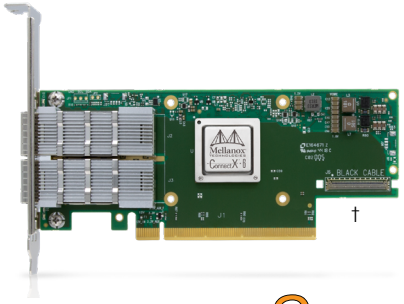
BOSTON
Servers | Storage | Solutions

Unit 5, Curo Park, Frogmore, AL2 2DD
Tel: 01727 876100
www.boston.co.uk



†

ConnectX®-6 VPI Card

200Gb/s InfiniBand & Ethernet Adapter Card

World's first 200Gb/s HDR InfiniBand and Ethernet network adapter card, offering industry-leading performance, smart offloads and In-Network Computing, leading to the highest return on investment for High-Performance Computing, Cloud, Web 2.0, Storage and Machine Learning applications

ConnectX-Virtual Protocol Interconnect® (VPI) is a groundbreaking addition to the Mellanox ConnectX series of industry-leading adapter cards. Providing two ports of 200Gb/s for InfiniBand and Ethernet connectivity, sub-600ns latency and 215 million messages per second, ConnectX-6 VPI enables the highest performance and most flexible solution aimed at meeting the continually growing demands of data center applications.

In addition to all the existing innovative features of past versions, ConnectX-6 offers a number of enhancements to further improve performance and scalability.

ConnectX-6 VPI supports HDR, HDR100, EDR, FDR, QDR, DDR and SDR InfiniBand speeds as well as 200, 100, 50, 40, 25, and 10Gb/s Ethernet speeds.

HPC Environments

Over the past decade, Mellanox has consistently driven HPC performance to new record heights. With the introduction of the ConnectX-6 adapter card, Mellanox continues to pave the way with new features and unprecedented performance for the HPC market.

ConnectX-6 VPI delivers the highest throughput and message rate in the industry. As the first adapter to deliver 200Gb/s HDR InfiniBand, 100Gb/s HDR100 InfiniBand and 200Gb/s Ethernet speeds, ConnectX-6 VPI is the perfect product to lead HPC data centers toward Exascale levels of performance and scalability.

ConnectX-6 supports the evolving Co-Design paradigm, which transforms the network into a distributed processor. With its In-Network Computing and In-Network Memory capabilities, ConnectX-6 offloads computation even further to the network, saving CPU cycles and increasing network efficiency.

ConnectX-6 VPI utilizes both IBTA RDMA (Remote Data Memory Access) and RoCE (RDMA over Converged Ethernet) technologies, delivering low-latency and high performance. ConnectX-6 enhances RDMA network capabilities even further by delivering end-to-end packet level flow control.

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 adapter card to deliver 200Gb/s throughput, ConnectX-6 is the perfect solution to provide machine learning applications with the levels of performance and scalability that they require.

ConnectX-6 utilizes the RDMA technology to deliver low-latency and high performance. ConnectX-6 enhances RDMA network capabilities even further by delivering end-to-end packet level flow control.

HIGHLIGHTS

FEATURES

- Up to 200Gb/s connectivity per port
- Max bandwidth of 200Gb/s
- Up to 215 million messages/sec
- Sub 0.6usec latency
- Block-level XTS-AES mode hardware encryption
- FIPS capable
- Advanced storage capabilities including block-level encryption and checksum offloads
- Supports both 50G SerDes (PAM4) and 25 SerDes (NRZ)-based ports
- Best-in-class packing with nsub-nanosecond accuracy
- PCIe Gen3 and PCIe Gen4 support
- RoHS-compliant
- ODCC-compatible

BENEFITS

- Industry-leading throughput, low CPU utilization and high message rate
- Highest performance and most intelligent fabric for compute and storage infrastructures
- Cutting-edge performance in virtualized networks including Network Function Virtualization (NFV)
- Mellanox Host Chaining technology for economical rack design
- Smart interconnect for x86, Power, Arm, GPU and FPGA-based compute and storage platforms
- Flexible programmable pipeline for new network flows
- Cutting-edge performance in virtualized networks, e.g., NFV
- Efficient service chaining enablement
- Increased I/O consolidation efficiencies, reducing data center costs & complexity

† For illustration only. Actual products may vary.

Security

ConnectX-6 block-level encryption offers a critical innovation to network security. As data in transit is stored or retrieved, it undergoes encryption and decryption. The ConnectX-6 hardware offloads the IEEE AES-XTS encryption/decryption from the CPU, saving latency and CPU utilization. It also guarantees protection for users sharing the same resources through the use of dedicated encryption keys.

By performing block-storage encryption in the adapter, ConnectX-6 excludes the need for self-encrypted disks. This allows customers the freedom to choose their preferred storage device, including byte-addressable and NVDIMM devices that traditionally do not provide encryption. Moreover, ConnectX-6 can support Federal Information Processing Standards (FIPS) compliance.

ConnectX-6 also includes a hardware Root-of-Trust (RoT), which uses HMAC relying on a device-unique key. This provides both a secure boot as well as cloning-protection. Delivering best-in-class device and firmware protection, ConnectX-6 also provides secured debugging capabilities, without the need for physical access.

Storage Environments

NVMe storage devices are gaining momentum, offering very fast access to storage media. The evolving NVMe over Fabric (NVMe-oF) protocol leverages RDMA connectivity to remotely access NVMe storage devices efficiently, while keeping the end-to-end NVMe model at lowest latency. With its NVMe-oF target and initiator offloads, ConnectX-6 brings further optimization to NVMe-oF, enhancing CPU utilization and scalability.

Cloud and Web2.0 Environments

Telco, Cloud and Web2.0 customers developing their platforms on Software Defined Network (SDN) environments are leveraging the Virtual Switching capabilities of the Operating Systems on their servers to enable maximum flexibility in the management and routing protocols of their networks.

Open V-Switch (OVS) is an example of a virtual switch that allows Virtual Machines to communicate among themselves and with the outside world. Software-based virtual switches, traditionally residing in the hypervisor, are CPU intensive, affecting system performance and preventing full utilization of available CPU for compute functions.

To address such performance issues, ConnectX-6 offers Mellanox ASAP^2-Accelerated Switch and Packet Processing® technology. ASAP^2 offloads the vSwitch/vRouter by handling the data plane in the NIC hardware while maintaining the control plane unmodified. As a result, significantly higher vSwitch/vRouter performance is achieved minus the associated CPU load.

The vSwitch/vRouter offload functions supported by ConnectX-5 and ConnectX-6 include encapsulation and de-capsulation of overlay network headers, as well as stateless offloads of inner packets, packet headers re-write (enabling NAT functionality), hairpin, and more.

In addition, ConnectX-6 offers intelligent flexible pipeline capabilities, including programmable flexible parser and flexible match-action tables, which enable hardware offloads for future protocols.

Standard Host Management

Mellanox host management and control capabilities include NC-SI over MCTP over SMBus, and MCTP over PCIe - Baseboard Management Controller (BMC) interface, as well as PLDM for Monitor and Control DSP0248 and PLDM for Firmware Update DSP0267.

Mellanox Socket Direct®

Mellanox Socket Direct technology improves the performance of dual-socket servers in numerous ways, such as by enabling each of their CPUs to access the network through a dedicated PCIe interface. As the connection from each CPU to the network bypasses the QPI (UPI) and the second CPU, Socket Direct reduces latency and CPU utilization. Moreover, each CPU handles only its own traffic (and not that of the second CPU), thus optimizing CPU utilization even further.

Socket Direct also enables GPUDirect® RDMA for all CPU/GPU pairs by ensuring that GPUs are linked to the CPUs closest to the adapter card. Socket Direct enables Intel® DDIO optimization on both sockets by creating a direct connection between the sockets and the adapter card.

Socket Direct technology is enabled by a main card housing the ConnectX-6 and an auxiliary PCIe card bringing in the remaining PCIe lanes. The ConnectX-6 Socket Direct card is installed into two PCIe x16 slots and connected using a 350mm long harness. The two PCIe x16 slots may also be connected to the same CPU. In this case the main advantage of the technology lies in delivering 200Gb/s to servers with PCIe Gen3-only support.

Compatibility

PCI Express Interface

- PCIe Gen 4.0, 3.0, 2.0, 1.1 compatible
- 2.5, 5.0, 8, 16GT/s link rate
- 32 lanes as 2x 16-lanes of PCIe
- Support for PCIe x1, x2, x4, x8, and x16 configurations
- PCIe Atomic
- TLP (Transaction Layer Packet) Processing Hints (TPH)
- PCIe switch Downstream Port Containment (DPC) enablement for PCIe hot-plug
- Advanced Error Reporting (AER)
- Access Control Service (ACS) for peer-to-peer secure communication
- Process Address Space ID (PASID) Address Translation Services (ATS)
- IBM CAPIv2 (Coherent Accelerator Processor Interface)
- Support for MSI/MSI-X mechanisms

Operating Systems/Distributions*

- RHEL, SLES, Ubuntu and other major Linux distributions
- Windows
- FreeBSD
- VMware
- OpenFabrics Enterprise Distribution (OFED)
- OpenFabrics Windows Distribution (WinOF-2)

Connectivity

- Interoperability with InfiniBand switches (up to HDR, as 4 lanes of 50Gb/s data rate)
- Interoperability with Ethernet switches (up to 200GbE, as 4 lanes of 50Gb/s data rate)
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

Features

InfiniBand
- HDR / HDR100 / EDR / FDR / QDR / DDR / SDR
- IBTA Specification 1.3 compliant
- RDMA, Send/Receive semantics
- Hardware-based congestion control
- Atomic operations
- 16 million I/O channels
- 256 to 4Kbyte MTU, 2Gbyte messages
- 8 virtual lanes + VL15

Ethernet
- 200GbE / 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
- 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
- Enhanced Atomic operations
- Advanced memory mapping support, allowing user mode registration and remapping of memory (UMR)
- Extended Reliable Connected transport (XRC)
- Dynamically Connected transport (DCT)
- 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 ASAP²
  - Flexible match-action flow tables
  - Tunneling encapsulation / de-capsulation
- Intelligent interrupt coalescence
- Header rewrite supporting hardware offload of NAT router

Storage Offloads
- Block-level encryption: XTS-AES 256/512 bit key
- 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, NVMe-oF

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)
  - 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
- HPC-X, OpenMPI, MVAPICH, MPICH, OpenSHMEM, PGAS and varied commercial packages

Management and Control
- NC-SI, MCTP over SMBus and MCTP over PCIe - Baseboard Management Controller interface
- PLDM for Monitor and Control DSP0248
- PLDM for Firmware Update DSP0267
- 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

Remote Boot
- Remote boot over InfiniBand
- 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 and firmware release notes for feature availability.

Table 1 - Part Numbers and Descriptions

OPN	InfiniBand Supported Speeds (Gb/s)	Ethernet Supported Speeds (GbE)	No. of Network Ports	Cage(s)	PCI Express Configuration
MCX653105A-ECAT	HDR100, EDR, FDR, QDR, DDR, SDR	100,50,40,25,10	1	QSFP56	PCIe 3.0/4.0 x16
MCX653106A-ECAT	HDR100, EDR, FDR, QDR, DDR, SDR	100,50,40,25,10	2	QSFP56	PCIe 3.0/4.0 x16
MCX653105A-HDAT	HDR, HDR100, EDR, FDR, QDR, DDR, SDR	200,100,50,40,25,10	1	QSFP56	PCIe 3.0/4.0 x16
MCX653106A-HDAT	HDR, HDR100, EDR, FDR, QDR, DDR, SDR	200,100,50,40,25,10	2	QSFP56	PCIe 3.0/4.0 x16
MCX653105A-EFAT	HDR100, EDR, FDR, QDR, DDR, SDR	100,50,40,25,10	1	QSFP56	PCIe 3.0/4.0 x16 Socket Direct 2x8 in a row
MCX653106A-EFAT	HDR100, EDR, FDR, QDR, DDR, SDR	100,50,40,25,10	2	QSFP56	PCIe 3.0/4.0 x16 Socket Direct 2x8 in a row
MCX654105A-HCAT	HDR, HDR100, EDR, FDR, QDR, DDR, SDR	200,100,50,40,25,10	1	QSFP56	PCIe3.0 x16 + PCIe3.0x16 aux. card Socket Direct
MCX654106A-HCAT	HDR, HDR100, EDR, FDR, QDR, DDR, SDR	200,100,50,40,25,10	2	QSFP56	PCIe3.0 x16 + PCIe3.0x16 aux. card Socket Direct
MCX654106A-ECAT	HDR100, EDR, FDR, QDR, DDR, SDR	100,50,40,25,10	2	QSFP56	PCIe3.0 x16 + PCIe3.0x16 aux. card Socket Direct

NOTE: Dimensions without brackets are 167.65mm x 68.90mm. All tall-bracket adapters are shipped with the tall bracket mounted and a short bracket as an accessory.

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

BOSTON Servers | Storage | Solutions
Unit 5, Curo Park, Frogmore, AL2 2DD
Tel: 01727 876100
www.boston.co.uk

53486PB
Rev 1.8