mode alignment requires 10x the precision of multimode VCSEL alignment

Laser, TIA and DRVR chip still hybridized





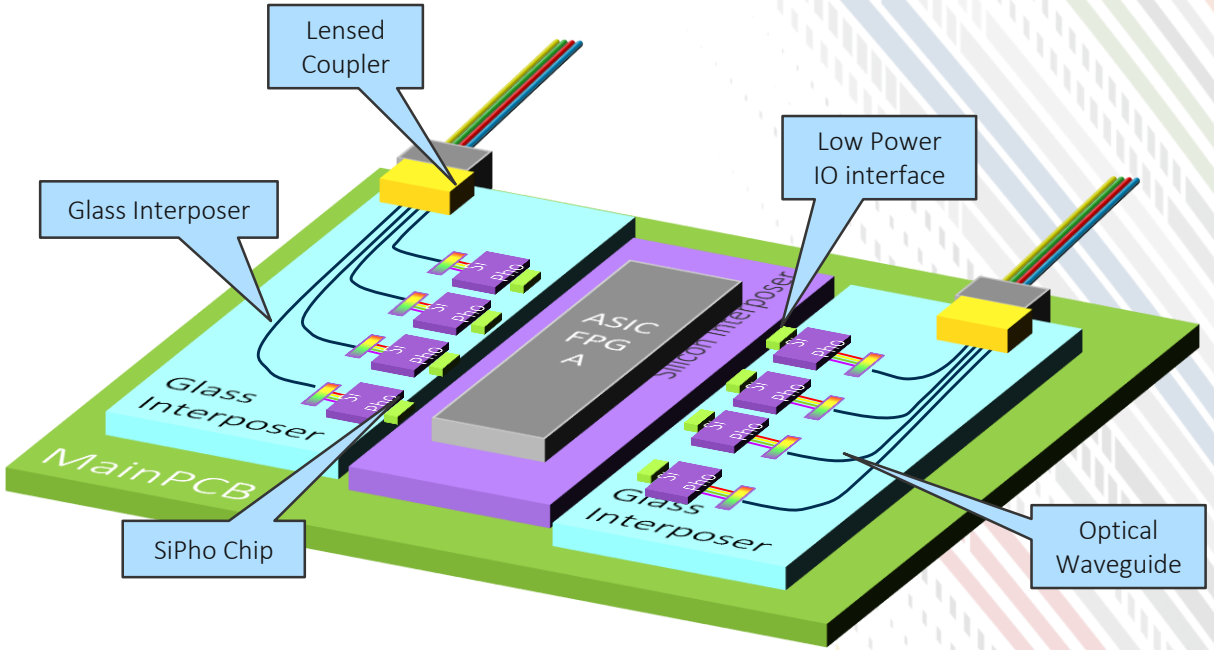
Long Term: Photonics Co-Packaging

Optics and ICs next to each other
- Improves signal integrity
- Decreases power consumption
- Increases density

Significant packaging challenge
- Opto-mechanical constraints
- Thermal challenge

New materials help
- Silicon Photonics Integration
- Glass interposers

Monolithic integration still a long ways off
- ICs much too optimized
- SiPho processes mostly incompatible



One last word on Copper…

Copper is not standing still…

Up to 56G PAM4 Performance

Cable-to-Cable

Move high-speed signals off the backplane/midplane and the line card

Cable-to-Board

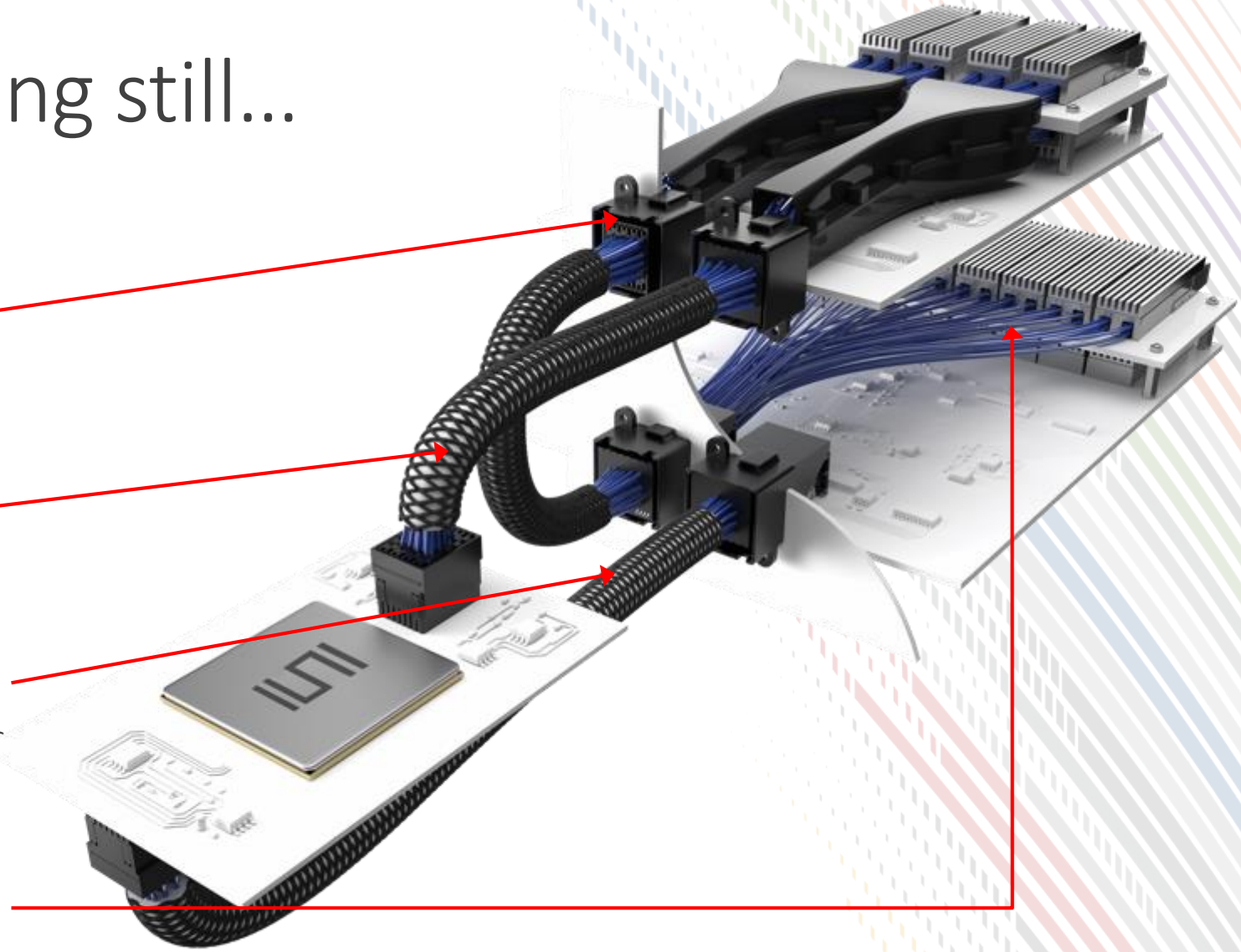
Flyover backplane/midplane applications or adjacent to the chip on a host board

Cable-to-ExaMAX®

Pluggable flyover backplane/midplane mates with existing ExaMAX® right-angle and vertical connectors

Cable-to-Panel

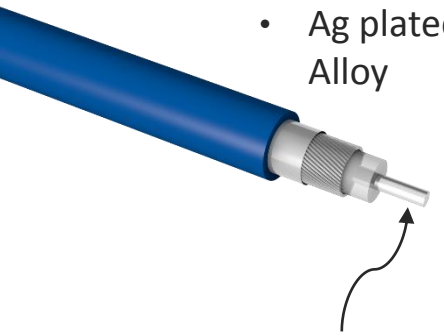
Route signals from the backplane/midplane directly to the panel for less transitions



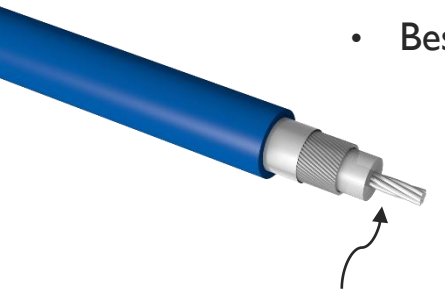
Configurable building blocks

Center Conductor

- 24awg through 32awg
  - Ag plated Cu
- 34awg through 38awg
  - Ag plated Cu high strength Alloy



- Solid
  - Best SI Performance



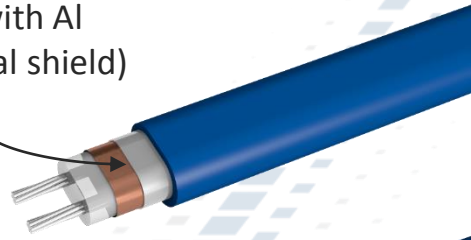
- Stranded
  - Better suited for dynamic environments

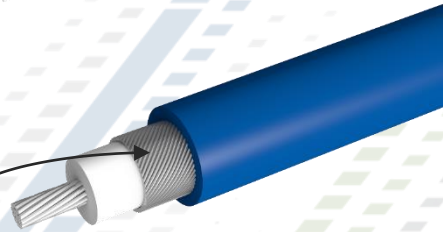
Dielectric

- **FEP (Ultimate in Performance and Stability)**
  - High Temp (200C)
  - Available in Both:
    - Solid (best dimensional stability)
    - Microcellular Foam (less material, smaller size, higher speed)
- **LSZH (Low Smoke Zero Halogen)**
  - Meets new emerging compliance demands
- **LDPE (Lowest Cost)**
  - Lower Temp applications 85˚C)

Shielding

- Copper + Metal Clad Combo
  - Copper combined with Al Mylar (best electrical shield)





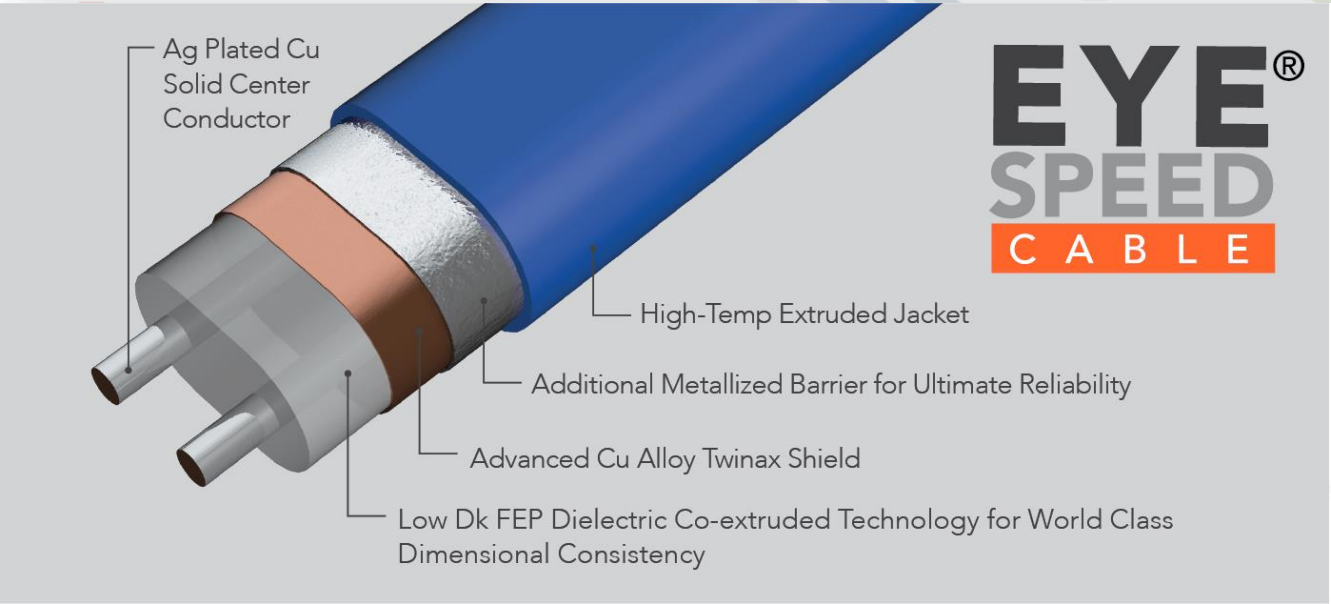
- Serve (spiral)
  - Effective for most applications, lowest-cost, flexible -- Samtec specialty

- Braid
  - Flexible, effective, expensive, slow to process

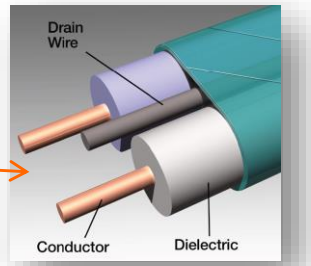
eye speed® ultra low skew twinax

Low Loss, High Performance Design Optimized for 28+ Gbps System Demands

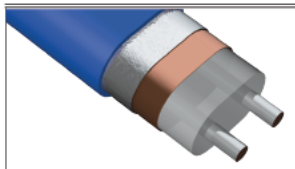
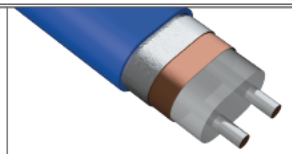
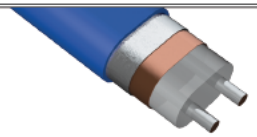
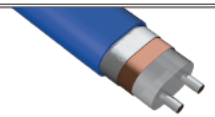
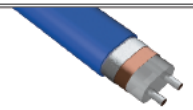
- Proprietary cable technology
- Ideal for 28 - 56+ Gbps applications
- Tight coupling between signal conductors
- Improved bandwidth and reach
- Improved signal integrity
- Ultra low skew twinax < 3.5 ps / meter



Contrast with
conventional approach



Ultra-low skew twinax

Nominal Performance Specifications			28 AWG	30 AWG	32 AWG	34 AWG	36 AWG
							
14 GHz (28 Gbps)	0.25 m	IL (dB)	-1.0	-1.2	-1.5	-1.8	-2.2
	1.00 m		-3.9	-4.7	-5.9	-7.2	-8.7
28 GHz (56 Gbps)	0.25 m		-1.5	-1.8	-2.2	-2.6	-3.2
	1.00 m		-6.0	-7.0	-8.7	-10.6	-12.7
Density / Flexibility			Good	Good	Better	Best	Best

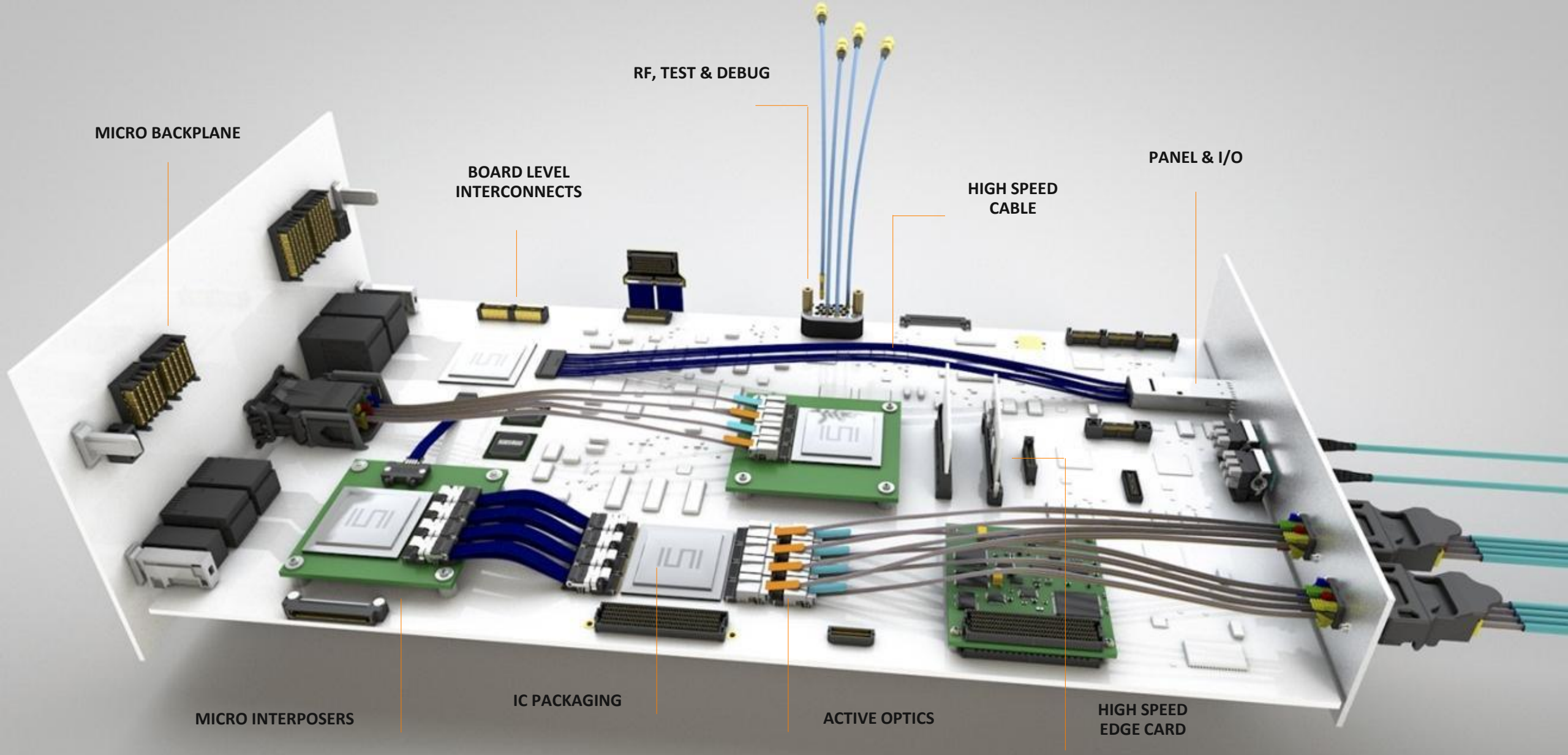
Eye Speed® Ultra Low Skew Twinax Cable is available in engineered impedance configurations of 85 Ω, 92 Ω and 100 Ω.

New 56G NRZ / 112G PAM4 connectors

- 56G NRZ capable / 112 PAM4 per channel
  - Very low crosstalk to 40 GHz+
  - Very tight impedance control
- BGA attach: high density, improved breakout region
- Target impedance is 92.5 ohms
- Mezzanine mated stack height: 7 mm to 20 mm
- Right angle options
- 8 to 72 signal pairs

NOVARAY™

samtec


RF, TEST & DEBUG
MICRO BACKPLANE
BOARD LEVEL INTERCONNECTS
HIGH SPEED CABLE
PANEL & I/O
MICRO INTERPOSERS
IC PACKAGING
ACTIVE OPTICS
HIGH SPEED EDGE CARD

Thank You!

Thank You!