DATASHEET

NVIDIA CONNECTX-5 INFINIBAND ADAPTER CARDS

Supporting up to 100Gb/s

NVIDIA® ConnectX®-5 InfiniBand adapter cards provide a high performance and flexible solution with up to two ports of 100Gb/s InfiniBand and Ethernet connectivity, low latency, and a high message rate, plus an embedded PCIe switch and NVMe over Fabrics offloads. These intelligent remote direct memory access (RDMA)-enabled adapters provide advanced application offload capabilities for high-performance computing (HPC), cloud hyperscale, and storage platforms.

ConnectX-5 adapter cards for PCIe Gen3 and Gen4 servers are available as stand-up PCIe cards and Open Compute Project (OCP) Spec 2.0 form factors. Selected models also offer NVIDIA Multi-Host™ and NVIDIA Socket Direct™ technologies.

HPC Environments

ConnectX-5 offers enhancements to HPC infrastructures by providing MPI and SHMEM/PGAS and rendezvous tag matching offload, hardware support for out-of-order RDMA write and read operations, as well as additional network atomic and PCIe atomic operations support.

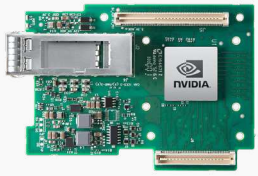
ConnectX-5 enhances RDMA network capabilities by completing the switch adaptive-routing capabilities and supporting data delivered out-of-order, while maintaining ordered completion semantics, providing multipath reliability, and efficient support for all network topologies, including DragonFly and DragonFly+.

ConnectX-5 also supports burst buffer offload for background checkpointing without interfering in the main CPU operations, and the innovative dynamic connected transport (DCT) service to ensure extreme scalability for compute and storage systems.

Storage Environments

NVMe storage devices are gaining popularity, offering very fast storage access. The NVMe over Fabrics (NVMe-oF) protocol leverages RDMA connectivity for remote access. ConnectX-5 offers further enhancements by providing NVMe-oF target offloads, enabling highly efficient NVMe storage access with no CPU intervention, and thus improved performance and lower latency.

Standard block and file access protocols can leverage RDMA for high-performance storage access. A consolidated compute and storage network achieves significant cost-performance advantages over multi-fabric networks.



KEY FEATURES

- Tag matching and rendezvous offloads
- Adaptive routing on reliable transport
- Burst buffer offloads for background checkpointing
- Embedded PCIe switch
- PCIe Gen4 support
- RoHS compliant

KEY BENEFITS

- Up to 100Gb/s connectivity per port
- Industry-leading throughput, low latency, low CPU utilization, and high message rate
- Smart interconnect for x86, Power, Arm, and GPU-based compute and storage platforms
- Advanced storage capabilities, including NVMe-oF offloads
- Efficient I/O consolidation, lowering data center costs and complexity

Images are for illustration only. Actual products may vary

Adapter Card Portfolio

ConnectX-5 InfiniBand adapter cards are available in several form factors, including low-profile stand-up PCIe, Open Compute Project (OCP) Spec 2.0 Type 1, and OCP 2.0 Type 2.

NVIDIA Multi-Host technology allows multiple hosts to be connected into a single adapter by separating the PCIe interface into multiple and independent interfaces.

The portfolio also offers NVIDIA Socket Direct configurations that enable servers without x16 PCIe slots to split the card's 16-lane PCIe bus into two 8-lane buses on dedicated cards connected by a harness. This provides 100Gb/s port speed even to servers without a x16 PCIe slot.

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

Compatibility

PCI Express Interface

- PCIe Gen 4.0, 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
- Transaction Layer Packet (TLP) Processing Hints (TPH)
- PCIe switch Downstream Port Containment (DPC)
- Access Control Service (ACS) for peer-to-peer secure communication
- Advanced 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[1]

- RHEL/CentOS
- Windows
- FreeBSD
- VMware
- OpenFabrics Enterprise Distribution (OFED)
- OpenFabrics Windows Distribution (WinOF-2)

Connectivity

- Interoperability with InfiniBand switches (up to 100Gb/s)
- Interoperability with Ethernet switches (up to 100GbE)
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

Standards[1]

- IEEE 802.3cd, 50, 100 and 200 Gigabit Ethernet
- IEEE 802.3bj, 802.3bm 100 Gigabit Ethernet
- IEEE 802.3by, Ethernet Technology 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 (supports only "Fast-Wake" mode)
- 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
- 25G/50G Ethernet Technology Consortium "Low Latency FEC" for 50/100/200GE PAM4 links
- PCI Express Gen 3.0 and 4.0

[1] This section describes hardware features and capabilities. Please refer to the driver and firmware release notes for feature availability.

[2] When using NVIDIA Socket Direct in virtualization or dual-port use cases, some restrictions may apply. For further details, contact NVIDIA Customer Support.

Features[1][2]

InfiniBand

- 100Gb/s and lower speed
- 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

- 100GbE / 50GbE / 40GbE / 25GbE / 10GbE / 1GbE
- Jumbo frame support (9.6KB)

Enhanced Features

- Hardware-based reliable transport
- Collective operations offloads
- Vector collective operations offloads
- NVIDIA 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 packets), 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 ASAP^2
  - Flexible match-action flow tables
  - Tunneling encapsulation/ decapsulation
- Intelligent interrupt coalescence
- Header rewrite supporting hardware offload of NAT router

Storage Offloads

- NVMe over Fabrics 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—NVIDIA ASAP^2

- Single root IO virtualization (SR-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

HPC Software Libraries

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

Management and Control

- NC-SI over MCTP over SMBus and NC-SI over 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
- I^{2}C 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)

[1] This section describes hardware features and capabilities. Please refer to the driver and firmware release notes for feature availability.

[2] When using NVIDIA Socket Direct in virtualization or dual-port use cases, some restrictions may apply. For further details, contact NVIDIA Customer Support.

Ordering Information

PCIe HHHL Form Factor

InfiniBand Supported Speeds	Ethenet Supported Speeds	Interface Type	Host Interface [PCIe]	Additional Features	Ordering Part Number
100Gb/s and lower	100GbE and lower	1x QSFP28	Gen3 x16	-	MCX555A-ECAT
		2x QSFP28	Gen3 x16	-	MCX556A-ECAT
		2x QSFP28	Gen3 x16	UEFI enabled	MCX556A-ECUT
		2x QSFP28	Gen4 x16	ConnectX-5 Ex	MCX556A-EDAT
100Gb/s and lower	100GbE and lower	2x QSFP28	2x Gen3 x8	Socket Direct for a dual-socket server. Dimensions without bracket: 16.7cm x 6.9cm (low profile), 11.3cm x 4.0cm, and 25cm harness	MCX556M-ECAT-S25
		2x QSFP28		Socket Direct for a dual-socket server, with active auxiliary PCIe connection card. Dimensions without bracket: 16.7cm x 6.9cm (low profile), 11.3cm x 4.0cm, and 35cm harness	MCX556M-ECAT-S35A

All tall-bracket adapters are shipped with the tall bracket mounted and a short bracket as an accessory.
Dimensions without bracket: 14.2cm x 6.9cm (low profile).
Contact NVIDIA regarding the support of lower speeds.
ConnectX-5 Ex is an enhanced performance version that supports PCIe Gen4 and higher throughput and lower latency.

OCP 2.0 Type 1 Form Factor

InfiniBand Supported Speeds	Ethenet Supported Speeds	Interface Type	Host Interface [PCIe]	Additional Features	Ordering Part Number
100Gb/s and lower	100GbE and lower	1x QSFP28	Gen3 x16	-	MCX545B-ECAN

Please refer to the Open Compute Project 2.0 Specifications.
OCP2.0 cards are shipped without a bracket.

OCP 2.0 Type 2 Form Factor

InfiniBand Supported Speeds	Ethenet Supported Speeds	Interface Type	Host Interface [PCIe]	Additional Features	Ordering Part Number
100Gb/s and lower	100GbE and lower	1x QSFP28	Gen3 x16	-	MCX545A-ECAN
		1x QSFP28	Gen3 x16	Multi-Host support	MCX545M-ECAN
		2x QSFP28 belly-to-belly	Gen4 x16	ConnectX-5 Ex	MCX546A-EDAN

Please refer to the Open Compute Project 2.0 Specifications.
OCP2.0 cards are shipped without a bracket.
25GbE or lower Ethernet port speeds require the use of a QSA (QSFP28 to SFP28) adapter. Please check ConnectX-5 firmware release notes for support status.

Standalone IC Ordering Information

InfiniBand Supported Speeds[1]	Ethenet Supported Speeds[1]	Description	Ordering Part Number
100Gb/s and lower	100GbE and lower	ConnectX-5 VPI, 2-port IC, 100Gb/s, PCIe 3.0 x16	MT27808A0-FDCF-EV
100Gb/s	100GbE and lower	ConnectX-5 VPI, 2-port IC, 100Gb/s, PCIe 3.0 x16, Industrial Temperature	MT27808A0-FCIF-EV
100Gb/s and lower	100GbE and lower	ConnectX-5 Ex VPI, 2-port IC, 100Gb/s, PCIe 4.0 x16	MT28808A0-FDCF-EV
100Gb/s and lower	100GbE and lower	ConnectX-5 Ex VPI, 2-port IC, 100Gb/s, Multi-Host, PCIe 4.0 x16	MT28808A0-FDCF-EVM

[1] Please refer to the driver and firmware release notes for feature availability.

Learn more

Find out more about InfiniBand adapters at **nvidia.com/en-us/networking/infiniband-adapters/**

NVIDIA