

Cable Interoperability Tests

1 Introduction

This document describes the interoperability tests a cable must pass.

2 Test Setup

There are five test scenarios, three of which are HCA to switch and the two are switch to switch test setup.

2.1 Mellanox QDR HCA to Intel QDR Switch

In this test setup, a Mellanox HCA on one server is connected to an Intel switch by the DUT and then another control cable connects the Intel switch to an Intel HCA on a second server to complete the connection (please see Figure 1 below).

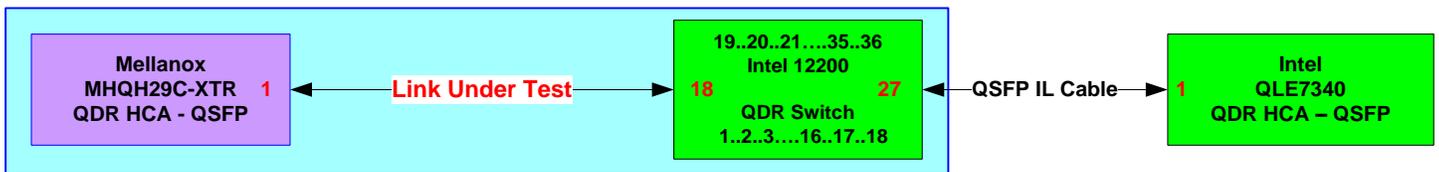


Figure 1

2.2 Intel QDR HCA to Mellanox QDR Switch

In this test setup, a Intel HCA on one server is connected to an Mellanox switch by the DUT and then another control cable connects the Mellanox switch to an HCA on a second server to complete the connection (please see Figure 2 below).

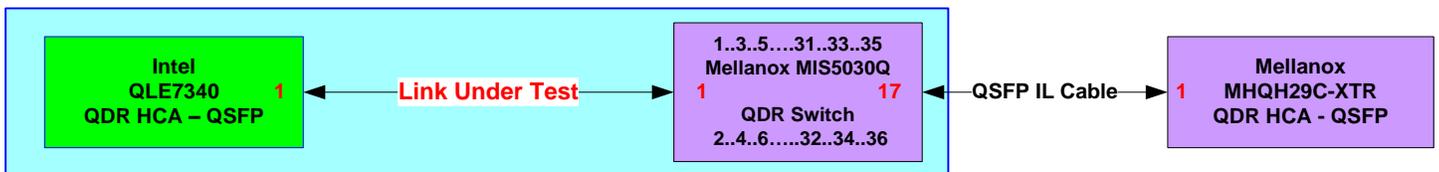


Figure 2

2.3 Mellanox FDR Switch to Intel QDR Switch

In this test scenario, a Mellanox FDR HCA is connected to a Mellanox FDR switch and an Intel QDR switch is connected to an Intel HCA on a second server using control cables. The DUT completes the connection between the Mellanox FDR switch and the Intel QDR switch (please see Figure 3 below).

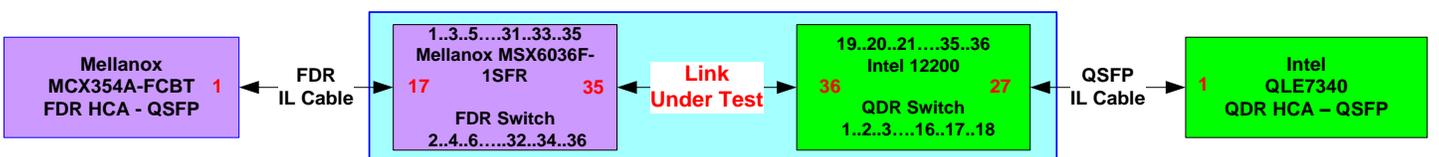


Figure 3

2.4 Mellanox FDR HCA to Intel QDR Switch

In this scenario, an Intel QDR switch is connected to an Intel HCA using a control cable. The DUT connects the Mellanox FDR HCA to the Intel QDR switch (please see Figure 4 below).

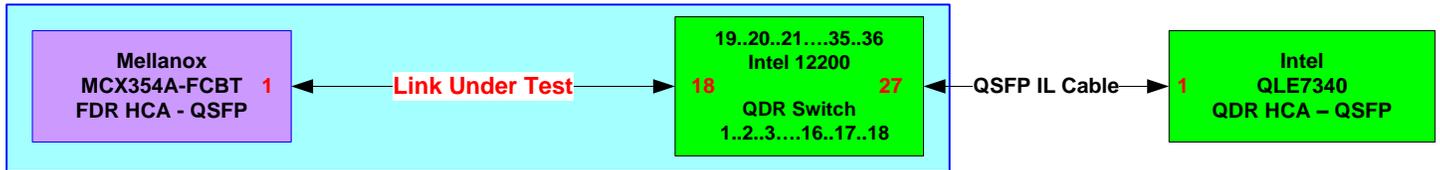


Figure 4

2.5 Mellanox FDR Switch to Mellanox FDR Switch

In this test scenario, a Mellanox FDR switch is connected to a Mellanox FDR HCA on one server and a similar FDR switch to HCA connection is made with another Mellanox FDR switch/HCA set on a second server. The DUT completes the connection between the two FDR switches (please see Figure 5 below).

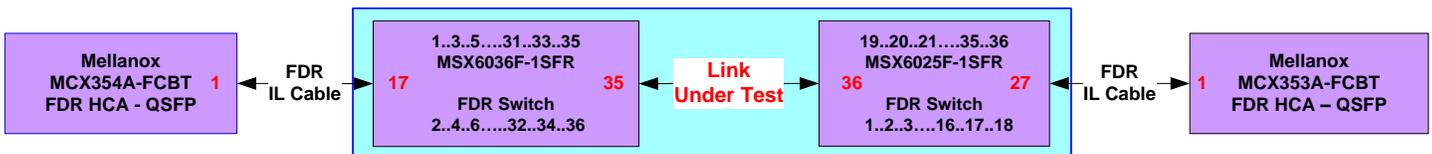


Figure 5

3 Tests

There are a number of different tests in this set up

3.1 ibdiagnet

This test scans the network using directed route packets to extract information about the fabric and its connectivity. Ibdagnet will also check for bad GUIDs/LIDs; Link State; Performance counters; Link Speed; Link Width; Partitions; and IPOIB Subnets.

Ibdagnet will run at the beginning of the MPI test run and also after MPI is completed.

3.2 MPI

Several benchmarks are completed as part of the MPI test run

3.2.1 PingPong

This repeatedly sends data packets back and forth between the paired nodes and records the latency, throughput measurements and statistics from the test.

3.2.2 PingPing

In this benchmark test, the paired nodes simultaneously send data packets, and receive the data from the other node and then wait on their send. The benchmark reports the time this entire process takes to complete.

3.2.3 Sendrecv

The send-receive operations combine in one call the sending of a message to one destination and the receiving of another message, from another process. The two (source and destination) are possibly the same.

3.2.4 Exchange

In this routine, each process exchanges data with both its left and right neighbor in the chain

3.2.5 Allreduce

This routine combines values from all processes and distributes the result back to all processes

3.2.6 Reduce

Reduces values on all processes to a single value

3.2.7 Reduce_Scatter

Combines values and then scatters the results

3.2.8 Allgather

This routine gathers data from all tasks and distributes the combined data to all tasks

3.2.9 Allgatherv

Gathers data from all tasks and delivers the combined data to all tasks

3.2.10 Gather

Gathers together values from a group of processes

3.2.11 Gatherv

Gathers values into specified locations from all processes in a group

3.2.12 Scatter

Sends data from one process to all processes in a communicator or group

3.2.13 Scatterv

Scatters a buffer in parts to all processes in a communicator

3.2.14 Alltoall

Sends data from all processes to all processes

3.2.15 Alltoallv

Sends data from all to all processes; each process may send a different amount of data and provide displacements for the input and output data

3.2.16 Bcast

Broadcasts a message from the process with rank "root" to all other processes of the communicator

3.2.17 Barrier

Blocks the caller until all processes in the communicator have called it; that is, the call returns at any process only after all members of the communicator have entered the call

3.3 Test Results

As stated in Section 3.2, an MPI test run may return a number of different errors from the benchmarks and ibdiagnet runs.

A DUT passes a given interoperability test run if the Link Width and Link Speed return correct values and there are no Link Recovery, Port, Symbol, or Port Transmit Discard errors.

Additionally, a DUT will fail if the test does not return in reasonable time and has to be stopped.

Link Width (4)	Speed (10)	Link Recovery	Port Errors	Symbol Errors	Port xmit Discard	MPI Test	Interop Pass/Fail
4	10	0	0	0	0	Pass	Yes

Figure 6

Figure 6 above shows the result of a specific test run for a QDR cable. This cable returned the correct link width, link speed and there were no link recovery, port, symbol or port transmit discard errors. If either the link width or link speed was incorrect or there were some errors on the run, the MPI Test column would have returned a **Fail** (see Figure 7).

Link Width (4)	Speed (10)	Link Recovery	Port Errors	Symbol Errors	Port xmit Discard	MPI Test	Interop Pass/Fail
4	10	0	1	5	0	Fail	No

Figure 7

In Figure 7, the cable fails the MPI test because the test returns port errors and symbol errors. If either the link width or the link speed is incorrect, the MPI test run will abort.