†

ConnectX®-6 VPI IC

200Gb/s InfiniBand & Ethernet Adapter IC

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® is a groundbreaking addition to the Mellanox ConnectX series of industry-leading adapters. Offering unprecedented world-class performance, ConnectX-6 provides two ports of 200Gb/s for InfiniBand and Ethernet connectivity, sub-600ns latency and 215 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 VPI supports HDR, HDR100, and lower InfiniBand speeds, as well as 200, 100, 50, 40, 25, and 10 Gb/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. Delivering the highest throughput and message rate in the industry, ConnectX-6 is the first adapter to deliver 200Gb/s HDR InfiniBand, 100Gb/s HDR100 InfiniBand and 200Gb/s Ethernet speeds. Al this makes ConnectX-6 VPI 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 Direct 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 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 RoCE (RDMA over Converged Ethernet) technology to deliver low-latency and high performance. ConnectX-6 enhances RDMA network capabilities even further by delivering RoCE Congestion Control, achieving end-to-end best performance.

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
- Mellanox Multi-Host with advanced quality of service (QoS) capabilities
- Advanced storage including block-level encryption and checksum offloads
- Supports both 50G SerDes (PAM4) and 25G SerDes (NRZ) based ports
- Best-in-class packet pacing with sub-nanosecond accuracy
- PCIe Gen 3.0 and Gen 4.0 support
- RoHS compliant

BENEFITS

- Most intelligent, highest performance fabric for compute and storage infrastructures
- Cutting-edge performance in virtualized HPC networks including Network Function Virtualization (NFV)
- 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 platforms
- Flexible programmable pipeline for new network flows
- Enabler of efficient service chaining
- Efficient I/O consolidation, reducing data center costs and complexity

©2020 Mellanox Technologies. All rights reserved.

† 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 CPUutilization. It also guarantees protection for users sharing the sameresources 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 thefreedom 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.

Additionally, as in 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.

Cloud and Web 2.0 Environments

Telco, Cloud and Web 2.0 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.

With Mellanox ASAP^2 - Accelerated Switch and Packet Processing® Direct technology, significantly higher vSwitch/vRouter performance is achieved without the associated CPU load.

ConnectX-6 supports various vSwitch/vRouter offload functions including:

- Encapsulation and de-capsulation of overlay network headers
- 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 will enable hardware offloads of 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, 16 GT/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

- 50G SerDes (PAM4) and 25G SerDes (NRZ) based ports
- Interoperability with InfiniBand switches (up to HDR, as 4 lanes of 50Gb/s data rate)
- Interoperability with Ethernet switches (up to 200GbE)
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

©2020 Mellanox Technologies. All rights reserved.

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
- Mellanox 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 / decapsulation
- 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
- 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 - Mellanox ASAP²
- Single Root IOV
- Address translation and protection
- VMware NetQueue support
  - SR-IOV: Up to 1K Virtual Functions
  - SR-IOV: Up to 8 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

Mellanox 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
- Mellanox 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

InfiniBand Speeds (Gb/s)	Ethernet Speeds (GbE)	No. of Network Ports	Crypto Support	PCI Express Configuration	IC Model	OPN
HDR100 and lower	100GbE and lower	2	No crypto	PCIe Gen 3.0/4.0 x32	ConnectX-6	MT28908A0-NCCF-EV
HDR, HDR100, and lower	200GbE and lower	1	No crypto			MT28904A0-NCCF-HVM
HDR, HDR100, and lower	200GbE and lower	2	Crypto enabled			MT28908A0-CCCF-HV
HDR, HDR100, and lower	200GbE and lower	2	No crypto			MT28908A0-NCCF-HV
HDR, HDR100, and lower	200GbE and lower	2	Crypto enabled			MT28908A0-CCCF-HVM
HDR, HDR100, and lower	200GbE and lower	2	No crypto			MT28908A0-NCCF-HVM
HDR and lower	200GbE and lower	2	No crypto		ConnectX-6 Ex	MT28918A0-NCCF-HVM
HDR and lower	200GbE and lower	2	Crypto enabled			MT28918A0-CCCF-HVM

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

© Copyright 2020 Mellanox Technologies. All rights reserved.
Mellanox, Mellanox logo, ConnectX, Mellanox Multi-Host, Mellanox PeerDirect, GPUDirect and ASAP² are registered trademarks of Mellanox Technologies, Ltd.
All other trademarks are property of their respective owners.

53442PB
Rev 1.9