

# 100 GBit Ethernet activities at DESY

- **Introduction**
- **FPGA Evaluation Board**
- **PC with 100GBit network**
- **Accessories**
- **Firmware / Software**
- **Test equipment**
- **Outlook**

Frantisek Krivan , Manfred Zimmer, DESY  
Workshop DTS Platform KIT 16-17.10.2017

# Introduction

- > Purpose of current activity is to get a first experience with 100Gbit technology

- > Massive data rates

Duration	Bits	Bytes
1 sec	100G	125M
1 min	6T	750G
1 h	360T	45T
1 day	8.64P	1.08P
1 week	60.48P	7.56P
1 month	1814.4P	226.8P
1 year	22075.2P	2759.4P

- > Possible future applications

- Typical data rate of

- > Technology is currently

- FPGAs, evaluation

- > Following developments

- Loopback test between

- Test between 2 Evaluation boards and between Evaluation board and PC

- Development of simple passive FMC board with QSFP adapters to get experience with >25 GBit routing

- Finally: FPGA based DAQ card with 100 GBit port and FMC connector

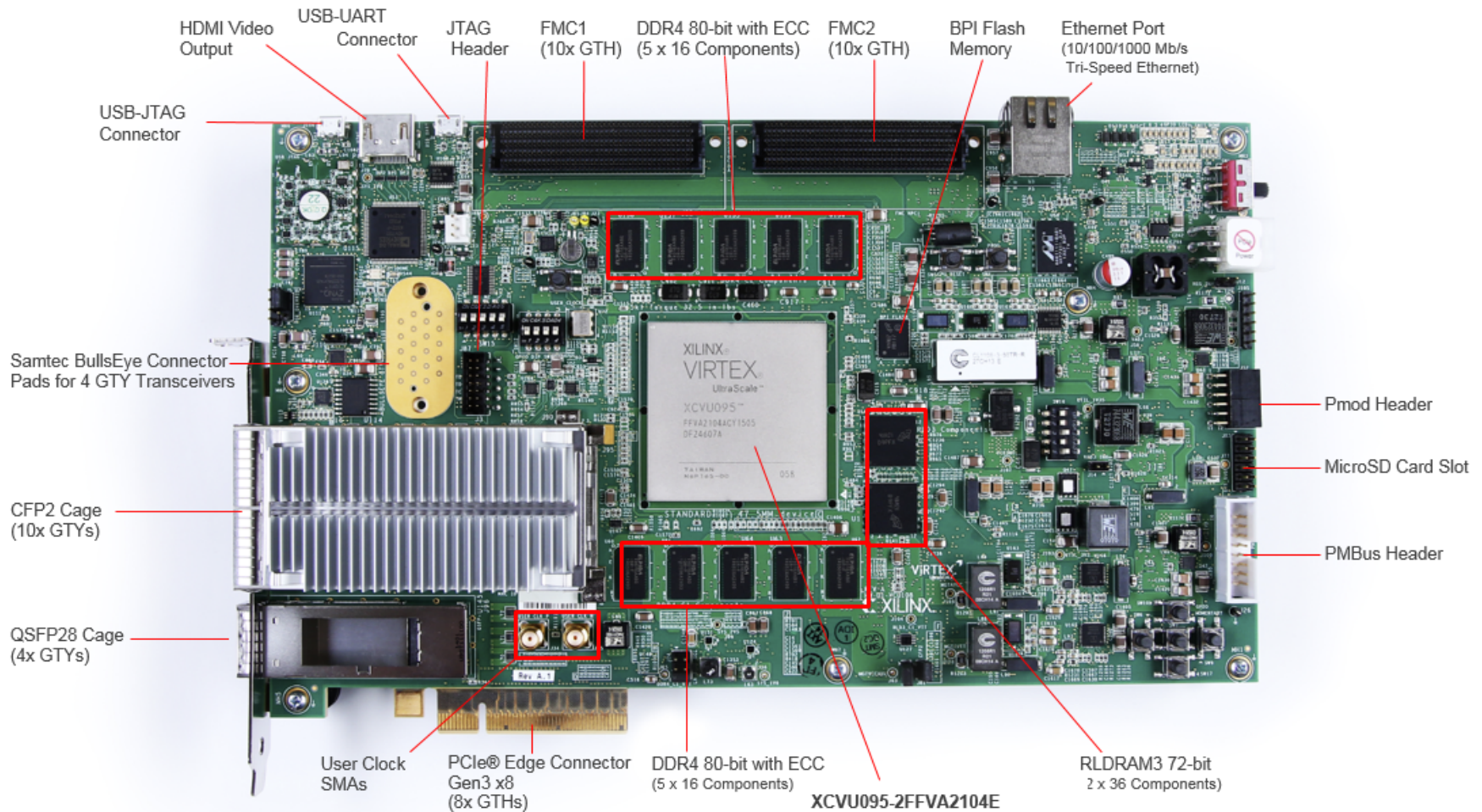


# Data vs. Time

Time	Bit size	Byte Size
1 sec	100G	125M
1 min	6T	750G
1 h	360T	45T
1 day	8.64P	1.08P
1 week	60.48P	7.56P
1 month	1814.4P	226.8P
1 year	22075.2P	2759.4P

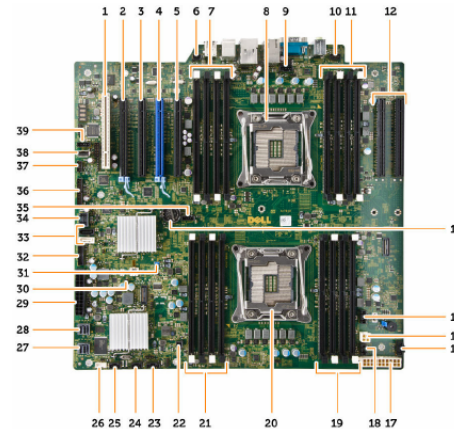


# Evaluation board - VCU108



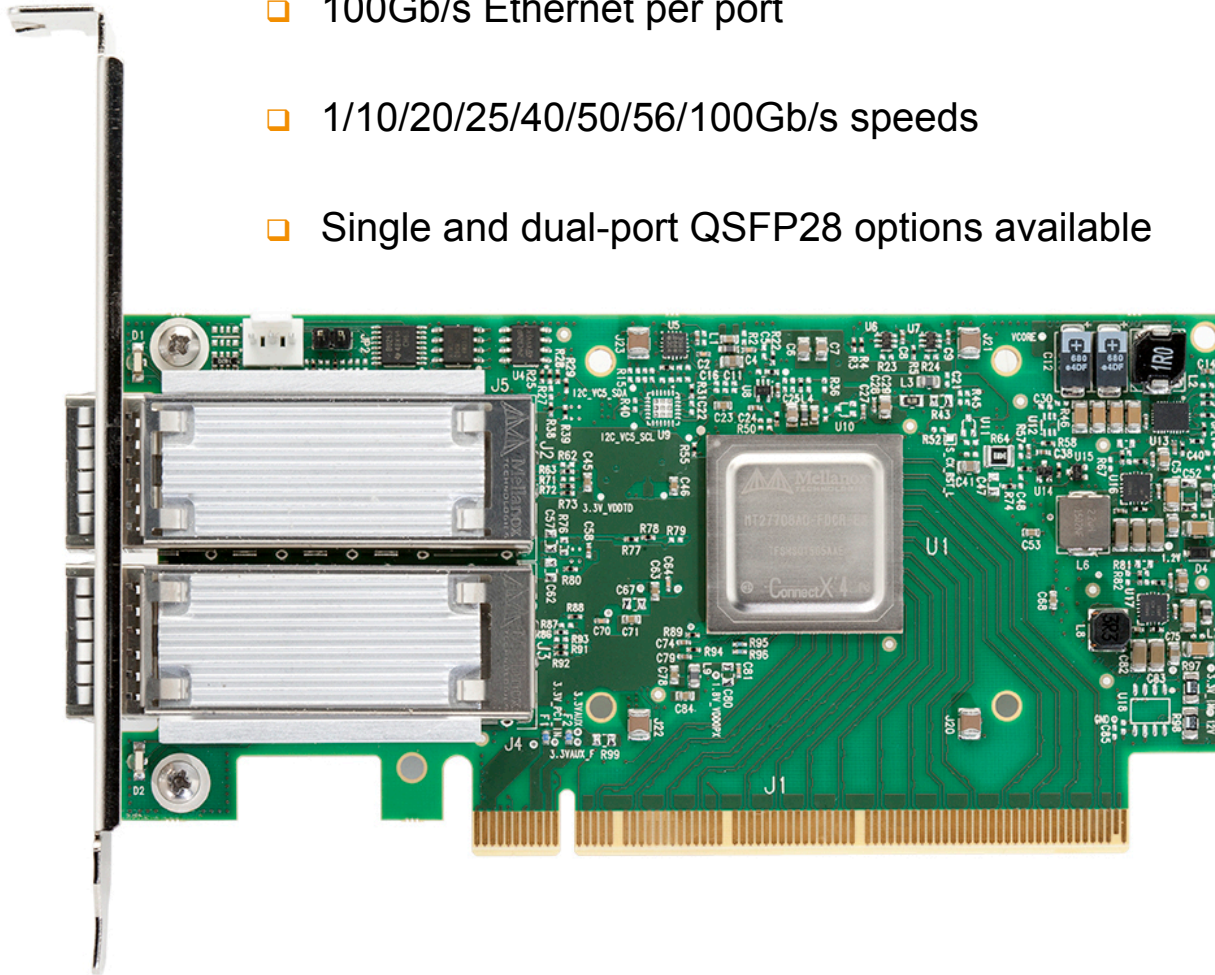
# DELL Precision Tower 7910

Feature	Precision Tower 7000 Series (7910) Technical Specification	
Processor Options	One or two Intel® Xeon® processor E5-2600 v4 Series with up to 22 cores and Intel Advanced Vector Extensions, Intel Trusted Execution Technology, Intel AES New instructions, Optimized Intel Turbo Boost and optional Intel vPro™ technology	
Operating System Options	Windows® 10 Professional (64 bit) Windows® 8.1 Professional (64-Bit) Windows® 7 Professional (64-Bit)	Red Hat® Enterprise Linux® 7.0 and 7.2 (required for Xeon E5-2600 v4 CPUs), RHEL 6.5 supported – future 6.8 for E5-2600 v4 ) Ubuntu 14.04 NeoKylin 6.0 (China only)
Chipset	Intel® C612	
Memory Options	Quad channel memory up to 1TB 2400MHz DDR4 ECC memory with dual CPUs, 16 DIMM Slots (8 DIMMs per CPU). Note: memory speed is dependent on specific Intel Xeon E5-2600 v4 Series processor installed	
Graphics <sup>4</sup> Options	Support for 4 PCI Express® x16 Gen 2/Gen 3 graphics cards - up to 675W (total 3 x 225W graphics cards in 3 slots)	
Chassis	<b>HxWxD:</b> 16.95 x 8.50 x 20.67"; 43.05cm x 21.59cm x 52.50cm <b>Bays:</b> (1) external slimline optical bay; (4) internal 3.5" bays (support total of (4) 3.5" or 2.5" drives); (1) external 5.25" bay (up to 4 2.5" drives) <b>Slots:</b> (2) PCIe x16 Gen 3 [(2) more with 2 <sup>nd</sup> CPU]; (1) PCIe x16 Gen 3 [wired as x4 – Slot 1], (1) PCIe x16 Gen 2 [wired as x4]; (1) PCI 32Bit <b>Power Supply:</b> 1300W (input voltage 181VAC-240AC) –90% efficient (80PLUS Gold Certified) Externally accessible/removable	



# Mellanox Technologies - ConnectX4

- ❑ 100Gb/s Ethernet per port
- ❑ 1/10/20/25/40/50/56/100Gb/s speeds
- ❑ Single and dual-port QSFP28 options available



# ConnectX4 – Feature Summary

## ETHERNET

- 100GbE / 56GbE / 50GbE / 40GbE / 25GbE / 10GbE / 1GbE
- IEEE 802.3bj, 802.3bm 100 Gigabit Ethernet
- 25G Ethernet Consortium 25, 50 Gigabit Ethernet
- IEEE 802.3ba 40 Gigabit Ethernet
- IEEE 802.3ae 10 Gigabit Ethernet
- IEEE 802.3az Energy Efficient Ethernet
- IEEE 802.3ap based auto-negotiation and KR startup
- Proprietary Ethernet protocols (20/40GBASE-R2, 50/56GBASE-R4)
- IEEE 802.3ad, 802.1AX Link Aggregation
- IEEE 802.1Q, 802.1P VLAN tags and priority
- IEEE 802.1Qau (QCN) – Congestion Notification
- IEEE 802.1Qaz (ETS)
- IEEE 802.1Qbb (PFC)
- IEEE 802.1Qbg
- IEEE 1588v2
- Jumbo frame support (9.6KB)

## ENHANCED FEATURES

- Hardware-based reliable transport
- Collective operations offloads
- Vector collective operations offloads
- Mellanox PeerDirect™ RDMA (aka GPUDirect®) communication acceleration
- 64/66 encoding
- Extended Reliable Connected transport (XRC)
- Dynamically Connected transport (DCT)
- Enhanced Atomic operations
- Advanced memory mapping support, allowing user mode registration and remapping of memory (UMR)

- On demand paging (ODP) – registration free RDMA memory access

## STORAGE OFFLOADS

- RAID offload - erasure coding (Reed-Solomon) offload
- T10 DIF - Signature handover operation at wire speed, for ingress and egress traffic

## OVERLAY NETWORKS

- Stateless offloads for overlay networks and tunneling protocols
- Hardware offload of encapsulation and decapsulation of NVGRE and VXLAN overlay networks

## HARDWARE-BASED I/O VIRTUALIZATION

- Single Root IOV
- Multi-function per port
- Address translation and protection
- Multiple queues per virtual machine
- Enhanced QoS for vNICs
- VMware NetQueue support

## VIRTUALIZATION

- SR-IOV: Up to 256 Virtual Functions
- SR-IOV: Up to 16 Physical Functions per port
- Virtualization hierarchies (e.g. NPAR)
  - » Virtualizing Physical Functions on a physical port
  - » SR-IOV on every Physical Function
- 1K ingress and egress QoS levels
- Guaranteed QoS for VMs

## CPU OFFLOADS

- RDMA over Converged Ethernet (RoCE)
- TCP/UDP/IP stateless offload
- LSO, LRO, checksum offload
- RSS (can be done on encapsulated packet), TSS, HDS, VLAN insertion / stripping, Receive

- flow steering
- Intelligent interrupt coalescence

## REMOTE BOOT

- Remote boot over Ethernet
- Remote boot over iSCSI
- PXE and UEFI

## PROTOCOL SUPPORT

- OpenMPI, IBM PE, OSU MPI (MVAPICH/2), Intel MPI,
- Platform MPI, UPC, Open SHMEM
- TCP/UDP, MPLS, VxLAN, NVGRE, GENEVE
- iSER, NFS RDMA, SMB Direct
- uDAPL

## MANAGEMENT AND CONTROL INTERFACES

- NC-SI, MCTP over SMBus and MCTP over PCIe - Baseboard Management Controller interface
- SDN management interface for managing the eSwitch
- I2C interface for device control and configuration
- General Purpose I/O pins
- SPI interface to Flash
- JTAG IEEE 1149.1 and IEEE 1149.6

\* This section describes hardware features and capabilities.  
Please refer to the driver release notes for feature availability



# ConnectX4 – Linux support

MLNX_OFED 3.3-1.0.4.0			
Supported OS(s)	Driver Version	Device	Firmware Version
RHEL: 6.2, 6.5, 6.6, 6.7, 6.8, 7.0, 7.1, 7.2 CentOS: 6.2, 6.5, 6.6, 6.7, 6.8, 7.0, 7.1, 7.2 SLES: 10 SP3; 11 SP2; SP3; SP4; 12 12 SP1 OEL: 6.5, 6.6, 6.7; 7.1 Ubuntu: 12.04, 14.04, 14.10, 15.04, 15.10, 16.04 Fedora: 19, 20, 21, 22, 23 Debian: 7.5 7.6, 8.0, 8.1, 8.2, 8.3 Xen: 6.5 Wind River: 6.0	3.3-1.0.4.0	Connect-IB®	10.16.1006 and above
		ConnectX®-4 Lx EN	14.16.1006 and above
		ConnectX®-4 VPI	12.16.1006 and above
		ConnectX®-4 EN	12.16.1006 and above
		ConnectX®-3 Pro VPI	2.36.5000 and above
		ConnectX®-3 Pro EN	2.36.5000 and above
		ConnectX®-3 VPI	2.36.5000 and above
		ConnectX®-3 EN	2.36.5000 and above





# ConnectX4 – Windows support

WinOF-2 Supported OS(s)	Driver Version	Device	Firmware Version
Windows Server 2012 R2 (x64 only)	1.40	ConnectX®-4 Lx EN	14.16.1006 and above
Windows Server 2012 (x64 only)		ConnectX®-4 VPI	Not Supported
		ConnectX®-4 EN	12.16.1006 and above



# TRANSCEIVER QSFP28

Part number: Q.851HG.02-MEL01

100 Gigabit SR4 QSFP28 Transceiver | 100m MM 850nm



Universal 100 Gigabit SR4 QSFP28 Transceiver - 100m MM 850nm, with Monitoring (DDM / DOM), Standard MSA Multimode Transceiver for 100Gigabit Ethernet LAN with MPO connector.

**68x**  
in stock

## €978.88\*

Compatibility<sup>2</sup>

▼

Buy

1

Quote

1

Is your Compatibility missing? [Request another one here.](#)

Additional Information	Product Description
TYPE	QSFP28
INTERFACE	Multimode
WAVELENGTH TX	850nm
WAVELENGTH RX	850nm
DISTANCE	100m
POWERBUDGET (DB)	1.90dB
PROTOCOLS	100Gigabit Ethernet LAN
BANDWIDTH FROM	25.781Gbit/s
BANDWIDTH TO	103.125Gbit/s
LASER	VCSEL
RECEIVER TYPE	PIN
CONNECTOR	MPO
POSSIBLE COMPATIBILITIES	Brocade, Mellanox, Cisco Systems, Juniper, Edge Core
SUPPORTED ORIGINAL ARTICLE NUMBERS	Juniper QSFP28-100G-SR4, Mellanox MMA1B00-C100, Cisco Systems QSFP-100G-SR4-S
WAVELENGTH TX MIN.	840nm
WAVELENGTH TX MAX.	860nm
WAVELENGTH RX MIN.	840nm
WAVELENGTH RX MAX.	860nm



# 100Gb/s QSFP28 Parallel Active Optical Cable (AOC) - 10m

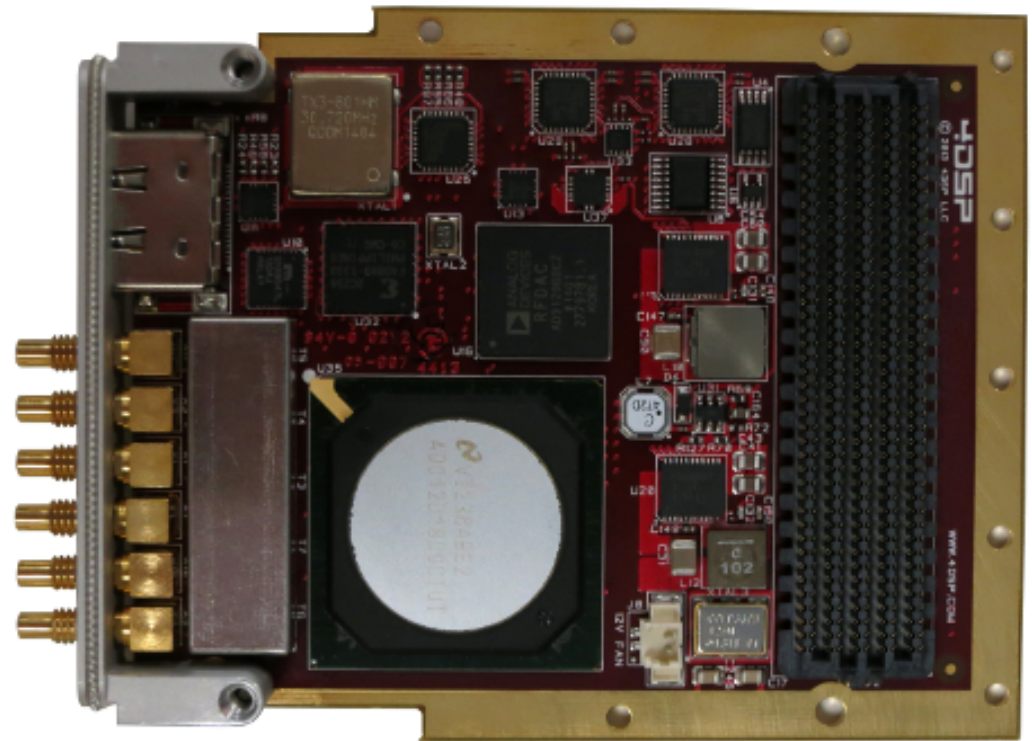
[https://hitechglobal.us/index.php?route=product/product&path=68\\_71&product\\_id=220](https://hitechglobal.us/index.php?route=product/product&path=68_71&product_id=220)

Description	Related Products (2)		<p><b>Brand:</b> InnoLight Technology Corp <b>Product Code:</b> TF-FC010-N00 <b>Availability:</b> In Stock</p>
<p><b>Features</b></p> <ul style="list-style-type: none"><li>- 4 independent full-duplex channels</li><li>- Up to 28Gb/s data rate per channel</li><li>- QSFP MSA compliant</li><li>- Up to 100m OM4 MMF transmission</li><li>- Operating case temperature: 0 to 70oC</li><li>- Single 3.3V power supply</li><li>- Maximum power consumption 3.5W each terminal</li><li>- RoHS-6 compliant</li></ul>			<p><b>Price: \$485.00</b></p> <p>Qty: <input type="text" value="1"/> <a href="#">Add to Cart</a></p>



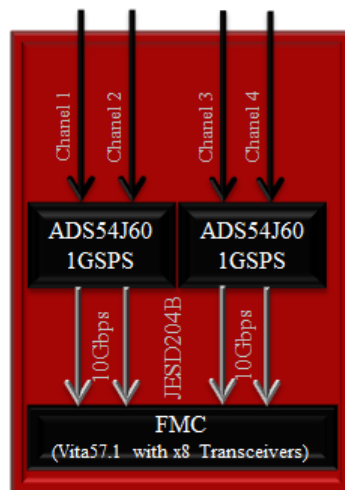
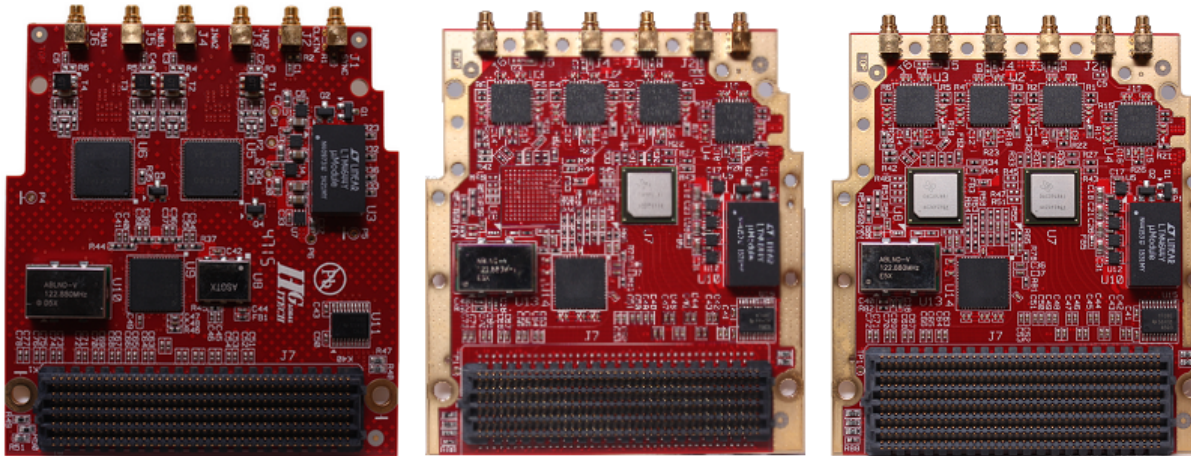
# 4DSP - FMC163

- V Single-channel 12-bit 4Gbps A/D
- Dual-channel 12-bit 2Gbps A/D
- Single-channel 14-bit 5.7Gbps D/A

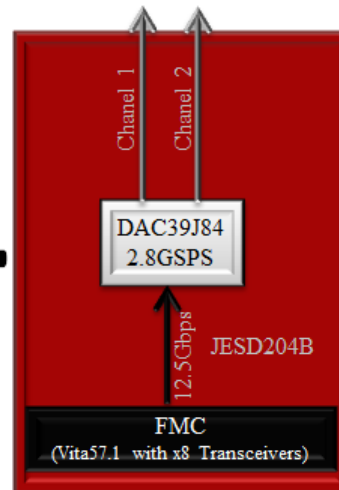


# HTG FMC ADC /DAC

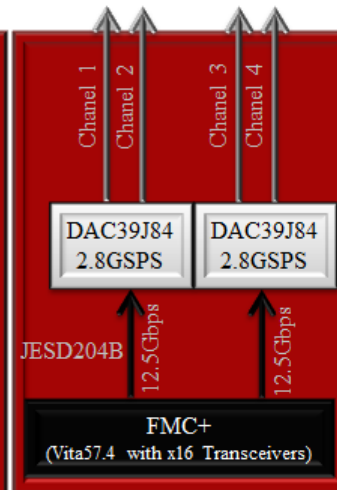
## 16-bit ADC/DAC FMC Modules



HTG-ADC16



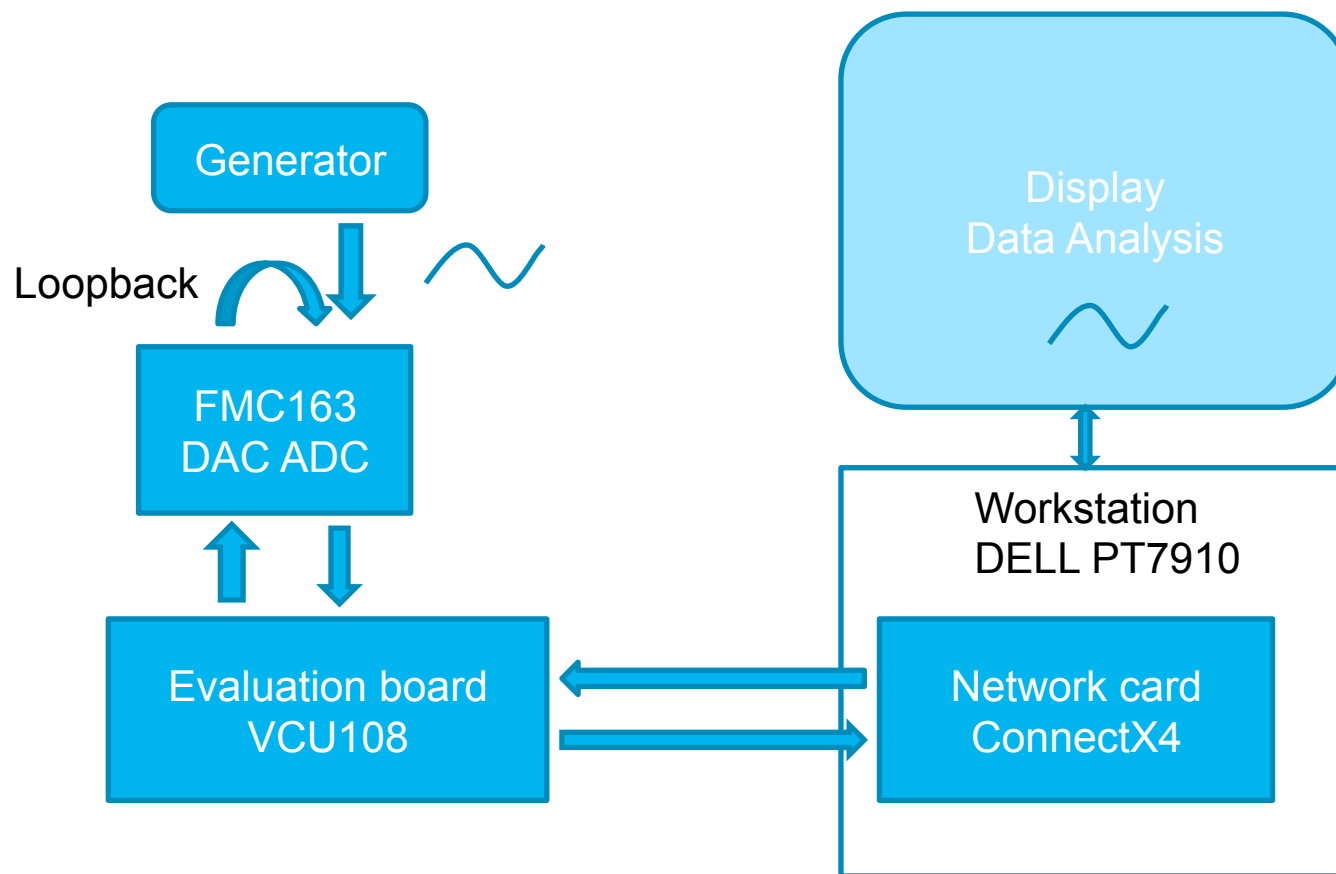
HTG-X2DAC16



HTG-X4DAC16



# Block Diagram



# IBERT – 4 x 25Gbps Loopback test

**write\_bitstream Complete**

**Unit Interval**

**BER**

**Summary**

Summary	Metrics	Settings
Name: SCAN_0	Open area: 5957	Link settings: N/A
Description: Scan 0		Horizontal increment: 1
Started: 2016-Jul-18 09:09:18		Horizontal range: -0.500 UI to 0.500 UI
Ended: 2016-Jul-18 09:09:58		Vertical increment: 1
		Vertical range: 100%

**Serial I/O Scans**

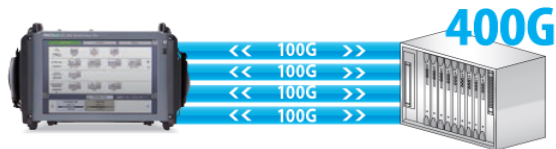
Name	Link	Link Settings	Scan Type	Status	Progress	Open Area	Horz Incr	Horz Range	Vert Incr	Vert Range	Dwell	Dwell BER	Dwell Time	Start Time	End Time
Scan 0 Found 0	2d_full_eye		2d_full_eye	Done	100%	5957	1	-0.500 UI to 0.500 UI	1	100%	BER	1e-5	0	2016-Jul-18 09:09:18	2016-Jul-18 09:09:58
Scan 1 Found 1	2d_full_eye		2d_full_eye	Done	100%	5929	1	-0.500 UI to 0.500 UI	1	100%	BER	1e-5	0	2016-Jul-18 09:26:52	2016-Jul-18 09:27:32
Scan 2 Found 2	2d_full_eye		2d_full_eye	Done	100%	1767	1	-0.500 UI to 0.500 UI	1	100%	BER	1e-5	0	2016-Jul-18 10:20:46	2016-Jul-18 10:21:26
Scan 3 Found 3	2d_full_eye		2d_full_eye	Done	100%	3111	1	-0.500 UI to 0.500 UI	1	100%	BER	1e-5	0	2016-Jul-18 10:14:17	2016-Jul-18 10:14:58

## Where to Use MT1100A

[Back to Index page](#)

### R&D on 400-Gbps Networks

- \* 4 x 100G client testing with all-in-one tester
- \* OTN multi-stage mapping and ODU-flex



### Manufacturing 100-Gbps Transport Equipment

- \* CFP, CFP2, CFP4, CXP, QSFP+, SFP+, SFP, CAUI, XLAUI interface
- \* MDIO control
- \* VOD, Pre-Emphasis, Rx Equalizer
- \* Multi-users log-in



### Network Commissioning Tests

- \* 100G core network to CPRI/OBSAI mobile fronthaul
- \* GPS-synchronized one-way latency test
- \* OTN-mapped client protocol testing
- \* Y.1564, RFC 2544, RFC 6349



### Troubleshooting

- \* Simultaneous two-way monitoring
- \* Channel statistics and Ethernet capture
- \* Long-term monitoring including remote boot, operation, file transfer and firmware update





## FTB-890 Highlights



### OTN

OTU1/OTU2/1e/2e/1f/2f  
OTU3/e1/e2 & OTU4  
OTN BERT  
OH, alarms, errors  
SDT and RTD measurements  
SONET/SDH client mapping

### Ethernet

10/100M, Gig-E, 10GE, 40GE, 100GE  
BERT, Multi-stream, RFC2544 and  
Y.1564  
Smart Loopback & Dual Test Set  
iSAM, RFC 6349, TGEN  
Advanced Filtering & Capture  
Traffic scan, MPLS, IPv4/6 up to  
100G, Ethernet OAM up to 10G

### SONET-SDH

OC1/STM0 up to OC-192/STM64  
BERT  
OH, alarms and errors  
APS, SDT and RTD measurements  
BER Threshold

### Fibre Channel

1/2/4/8/10x/16x FC BERT  
Latency measurement  
Buffer-to-Buffer credit estimation  
Alarms and Errors

### Physical Layer

Per  $\lambda$  laser control & power  
measurement  
Per lane frequency & offset  
CFP4/QSFP MDIO read/write access  
Per lane skew, alarms and errors

### CPRI/OBSAI

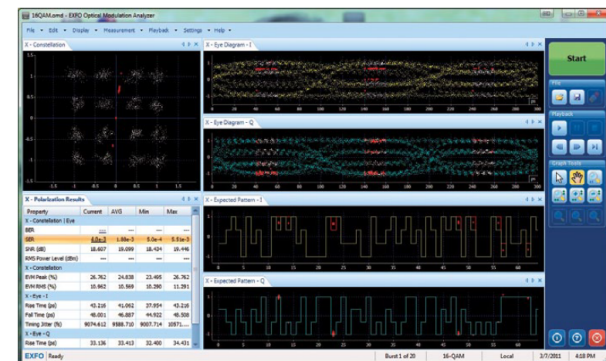
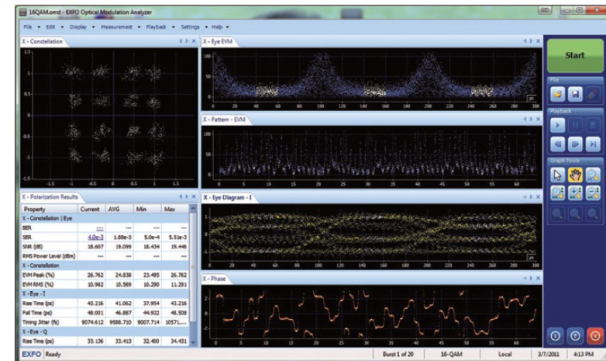
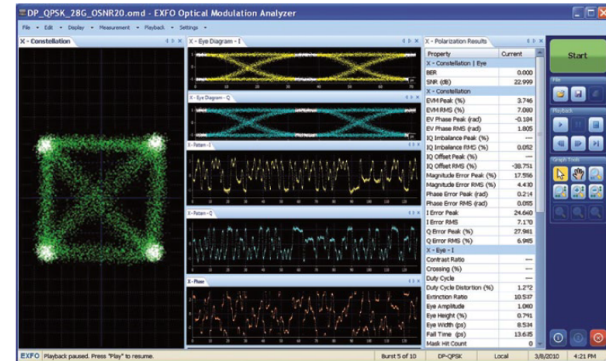
1.2/2.4/3.1/4.9/6.1/9.8G CPRI BERT  
CPRI SDT and RDT  
RRH and BBU Emulation

# PSO-200 OPTICAL MODULATION ANALYZER

Constellation, Eye Diagrams and Patterns

Error Vector Magnitude

Bit Error-Rate Analysis



# Documents exchange / sharing

- ❑ Windows group disk, intern part - N:\4all\intern\fea\UHS\_DAQ\
- ❑ FE wiki page - [https://fe-wiki.desy.de/Main\\_Page](https://fe-wiki.desy.de/Main_Page)
- ❑ DESY cloud - <https://desycloud.desy.de/index.php>
- ❑ Windows member disk, group intern part - N:\accountname\fe
- ❑ Atlassian Software for Software Development and Collaborative Documentation
  - [Atlassian Confluence](#), a wiki system with fine grained access controls and very comfortable editing functions
  - [Atlassian Jira](#) with Agile Tools, a ticket system / bug tracker for software development or project management
  - [Atlassian Stash](#), a git repository with a web interface for software development or as a document store for workgroups
  - [Atlassian Bamboo](#), a Continuous Integration System for software development

