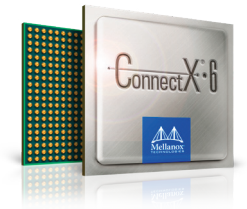
ADAPTER SILICON
PRODUCT BRIEF



ConnectX®-6 VPI IC

200Gb/s InfiniBand & Ethernet Adapter IC

ConnectX® 6 †

World's first 200Gb/s HDR InfiniBand and Ethernet network adapter offering world-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-6 Virtual Protocol Interconnect® offers unprecedented world-class performance, providing two ports of 200Gb/s for InfiniBand and Ethernet connectivity, sub-600ns latency and 200 million messages per second. With Mellanox Multi-Host™ supporting up-to 8 independent hosts, integrated PCIe switch, NVMe over Fabric and security offloads, ConnectX-6 offers the highest performance and most flexible solution for the most demanding data center applications.

ConnectX-6 is a groundbreaking addition to the Mellanox ConnectX series of industry-leading adapters. 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, 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 with which the network becomes a distributed processor. With its In-Network Computing and In-Network Memory capabilities, ConnectX-6 offloads even further computation to the network, saving CPU cycles and increasing the efficiency of the network.

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 is an essential function within many enterprise data centers, clouds and Hyperscale platforms. Many of these use Machine learning tools, which rely on especially high throughput and low latency to train deep neural networks and to improve recognition and classification accuracy. As the first adapter to deliver 200Gb/s throughput, ConnectX-6 is the perfect solution to provide machine learning applications with the performance level and scalability they require.

HIGHLIGHTS

FEATURES

- Highest performance
  - Up to 200Gb/s connectivity per port
  - Max bandwidth 200Gb/s
  - Up to 200 million messages/sec
  - Sub 0.6usec latency
- Block-level XTS-AES mode hardware encryption
- FIPS compliant adapter
- Multi-Host of up to 8 independent hosts
- Embedded PCIe switch
- PCIe Gen3 and PCIe Gen4 support

BENEFITS

- Industry-leading throughput, low latency, low CPU utilization and high message rate
- Highest performance and most intelligent fabric for compute and storage infrastructures
- Maximizes data center ROI with Multi-Host technology
- Advanced storage capabilities including block-level encryption and checksum offloads
- Host Chaining technology for economical rack design
- Smart interconnect for x86, Power, Arm, GPU and FPGA-based compute and storage platforms
- 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-6 utilizes 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.

ENHANCED 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 XTS-AES 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 encryption in the adapter, ConnectX-6 renders encryption unnecessary elsewhere in the network (e.g., storage). ConnectX-6 also supports Federal Information Processing Standards (FIPS) compliance, mitigating 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.

STORAGE ENVIRONMENTS

NVMe storage devices are gaining momentum, offering very fast access to storage media. The evolving NVMe over Fabric (NVMeOF) 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.

Storage customers will benefit from the embedded 16-lane PCIe switch, which allows them to create standalone appliances in which ConnectX-6 is directly connected to the SSDs. By leveraging ConnectX-6 PCIe Gen3/Gen4 capability, customers can build large, efficient high speed storage appliances with PCIe Gen3/Gen4 NVMe devices.

Similar to previous ConnectX generations, ConnectX-6 enables Host Chaining, an innovative storage rack design by which different servers can be connected with no need for a switch. Mellanox Multi-Host technology has been improved even further in ConnectX-6, allowing for up to eight hosts to be connected to a single adapter by segmenting the PCIe interface into multiple and independent interfaces. With a variety of new rack design alternatives, ConnectX-6 lowers the total cost of ownership (TCO) in the data center by reducing CAPEX (cables, NICs, and switch port expenses), and OPEX (cutting down on switch port management and overall power usage).

CLOUD AND WEB2.0 ENVIRONMENTS

Telco, Cloud and Web2.0 customers developing platforms on Software Defined Network (SDN) environments are leveraging the virtual switching capabilities of their servers' operating systems, to maximize the flexibility of managing and routing their network protocols.

Open VSwitch (OVS) allows virtual machines to communicate among themselves and with the outside world. Software-based virtual switches, traditionally residing in the hypervisor, are CPU-intensive; they affect system performance and prevent full CPU utilization for compute operations.

To address this, ConnectX-6 offers Mellanox Accelerated Switching And Packet Processing (ASAP^2) Direct technology to offload 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 without 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 & MULTI-HOST MANAGEMENT

Mellanox's host management technology for standard and multi-host platforms optimizes board management and power, performance and memory management via NC-SI, MCTP over SMBus and MCTP over PCIe, as well as PLDM for Monitor and Control DSP0248 and PLDM for Firmware Update DSP0267.

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)
- Embedded PCIe switch
- 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
- Advanced memory mapping support, allowing user mode registration and remapping of memory (UMR)
- Enhanced Atomic operations
- 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, NVMeOF

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)
  - Virtualizing Physical Functions on a physical port
  - SR-IOV on every Physical Function
- Configurable and user-programmable QoS
- Guaranteed QoS for VMs

Multi-Host
- Independent PCIe interfaces to independent hosts
  - Two PCIe x16 to two hosts, or four PCIe x8 to four hosts, or eight PCIe x4 to eight hosts
  - Independent NC-SI SMBus interfaces
  - Independent stand-by and wake-on-LAN signals
- Multi-Host Socket Direct – overcoming the QPI bottlenecks

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 DSP026
- 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	Description
MT28908A0-FCCF-EV	ConnectX-6 VPI, 2-port IC, HDR100/EDR/100GbE, PCIe 4.0 x32, No Crypto (ROHS2 R6)
MT28904A0-FCCF-EVM	ConnectX-6 VPI, 1-port IC, HDR100/EDR/100GbE, Multi-Host, PCIe 4.0 x32, No Crypto (ROHS2 R6)
MT28908A0-FCCF-HV	ConnectX-6 VPI, 2-port IC, HDR/200GbE, PCIe 4.0 x32 (ROHS2 R6)
MT28908A0-FCCF-HVM	ConnectX-6 VPI, 2-port IC, HDR/200GbE, Multi-Host, PCIe 4.0 x32 (ROHS2 R6)

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