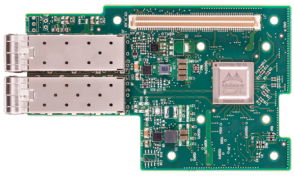
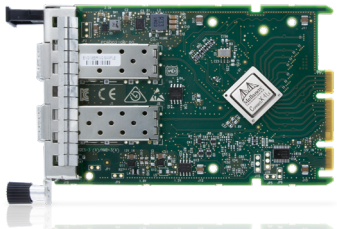
ADAPTER CARD
PRODUCT BRIEF





ConnectX®-4 Lx EN Adapter Card for Open Compute Project (OCP)

Single-port 10/25/40/50 Gigabit Ethernet Adapter Cards supporting Mellanox Multi-Host® Technology, RDMA, Overlay Networks and more

ConnectX-4 Lx EN network controller with a 10/25/40/50Gb/s Ethernet interface provides industry-leading, high bandwidth Ethernet connectivity for Open Compute Project (OCP) server and storage applications in Web 2.0, Enterprise Data Centers and Cloud infrastructures. Key capabilities of ConnectX-4 Lx EN Adapter Card for OCP include:

- Enables the achievement of significant throughput and latency improvements for server and storage applications, resulting in faster access, real-time response and more virtual machines hosted per server.
- Improves network performance by increasing available bandwidth and decreasing the associated transport load on the CPU, especially in virtualized server environments.
- Offers various form factors to meet the needs of every data center including OCP Specifications 2.0 and 3.0.
- Supports Mellanox Multi-Host® technology, which enables a new innovative rack design that achieves maximum CAPEX and OPEX savings without compromising on network performance.

Mellanox Multi-Host® Technology

ConnectX-4 Lx's innovative Multi-Host technology enables data centers to design and build scale-out heterogeneous compute and storage racks, with direct connectivity between compute elements and the network. Significantly improving cost savings, flexibility, and Total Cost of Ownership (TCO), Multi-Host technology provides better power and performance, while achieving maximum data processing and data transfer at minimum capital and operational expenses.

Mellanox's ConnectX-4 Lx's Multi-Host works by allowing multiple hosts to connect into a single interconnect adapter, by separating the adapter PCIe interface into several independent PCIe interfaces. Each interface connects to a separate host CPU—with no performance degradation. Reducing Data Center CAPEX and OPEX, Multi-Host slashes switch port management and power usage by reducing the number of cables, NICs and switch ports required by four independent servers, from four to one of each. Additional Features & Benefits of ConnectX-4 Lx's Multi-Host Technology:

- Enables IT managers to remotely control the configuration and power state of each host individually; guaranteeing host security and isolation, the management of one host does not affect host traffic performance nor the management of other hosts.
- Lowering total cost of ownership (TCO), Mellanox Multi-Host uses a single BMC, with independent NC-SI/MCTP management channels for each of the managed hosts.
- Multi-Host also supports a heterogeneous data center architecture; the various hosts connected to the single adapter can be x86, Power, GPU, Arm or FPGA, thereby removing any limitations in passing data or communicating between compute elements.

HIGHLIGHTS

FEATURES

- Multi-Host technology
- Connects up to 4 independent hosts
- OCP Specification 2.0 and 3.0, as applicable
- 10/25/40/50Gb/s speeds
- Virtualization
- Low-latency RDMA over Converged Ethernet
- Hardware offloads for NVGRE and VXLAN encapsulated traffic
- CPU offloading of transport operations
- Application offloading
- Mellanox PeerDirect™ communication acceleration
- End-to-end QoS and congestion control
- Hardware-based I/O virtualization
- Erasure Coding offload
- RoHS compliant
- ODCC compatible

BENEFITS

- 10/25/40/50GbE connectivity for servers and storage
- Open Compute Project form factor
- Industry-leading throughput and low latency for Web access and storage
- Maximizing data centers' return on investment (ROI) with Multi-Host technology
- Smart interconnect for x86, Power, Arm, and GPU-based compute and storage platform
- Cutting-edge performance in virtualized Overlay Networks (VXLAN and NVGRE)
- Efficient I/O consolidation, lowering data center costs and complexity
- Virtualization acceleration
- Power efficiency

† For illustration only. Actual products may vary.

I/O Virtualization

ConnectX-4 Lx EN SR-IOV technology provides dedicated adapter resources and guaranteed isolation and protection for virtual machines (VMs) within the server. I/O virtualization with ConnectX-4 Lx EN gives data center administrators better server utilization while reducing cost, power, and cable complexity, allowing more Virtual Machines and more tenants on the same hardware. ConnectX-4 Lx EN's SR-IOV capability and Multi-Host technology are mutually exclusive, and each host in a Multi-Host server can leverage an individual SR-IOV implementation.

Overlay Networks

In order to better scale their networks, data center operators often create overlay networks that carry traffic from individual virtual machines over logical tunnels in encapsulated formats such as NVGRE and VXLAN. While this solves network scalability issues, it hides the TCP packet from the hardware offloading engines, placing higher loads on the host CPU. ConnectX-4 Lx EN effectively addresses this by providing advanced NVGRE and VXLAN hardware offloading engines that encapsulate and de-capsulate the overlay protocol headers, and enable the traditional offloads to be performed on the encapsulated traffic for these and other tunneling protocols (GENEVE, MPLS, QinQ, and so on). With ConnectX-4 Lx EN, data center operators can achieve native performance in the new network architecture.

OVS Offload

ConnectX-4 Lx EN adds offloads to OpenVSwitch (OVS) functionalities, maximizing CPU efficiency while maintaining high flexibility.

OVS allows virtual machines to communicate both with each other and the outside world, and typically resides in the hypervisor. Switching is based on 12-tuple matching on-flows. The OVS software-based solution is CPU-intensive, affecting system performance and preventing full utilization of available bandwidth.

RDMA over Converged Ethernet (RoCE)

ConnectX-4 Lx EN supports RoCE specifications, delivering low-latency and high performance over Ethernet networks. Leveraging data center bridging (DCB) capabilities as well as ConnectX-4 Lx EN advanced congestion control hardware mechanisms, RoCE provides efficient low-latency RDMA services over Layer 2 and Layer 3 networks.

Mellanox PeerDirect™

PeerDirect communication provides high efficiency RDMA access by eliminating unnecessary internal data copies between components on the PCIe bus (for example, from GPU to CPU), and therefore significantly reduces application run time. ConnectX-4 Lx EN advanced acceleration technology enables higher cluster efficiency and scalability to tens of thousands of nodes.

Storage Acceleration

With ConnectX-4 Lx EN, Storage applications enjoy improved performance with higher bandwidth. Moreover, 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.

Distributed RAID

ConnectX-4 Lx EN delivers advanced Erasure Coding offloading capability, enabling distributed Redundant Array of Inexpensive Disks (RAID), a data storage technology that combines multiple disk drive components into a logical unit for the purposes of data redundancy and performance improvement.

ConnectX-4 Lx EN's Reed-Solomon capability introduces redundant block calculations, which, together with RDMA, achieves high performance and reliable storage access.

Standard & Multi-Host Management

Mellanox's host management technology for standard and multi-host platforms optimizes board management and power, performance and 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.

Software Support

All Mellanox adapter cards are supported by Windows, Linux distributions, VMware, FreeBSD, and Citrix XENServer. ConnectX-4 Lx EN supports various management interfaces and has a rich set of tools for configuration and management across operating systems.

Compatibility

PCI Express Interface
- PCIe Gen 3.0 compliant, 1.1 and 2.0 compatible
- 2.5, 5.0, or 8.0GT/s link rate x8
- Auto-negotiates to x8, x4, x2, or x1
- Support for MSI/MSI-X mechanisms

Operating Systems/Distributions*
- RHEL/CentOS
- Windows
- FreeBSD
- VMware
- OpenFabrics Enterprise Distribution (OFED)
- OpenFabrics Windows Distribution (WinOF-2)

Connectivity
- Interoperable with 10/25/40/50Gb Ethernet switches
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

FEATURES

Ethernet
- 50GbE / 40GbE / 25GbE / 10GbE
- 25G Ethernet Consortium 25, 50 Gigabit Ethernet
- 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)

Multi-Host
- Up to 4 separate PCIe interfaces to 4 independent hosts
- Two PCIe x4 to two hosts or four PCIe x4 to four hosts or four PCIe x2 to four hosts
- Independent NC-SI SMBus interfaces
- Independent stand-by and wake-on-LAN signals

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)
- Registration free RDMA memory access

Offloads
- RDMA over Converged Ethernet (RoCE)
- TCP/UDP/IP stateless offload
- LSO, LRO, checksum offload
- RSS (can be done on encapsulated packet), TSS, HDS, VLAN insertion/stripping, Receive flow steering
- Intelligent interrupt coalescence

Storage Offloads
- RAID offload - erasure coding (Reed-Solomon) offload

Overlay Networks
- Stateless offloads for overlay networks and tunneling protocols
- Hardware offload of encapsulation and decapsulation of NVGRE and VXLAN overlay networks

Hardware-Based I/O Virtualization
- Single Root IOV
- Multi-function per port
- Address translation and protection
- Multiple queues per virtual machine
- Enhanced QoS for vNICs
- VMware NetQueue support

Virtualization
- SR-IOV: Up to 256 Virtual Functions
- SR-IOV: Up to 16 Physical Functions per port
- Virtualization hierarchies (e.g. NPAR and Multi-Host)
  - Virtualizing Physical Functions on a physical port
  - SR-IOV on every Physical Function
- 1K ingress and egress QoS levels
- Guaranteed QoS for VMs

Protocol Support
- OpenMPI, IBM PE, OSU MPI (MVAPICH/2), Intel MPI
- Platform MPI, UPC, Open SHMEM
- TCP/UDP, MPLS, VxLAN, NVGRE, GENEVE
- iSER, NFS RDMA, SMB Direct
- uDAPL

Management and Control Interfaces
- 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
- PXE and UEFI

* 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	OCP Specification
MCX4421A-XCQN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, with Host Management, 10GbE dual-port SFP28, PCIe3.0 x8, no bracket,	OCP Spec 2.0
MCX4421A-ACQN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1,with Host Management, 25GbE dual-port SFP28, PCIe3.0 x8, no bracket	
MCX4411A-ACQN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, with Host Management, 25GbE single-port SFP28, PCIe3.0 x8, no bracket	
MCX4421A-ACAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, without Host Management 25GbE dual-port SFP28, PCIe3.0 x8, no bracket	
MCX4421A-ACUN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, without Host Management, 25GbE dual-port SFP28, PCIe3.0 x8, UEFI Enabled, no bracket	
MCX4411A-ACAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, without Host Management, 25GbE single-port SFP28, PCIe3.0 x8, no bracket	
MCX4411A-ACUN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, without Host Management, 25GbE single-port SFP28, PCIe3.0 x8, UEFI Enabled, no bracket	
MCX4431M-BCAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 2, with Multi-Host and Host Management, 40GbE single-port QSFP28, PCIe3.0 x8, no bracket	
MCX4431A-GCAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 2, with Host Management, 50GbE single-port QSFP28, PCIe3.0 x8, no bracket	
MCX4431A-BCAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, with Host Management, 40GbE single-port QSFP28, PCIe3.0 x8, no bracket	
MCX4431M-GCAN	ConnectX®-4 Lx EN network interface card for OCP2.0, Type 1, with Multi-Host and Host Management, 50GbE single-port QSFP28, PCIe3.0 x8, no bracket	
MCX4621A-ACAB	ConnectX®-4 Lx EN network interface card for OCP 3.0, with Host management, 25GbE Dual-port SFP28, PCIe3.0 x8, Thumbscrew bracket	OCP Spec 3.0

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

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

53973PB
Rev 1.8