

ConnectX[®]-5 EN IC

100Gb/s Ethernet Adapter IC

Intelligent RDMA-enabled network adapter IC with advanced application offload capabilities—the ideal solution for today’s demanding applications and markets, e.g., Web 2.0, Cloud Data Centers, Telecoms, Storage, Machine Learning & Data Analytics

Mellanox ConnectX-5 accelerates hyperscale network performance, providing up to two ports of 100Gb Ethernet connectivity, 750ns latency, 200 million messages per second, a PCIe switch, and NVMe over Fabric (NVMe-oF) offloads. Moreover, with Mellanox Multi-Host[®] technology powered by a shared buffer architecture, and RoCE enhancements, ConnectX-5 offers a very flexible solution for demanding data center and cloud applications. ConnectX-5 Ethernet adapter ICs are available in various port types to meet the needs of diverse data centers.

Web 2.0 and Cloud Environments

Web 2.0 & Cloud service provider developers of Software Defined Network (SDN) environments leverage the virtual-switching capabilities of their servers’ operating systems to enable maximum flexibility in deploying and managing virtual servers, and rolling out new functionality.

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

By taking advantage of the following capabilities, data center administrators enjoy better server utilization, while reducing costs, power, and cable complexity, and allowing more Virtual Appliances, Virtual Machines and more tenants to run on the same hardware:

Accelerated offloading of virtual switches and virtual routers, such as Open vSwitch (OVS), powered by advanced ASAP² - Switching and Packet Processing[®], enables significantly higher data transfer performance without overloading the CPU. In addition, ConnectX-5 when combined with native RoCE (RDMA over Converged Ethernet) and DPDK (Data Plane Development Kit) support, dramatically improves Cloud and NFV platform efficiencies.

The vSwitch/vRouter offload functions supported by ConnectX-5 include Overlay Networks (e.g., VXLAN, NVGRE, MPLS, GENEVE, and NSH) headers’ encapsulation and de-encapsulation, as well as stateless offloads of inner packets, rewriting packet headers to enable NAT functionality, and more. In addition, the intelligent ConnectX-5 flexible pipeline capabilities, which include flexible parser and flexible match-action tables, are programmable, enabling hardware offloads adapted to 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.



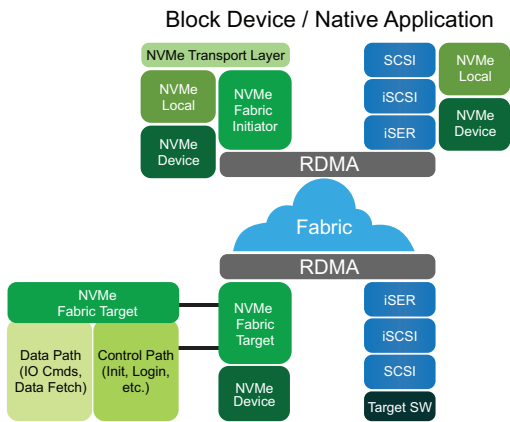
HIGHLIGHTS

FEATURES

- Tag matching and rendezvous offloads
- Adaptive routing on reliable transport
- Burst buffer offloads for background checkpointing
- NVMe over Fabric (NVMe-oF) offloads
- Back-end switch elimination by host chaining
- Embedded PCIe switch
- Enhanced vSwitch/vRouter offloads
- Flexible pipeline
- RoCE for overlay networks
- PCIe Gen 4 support
- RoHS compliant

BENEFITS

- Up to 100Gb/s connectivity per port
- CPU utilization 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



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 Mellanox Multi-Host technology that was first introduced with ConnectX-4 can be used. Mellanox's Multi-Host technology allows multiple hosts to be connected into a single adapter by separating the PCIe interface into multiple and independent interfaces.

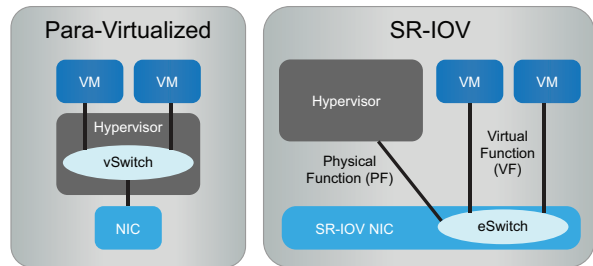
With the various new rack design alternatives, ConnectX-5 lowers the total cost of ownership (TCO) in the data center by reducing CAPEX (cables, NICs, and switch port expenses), and by reducing OPEX by cutting down on switch port management and overall power usage.

Storage Environments

NVMe storage devices are gaining popularity by offering very fast storage access. The evolving NVMe over Fabric (NVMe-oF) protocol leverages RDMA connectivity to enable remote access.

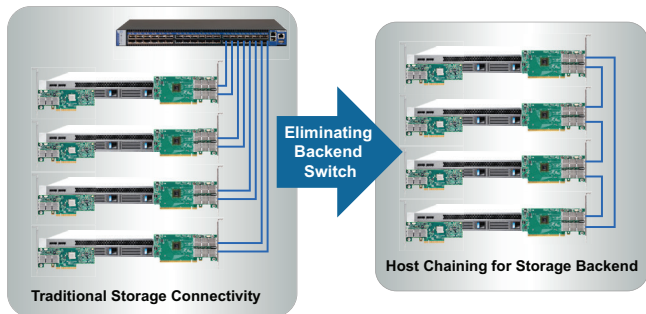
ConnectX-5 further enhancements to NVMe-of by providing NVMe-of target offloads, enabling very efficient NVMe storage access with no CPU intervention, thus improving performance and latency.

The embedded PCIe switch enables customers to build standalone storage or Machine Learning appliances. 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.



Standard & Multi-Host Management

Mellanox's host management technology for standard and multi-host platforms optimizes board management and power, performance and secure firmware update 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
- PCIe Gen 3.0, 2.0 and 1.1 compatible
- 2.5, 5.0, 8, 16 GT/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 Reporting (AER)
- Process Address Space ID (PASID) Address Translation 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*

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 (fast wake)
- 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 802.1Qaz (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)
- 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, 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

- 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 512 Virtual Functions
 - SR-IOV: Up to 8 Physical Functions per host
- Virtualization hierarchies (e.g., NPAR when enabled)
 - 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 x8 to two hosts, or four PCIe x4 to four 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

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

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 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 - Ordering Part Numbers and Descriptions

OPN	Description	Supported Speeds* (GbE)
MT27808A0-FDCF-AE	ConnectX [®] -5 EN, 2-port IC, 25GbE, PCIe 3.0 x16	1/10/25
MT27808A0-FDCF-GE	ConnectX [®] -5 EN, 2-port IC, 50GbE, PCIe 3.0 x16	1/10/25/40/50
MT27804A0-FDCF-CE	ConnectX [®] -5 EN, 1-port IC, 100GbE, PCIe 3.0 x16	1/10/25/40/50/100
MT27808A0-FDCF-CE	ConnectX [®] -5 EN, 2-port IC, 100GbE, PCIe 3.0 x16	1/10/25/40/50/100
MT27808A0-FDCF-CEM	ConnectX [®] -5 EN, 2-port IC, 100GbE, Mellanox Multi-Host, PCIe 3.0 x16	1/10/25/40/50/100
MT27808A0-FCIF-AE	ConnectX [®] -5 EN, 2-port IC, 25GbE, PCIe 3.0 x16, Industrial Temperature	1/10/25
MT27808A0-FCIF-CE	ConnectX [®] -5 EN, 2-port IC, 100GbE, PCIe 3.0 x16, Industrial Temperature	1/10/25/40/50/100

* Please refer to the driver and firmware release notes for feature availability.