Need for Speed: Comparing FDR and EDR InfiniBand (Part 1)

By Olumide Olusanya and Munira Hussain

The goal of this blog is to evaluate the performance of Mellanox Technologies' FDR (Fourteen Data Rate) Infiniband and their latest EDR (Enhanced Data Rate) Infiniband with speeds of 56Gb/s and 100Gb/s respectively. This is the first of our two series blog and we will be showing how these interconnects perform on a cluster using industry-wide micro-level benchmarks and applications on HPC cluster configuration. In this part, we will show latency, bandwidth and HPL results for FDR vs EDR and in part 2 we will share more results with other applications which include ANSYS Fluent, WRF, and NAS Parallel Benchmarks. You should also keep in mind that while some applications would benefit from the higher bandwidth in EDR, other applications which have low communication overhead would show little performance improvement in comparison.

General Overview:

Mellanox EDR adapters are based on a new generation ASIC also known as ConnectX-4 while the FDR adapters are based on ConnectX-3. The theoretical uni-directional bandwidth for EDR is 100 Gb/s versus FDR which is 56Gb/s. Another difference is that EDR adapters are x16 adapters while FDR adapters are available in x8 and x16. Both of these adapters operate at a bus width of 4X link. The messaging rate for EDR can reach up to 150 million messages per second compared with FDR ConnectX-3 adapters which deliver more than 90 million messages per second.

Table 1 below shows the difference between EDR and FDR and **Table 2** describes the configuration of the cluster used in the test while **Table 3** lists the applications and benchmarks used for this test.

Table 1 - Difference between EDR and FDR

	FDR	EDR
Chipset	ConnectX-3	ConnectX-4
Link	x8 and x16 Gen3	x16 Gen3
Theoretical BW	56 Gb/s	100 Gb/s
Messaging rate	90 MMS	150 MMS
Port	QSFP	QSFP28

Table 2 - Cluster configuration

Components	Details
Server	16 nodes x PowerEdge C6320 [4 chassis]
Processor	Intel®Xeon®Intel Xeon E5-2660 v3 @2.6/2.2 GHz , 10 cores, 105W

BIOS	1.1.3
Memory	128 GB – 8 x16 GB @ 2133MHz
Operating System	Red Hat Enterprise Linux Server release 6.6.z (Santiago)
Kernel	2.6.32-504.16.2.el6.x86_64
MPI	Intel® MPI 5.0.3.048
Drivers	MLNX_OFED_LINUX-3.0-1.0.1
BIOS settings	• System Profile: Performance Optimized • Turbomode: Enabled • Cstates: Disabled • Nodeinterleave: Disabled • Hyper threading: Disabled • Snoop mode: Early/Home/COD snoop
Interconnect	**EDR** • Mellanox ConnectX-4 EDR 100Gbps • Mellanox Switch-IB SB7790 • PCI-E x16 Gen3 riser slot • HCA firmware: 12.0012.1100 • PSID: MT_2180110032 **FDR** • Mellanox ConnectX-3 FDR 56Gbps • Mellanox SwitchX SX6025 • PCI-E x8 Gen3 Mezz slot • HCA firmware: 2.30.8000 • PSID: DEL0A30000019

Table 3 - Applications and Benchmarks

Application	Domain	Version	Benchmark
OSU Micro-Benchmarks	Efficiency of MPI implementation	From Mellanox OFED 3.1	Latency, Bandwidth
HPL	Random dense linear system	From Intel MKL	Problem size 90% of total memory
Ansys Fluent	Computational Fluid Dynamics	V16.0	Eddy_417k
WRF	Weather Research and Forecasting	V3.5.1	Conus 12km

NAS Parallel Benchmarks	Computational Fluid Dynamics	3.3.1	CG, MG, IS, FT

Results

OSU Micro-Benchmarks

To find the latency and bandwidth, we used the tests from the OSU Micro-Benchmark suite. These tests use the MPI message passing performance to check the quality of a network fabric. Using the same system configuration for EDR and FDR fabrics, we got latency results as shown in **Figure 1** below.

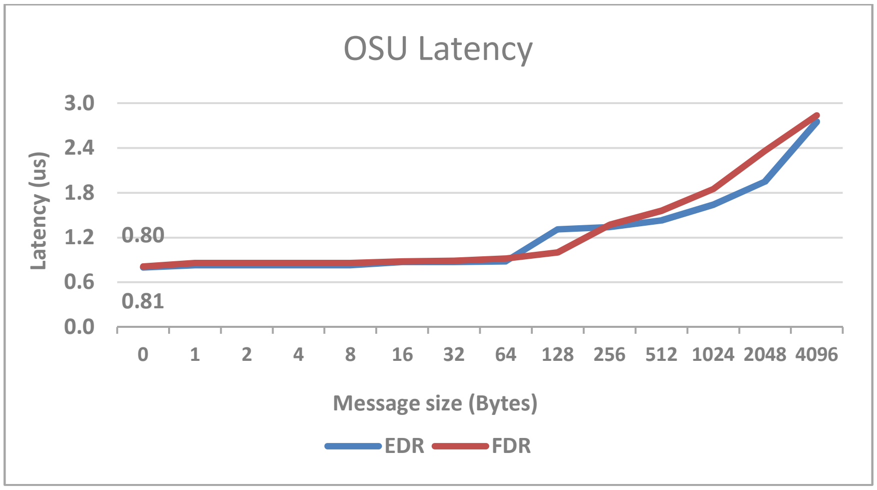

Figure 1 - OSU Latency (using MPI from Mellanox HPC-X Toolkit)

Figure 1 shows a simple OSU node-to-node latency result for EDR vs FDR. Latency numbers are typically taken from the lowest data points (usually the point with the lowest message size). Hence, the lower the data points, the better. In the above OSU latency graph, EDR shows a latency of 0.80us while FDR shows 0.81us. As the message size increases past 512 Bytes, EDR provides an even lower latency of 2.75us compared with FDR's 2.84us for a 4KB message size. When we did a further latency study using RDMA, EDR measured 0.61us and FDR measured 0.65us.

Figure 2 below plots the OSU unidirectional and bidirectional bandwidth achieved by both EDR and FDR at different message sizes from 1- 4MB.

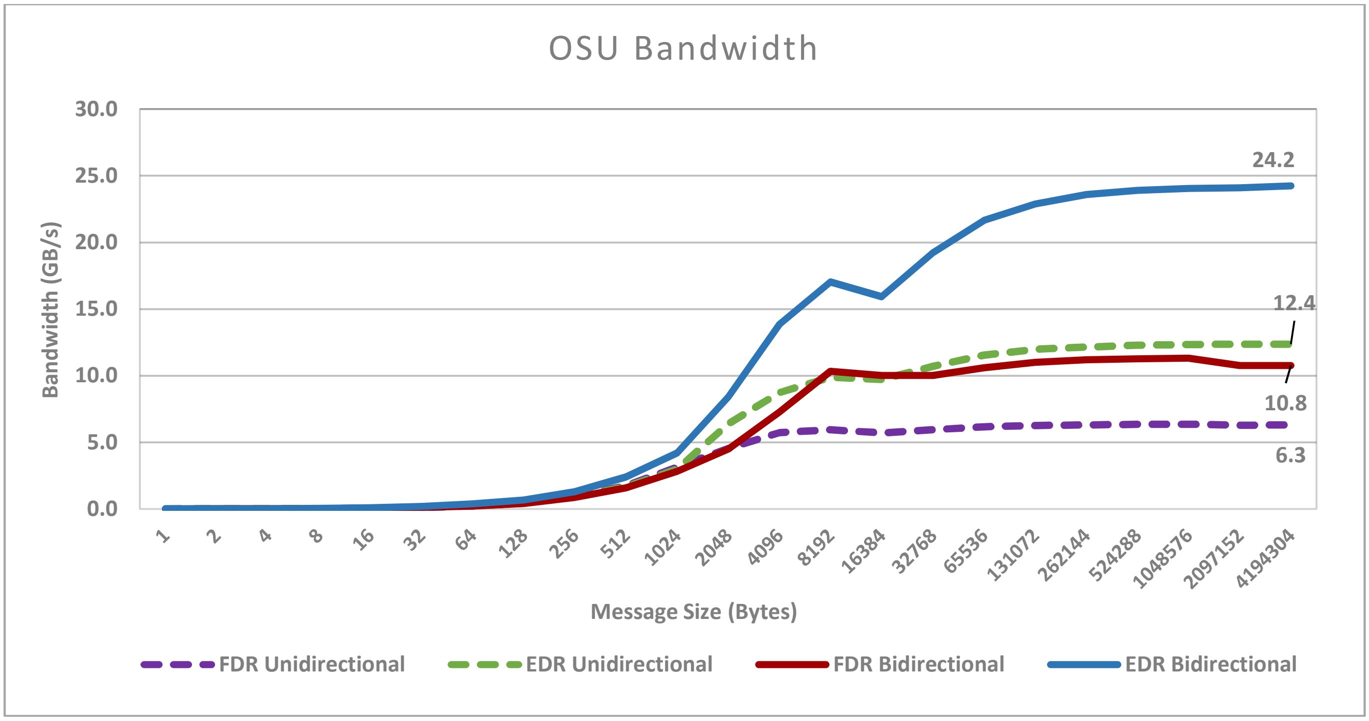

Figure 2 - OSU Bandwidth (using MPI from Mellanox HPC-X Toolkit)

OSU unidirectional bandwidth is a ping-pong type of communication test where the sender sends a fixed size of messages back-to-back to a receiver and then the receiver responds only after receiving all the messages. This test measures the maximum data rate of the network one–way or the unidirectional bandwidth. The result is taken from the achieved bandwidth of the maximum message size which is 4MB. In the above test, EDR achieves a maximum unidirectional data rate of 12.4GB/s (99.2Gb/s) and FDR achieves 6.3GB/s (50.4Gb/s). This is a 97% performance improvement in EDR over FDR.

OSU bidirectional bandwidth is very similar to the unidirectional test, but in this case, both nodes send messages to each other and await a reply. From the above graph, EDR achieves a bidirectional data rate of 24.2GB/s (193.6Gb/s) compared with FDR’s 10.8GB/s (86.4Gb/s) which gives us a 124% improvement with EDR over FDR.

HPL

Figure 3 below shows the HPL performance between EDR and FDR using COD (Cluster on Die) snoop mode. Previous studies have shown that COD gives the best performance over Home and Early snoop.

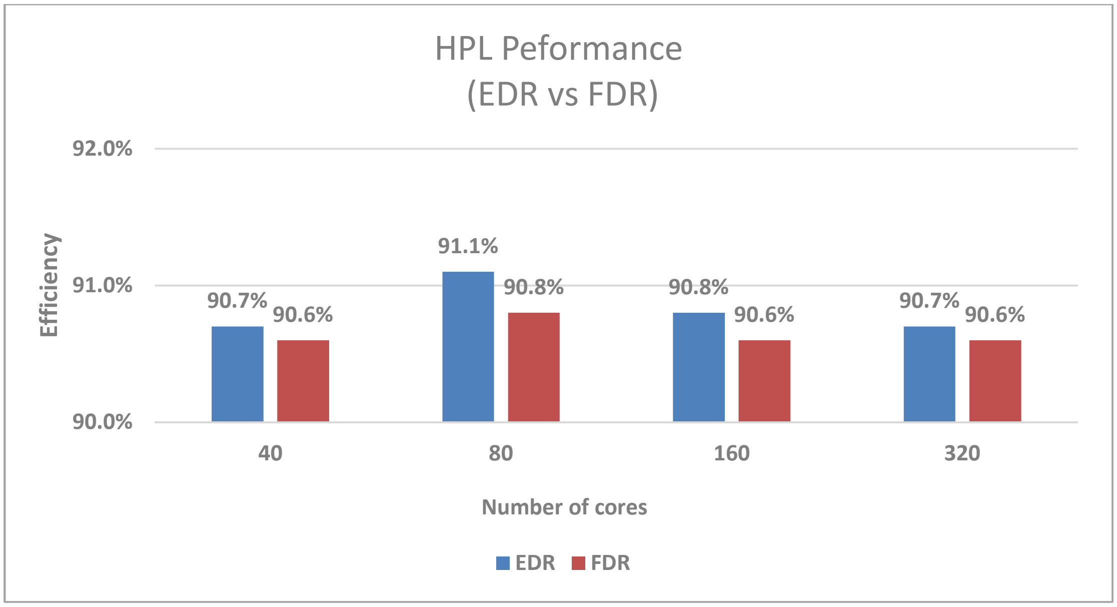

Figure 3 - HPL Performance

HPL benchmark is a compute-intensive application. It could spend more than 80% of its runtime on computation depending on how you tune it. During the bulk of its communication time, it sends messages of small sizes across the cluster which may not benefit from a higher speed network. Hence, you should not expect a huge performance difference between EDR and FDR. Even though EDR seems to perform slightly better than FDR by 0.33% in the 80-core run, this difference is within our run-run variation for successive tests with either EDR or FDR. As a result, this performance gain cannot be attributed to an EDR advantage. This also makes it is difficult to test accurately the effect of one interconnect over the other with HPL.

Conclusion

From our tests so far, EDR has shown a clear bandwidth advantage when compared with FDR – 97% in unidirectional and 124% in bidirectional bandwidth. In the second part of this blog, we will share more results from other applications (ANSYS Fluent, WRF, and NAS Parallel Benchmarks) to compare performance between EDR and FDR.