ETHERNET
ADAPTER CARDS

PRODUCT BRIEF

ConnectX®-5 EN for Open Compute Project (OCP)

Single and dual-port 100GbE PCIe Gen4 intelligent RDMA-enabled network adapter card with advanced application offload and Multi-Host capabilities for Machine Learning, Web2.0, Cloud, and Storage platforms

ConnectX-5 EN supports up to two ports of 100Gb Ethernet connectivity, sub-600 ns latency, and very high message rate, plus PCIe Gen4 support and NVMe over Fabric offloads, providing the highest performance and most flexible solution for Open Compute Project servers and storage appliances while supporting the most demanding applications and markets: Machine Learning, Data Analytics, and more.

Machine Learning and Big Data Environments

Data analytics has become an essential function within many enterprise data centers, clouds and Hyperscale platforms. Machine learning relies on especially high throughput and low latency to train deep neural networks and to improve recognition and classification accuracy. As the first OCP card to deliver 200Gb/s throughput, ConnectX-5 dual-port 100GbE is the perfect solution to provide machine learning applications with the levels of performance and scalability that they require.

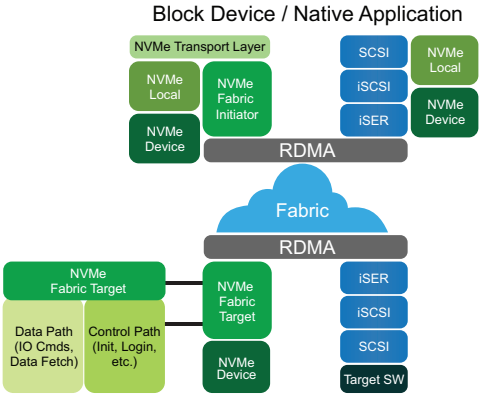
ConnectX-5 EN for Open Compute Project (OCP) utilizes RoCE (RDMA over Converged Ethernet) technology, delivering low-latency and high performance.

ConnectX-5 also supports GPUDirect® and Burst Buffer offload for background checkpointing without interfering in the main CPU operations, and the innovative transport service Dynamic Connected Transport (DCT) to ensure extreme scalability for compute and storage systems.

Storage Environments

NVMe storage devices are gaining popularity, offering very fast storage access. The evolving NVMe over Fabric (NVMf) protocol leverages the RDMA connectivity for remote access. ConnectX-5 offers further enhacements by providing NVMf target offloads, enabling very efficient NVMe storage access with no CPU intervention, and thus improved performance and lower latency.

As with the earlier generations of ConnectX adapters, standard block and file access protocols can leverage RoCE for high-performance storage access. A consolidated compute and storage network achieves significant cost-performance advantages over multi-fabric networks.



HIGHLIGHTS

NEW FEATURES

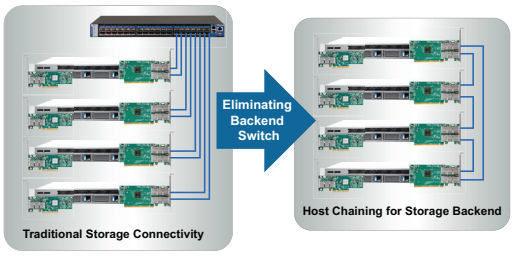
- PCIe Gen4 support, delivering 200Gb/s throughput
- Belly-to-belly 2 ports of 100GbE
- Burst Buffer Offloads for Background Checkpointing
- NVMe over Fabric (NVMf) Offloads
- Back-End Switch Elimination by Host Chaining
- Enhanced vSwitch/vRouter Offloads
- Flexible Pipeline
- RoCE for Overlay Networks

BENEFITS

- Up to 100Gb/s connectivity per port, for a total of 200Gb/s with PCIe Gen4 servers
- Open Compute Project form factor
- OCP Specification 2.0, type 2
- Industry-leading throughput, low latency, low CPU utllization and high message rate
- Maximizes data center ROI with Multi-Host technology
- Innovative rack design for storage and Machine Learning based on Host Chaining technology
- Smart interconnect for x86, Power, ARM, and GPU-based compute and storage platforms
- Advanced storage capabilities including NVMe over Fabric offloads
- Intelligent network adapter supporting flexible pipeline programmability
- Cutting-edge performance in virtualized networks including Network Function Virtualization (NFV)
- Enabler for efficient service chaining capabilities
- Efficient I/O consolidation, lowering data center costs and complexity

† For illustration only. Actual products may vary.

ConnectX-5 enables an innovative storage rack design, Host Chaining, by which different servers can interconnect directly without involving the Top of the Rack (ToR) switch. Alternatively, the Multi-Host technology that was first introduced with ConnectX-4 can be used. Mellanox Multi-Host™ technology, when enabled, allows multiple hosts to be connected into a single adapter by separating the PCIe interface into multiple and independent interfaces. With the various new rack design alternatives, ConnectX-5 lowers the total cost of ownership (TCO) in the data center by reducing CAPEX (cables, NICs, and switch port expenses), and by reducing OPEX by cutting down on switch port management and overall power usage.

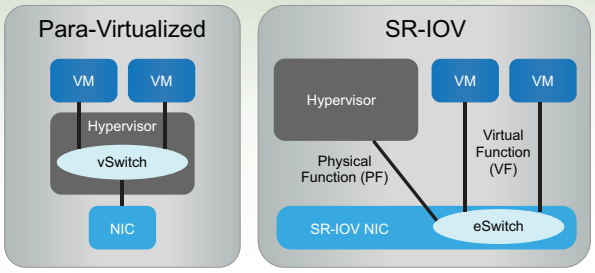


Cloud and Web2.0 Environments

Cloud and Web2.0 customers that are developing their platforms on Software Defined Network (SDN) environments, are leveraging their servers' Operating System Virtual-Switching capabilities to enable maximum flexibility.

Open V-Switch (OVS) is an example of a virtual switch that allows Virtual Machines to communicate with each other and with the outside world. Virtual switch traditionally resides in the hypervisor and switching is based on twelve-tuple matching on flows. The virtual switch or virtual router software-based solution is CPU intensive, affecting system performance and preventing fully utilizing available bandwidth.

Mellanox Accelerated Switching And Packet Processing (ASAP²) Direct technology allows to offload vSwitch/vRouter by handling the data plane in the NIC hardware while maintaining the control plane unmodified. As a result there is significantly higher vSwitch/vRouter performance without the associated CPU load.



The vSwitch/vRouter offload functions that are supported by ConnectX-5 include Overlay Networks (for example, VXLAN, NVGRE, MPLS, GENEVE, and NSH) headers' encapsulation and de-capsulation, as well as Stateless offloads of inner packets, packet headers' re-write enabling NAT functionality, and more.

Moreover, the intelligent ConnectX-5 flexible pipeline capabilities, which include flexible parser and flexible match-action tables, can be programmed, which enable hardware offloads for future protocols.

ConnectX-5 SR-IOV technology provides dedicated adapter resources and guaranteed isolation and protection for virtual machines (VMs) within the server. Moreover, with ConnectX-5 Network Function Virtualization (NFV), a VM can be used as a virtual appliance. With full data-path operations offloads as well as hairpin hardware capability and service chaining, data can be handled by the Virtual Appliance with minimum CPU utilization.

With these capabilities data center administrators benefit from better server utilization while reducing cost, power, and cable complexity, allowing more Virtual Appliances, Virtual Machines and more tenants on the same hardware.

COMPATIBILITY

PCI EXPRESS INTERFACE

- PCIe Gen 4
- PCIe Gen 3.0, 1.1 and 2.0 compatible
- 2.5, 5.0, 8, 16GT/s link rate
- Auto-negotiates to x16, x8, x4, x2, or x1 lanes
- PCIe Atomic
- TLP (Transaction Layer Packet) Processing Hints (TPH)
- Embedded PCIe Switch: Up to 8 bifurcations
- PCIe switch Downstream Port Containment (DPC) enablement for PCIe hot-plug
- Access Control Service (ACS) for peer-to-peer secure communication
- Advance Error Reportng (AER)
- Process Address Space ID (PASID) Address Transaltion Services (ATS)
- IBM CAPI v2 support (Coherent Accelerator Processor Interface)
- Support for MSI/MSI-X mechanisms

OPERATING SYSTEMS/DISTRIBUTIONS*

- RHEL/CentOS
- Windows
- FreeBSD
- VMware
- OpenFabrics Enterprise Distribution (OFED)
- OpenFabrics Windows Distribution (WinOF-2)

CONNECTIVITY

- Interoperability with Ethernet switches (up to 100GbE)
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

FEATURES SUMMARY*

ETHERNET
- 100GbE / 50GbE / 40GbE / 25GbE / 10GbE / 1GbE
- IEEE 802.3bj, 802.3bm 100 Gigabit Ethernet
- IEEE 802.3by, Ethernet Consortium 25, 50 Gigabit Ethernet, supporting all FEC modes
- 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
- 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)
- MPI Tag Matching
- Rendezvous protocol offload
- Out-of-order RDMA supporting Adaptive Routing
- Burst buffer offload
- In-Network Memory registration-free RDMA memory access

CPU OFFLOADS
- RDMA over Converged Ethernet (RoCE)
- TCP/UDP/IP stateless offload
- LSO, LRO, checksum offload
- RSS (also on encapsulated packet), TSS, HDS, VLAN and MPLS tag insertion / stripping, Receive flow steering
- Data Plane Development Kit (DPDK) for kernel bypass applications
- Open VSwitch (OVS) offload using ASAP^2
  - *Flexible match-action flow tables*
  - *Tunneling encapsulation / de-capsulation*
- Intelligent interrupt coalescence
- Header rewrite supporting hardware offload of NAT router

STORAGE OFFLOADS
- NVMe over Fabric offloads for target machine
- Erasure Coding offload - offloading Reed Solomon calculations
- T10 DIF - Signature handover operation at wire speed, for ingress and egress traffic
- Storage Protocols:
  - *SRP, iSER, NFS RDMA, SMB Direct, NVMf*

OVERLAY NETWORKS
- RoCE over Overlay Networks
- Stateless offloads for overlay network tunneling protocols
- Hardware offload of encapsulation and decapsulation of VXLAN, NVGRE, and GENEVE overlay networks

HARDWARE-BASED I/O VIRTUALIZATION
- Single Root IOV
- Address translation and protection
- VMware NetQueue support
- SR-IOV: Up to 512 Virtual Functions
- SR-IOV: Up to 16 Physical Functions per host
- Virtualization hierarchies (e.g., NPAR and Multi-Host, when enabled)
  - *Virtualizing Physical Functions on a physical port*
  - *SR-IOV on every Physical Function*
- Configurable and user-programmable QoS
- Guaranteed QoS for VMs

HPC SOFTWARE LIBRARIES
- Open MPI, IBM PE, OSU MPI (MVAPICH/2), Intel MPI
- Platform MPI, UPC, Open SHMEM

MANAGEMENT AND CONTROL
- NC-SI over MCTP over SMBus and NC-SI over MCTP over PCIe - Baseboard Management Controller interface
- SDN management interface for managing the eSwitch
- I^{2}C interface for device control and configuration
- General Purpose I/O pins
- SPI interface to Flash
- JTAG IEEE 1149.1 and IEEE 1149.6

REMOTE BOOT
- Remote boot over Ethernet
- Remote boot over iSCSI
- Unified Extensible Firmware Interface (UEFI)
- Pre-execution Environment (PXE)

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

Ordering Part Number	Description***	Dimensions w/o Bracket
MCX546A-CDAN	ConnectX®-5 Ex EN network interface card for OCP, 100GbE dual-port QSFP28, PCIe4.0 x16, no bracket, ROHS R6	OCP 2.0 type 2**
MCX546M-CDAN	ConnectX®-5 Ex EN network interface card for OCP with Multi-Host, 100GbE dual-port QSFP28, PCIe4.0 x16, no bracket, ROHS R6	OCP 2.0 type 2**
MCX546A-BCAN	ConnectX®-5 EN network interface card for OCP 40GbE dual-port QSFP28, PCIe3.0 x16, no bracket, ROHS R6	OCP 2.0 type 2**

** For more details, please refer to the Open Compute Project 2.0 Specifications.

*** All listed speeds are the maximum supported and include all lower supported speeds as well.

Mellanox TECHNOLOGIES
350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

MLNX-15-51095PB
Rev 1.1