



MetaComp ICFD++ Performance Benchmarking and Profiling

Oct 2018

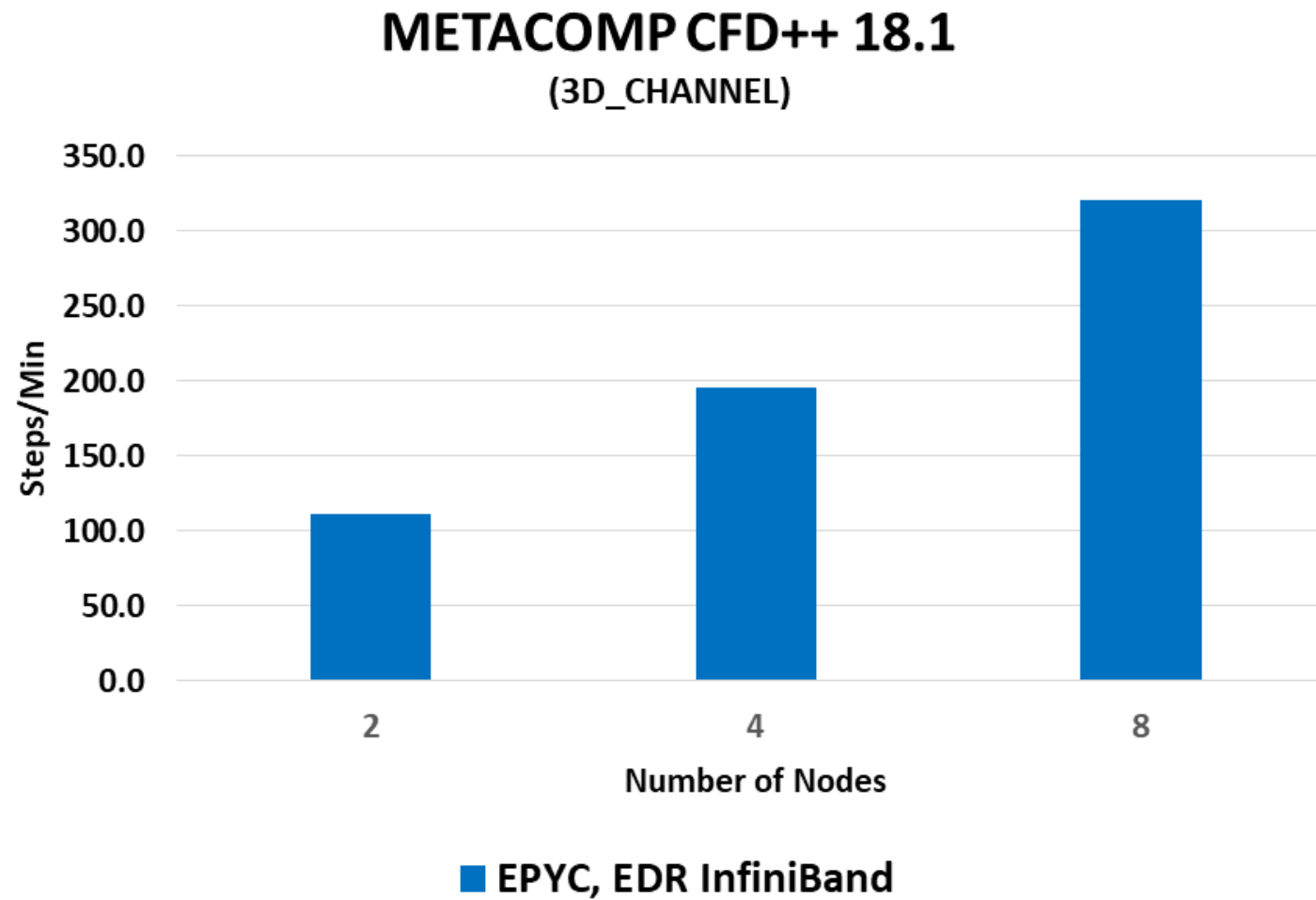
- **The following research was performed under the HPC Advisory Council activities**
 - Compute resource - HPC Advisory Council Cluster Center
- **The following was done to provide best practices**
 - MetaComp ICFD++ performance overview over AMD EPYC based platforms
 - Understanding MetaComp ICFD++ communication patterns
- **More info on MetaComp ICFD++ Application**
 - <http://www.metacomptech.com/index.php/features/icfd>

- **Computational Fluid Dynamics (CFD)**
 - Enables the study of the dynamics of things that flow
 - Enable better understanding of qualitative and quantitative physical phenomena in the flow which is used to improve engineering design
- **CFD brings together a number of different disciplines**
 - Fluid dynamics, mathematical theory of partial differential systems, computational geometry, numerical analysis, Computer science
- **MetaComp ICFD++ is a part of MetaComp's CFD software suite**
 - ICFD++ can be used to simulate compressible and incompressible fluids and flows, unsteady and steady flows, large range of speed regimes including low speeds through subsonic, transonic, supersonic and hypersonic speeds, laminar and turbulent flows, various equations of state



- **Venus cluster**
 - Supermicro AS -2023US-TR4 8-node cluster
 - Dual Socket AMD EPYC 7551 32-Core Processor @ 2.00GHz
 - Mellanox ConnectX-5 EDR 100Gb/s InfiniBand
 - Mellanox Switch-IB 2 SB7800 36-Port 100Gb/s EDR InfiniBand switch
 - Memory: 256GB DDR4 2677MHz RDIMMs per node
 - 240GB 7.2K RPM SSD 2.5" hard drive per node

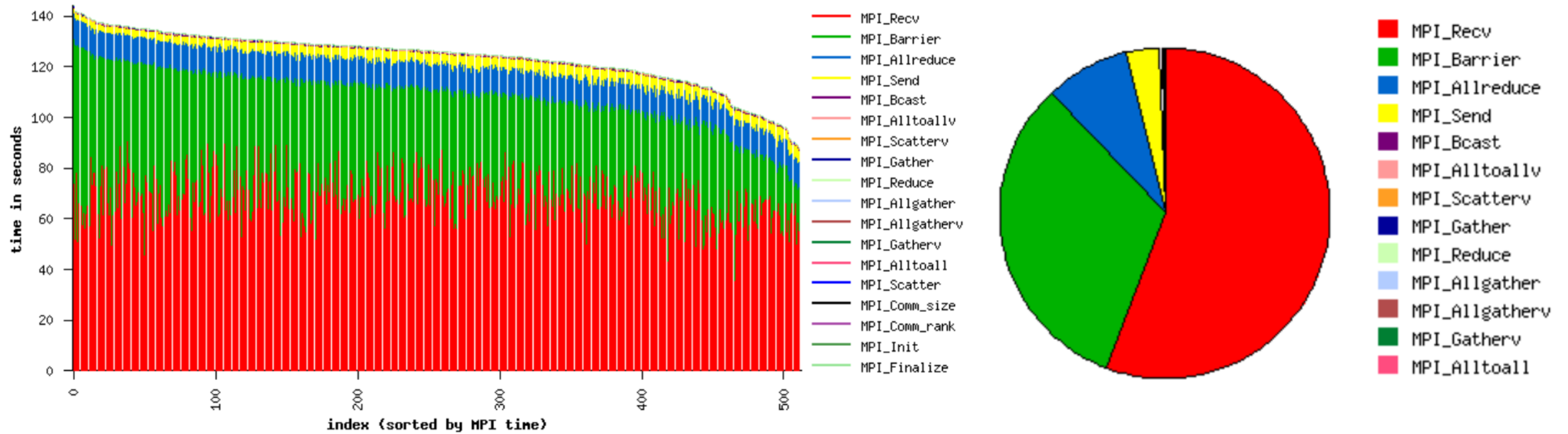
- **Software**
 - OS: RHEL 7.5, MLNX_OFED 4.4
 - MPI: HPC-X 2.2
 - MetaComp ICFD++ 18.1



Higher is better

MetaComp CDF++ Application Profile (8 nodes AMD EYPC)

- **16.14% MPI and WallClock of 765 seconds**



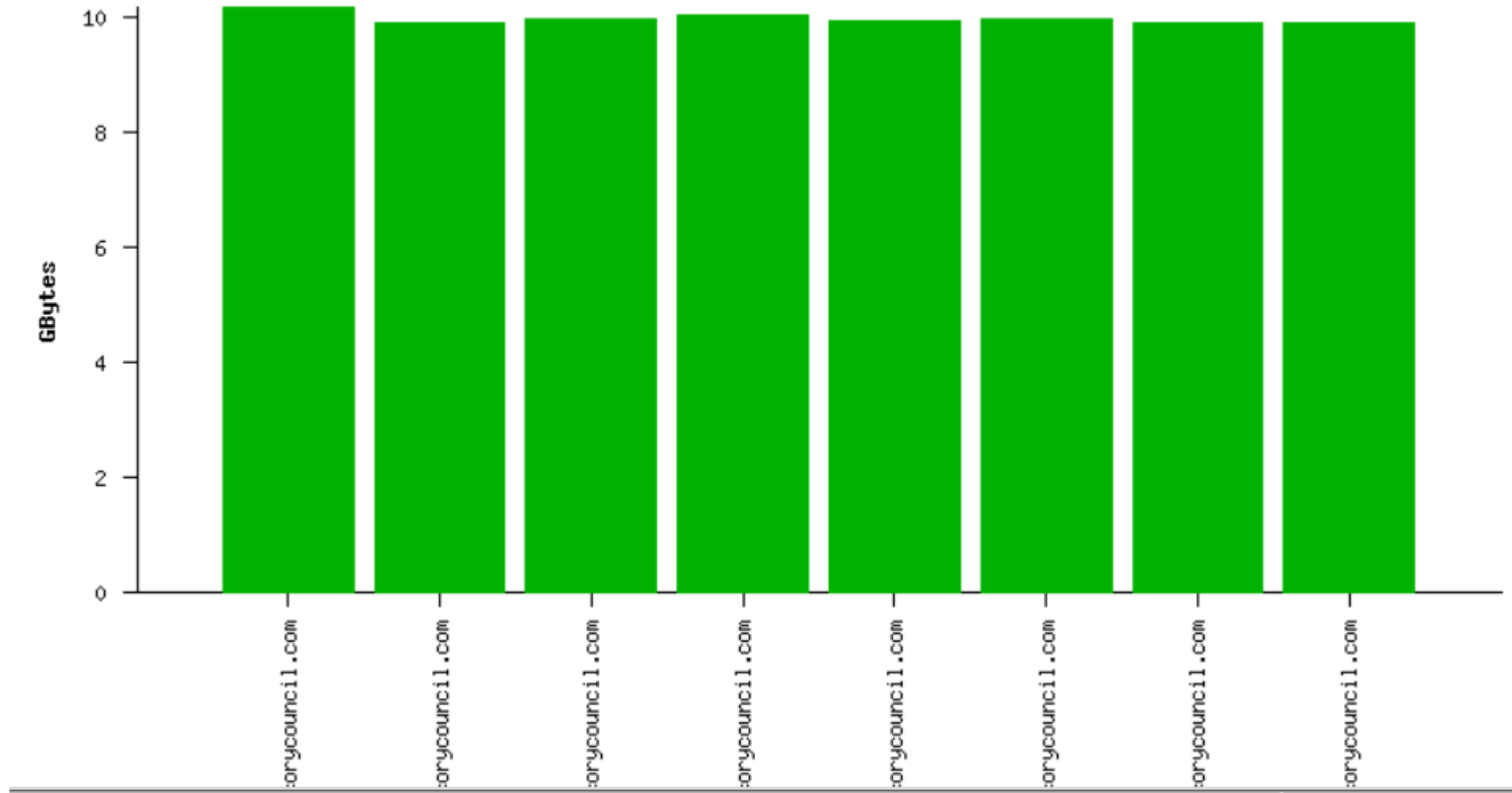
MetaComp CDF++ Application Profile (8 nodes AMD EYPC)

- Communication pattern**

Communication Event Statistics (% detail, --- error)									
	Comm Size	Buffer Size	Ncalls	Total Time	Avg Time	Min Time	Max Time	%MPI	%Wall
MPI_Barrier	512	0	24601088	2.038838e+04	8.287593e-04	3.814700e-06	3.022800e-02	32.23	5.20
MPI_Recv	0	4	275908153	1.772183e+04	6.423089e-05	0.000000e+00	4.838200e+00	28.01	4.52
MPI_Recv	0	10240	11949424	5.059036e+03	4.233707e-04	0.000000e+00	6.215200e+00	8.00	1.29
MPI_Allreduce	512	4	18444800	1.630818e+03	8.841613e-05	4.053100e-06	1.618500e-02	2.58	0.42
MPI_Recv	0	32768	22384000	1.161753e+03	5.190105e-05	1.907300e-06	1.501700e-02	1.84	0.30
MPI_Recv	0	24576	18656000	9.818093e+02	5.262700e-05	9.536700e-07	1.501500e-02	1.55	0.25
MPI_Recv	0	16384	20816000	9.191026e+02	4.415366e-05	0.000000e+00	1.475600e-02	1.45	0.23
MPI_Allreduce	512	768	8192000	9.161137e+02	1.118303e-04	1.382800e-05	1.460800e-02	1.45	0.23
MPI_Recv	0	40960	15232000	8.732476e+02	5.732981e-05	2.861000e-06	1.526500e-02	1.38	0.22
MPI_Allreduce	512	8	12292608	8.728588e+02	7.100681e-05	3.814700e-06	1.426100e-02	1.38	0.22
MPI_Recv	0	20480	18224000	8.516268e+02	4.673106e-05	9.536700e-07	1.518600e-02	1.35	0.22
MPI_Recv	0	28672	13552000	6.714251e+02	4.954436e-05	9.536700e-07	1.502000e-02	1.06	0.17
MPI_Recv	0	8192	13520144	6.513826e+02	4.817867e-05	0.000000e+00	6.144100e-01	1.03	0.17
MPI_Allreduce	512	24	8194048	5.601927e+02	6.836581e-05	5.960500e-06	1.443600e-02	0.89	0.14
MPI_Recv	0	4096	63072189	5.588526e+02	8.860523e-06	0.000000e+00	3.298100e-01	0.88	0.14
MPI_Allreduce	512	64	8192000	5.537490e+02	6.759632e-05	5.960500e-06	7.785100e-03	0.88	0.14
MPI_Allreduce	512	2560	2048000	5.508098e+02	2.689501e-04	4.196200e-05	7.550000e-03	0.87	0.14
MPI_Recv	0	49152	7040000	5.062815e+02	7.191499e-05	4.053100e-06	1.497900e-02	0.80	0.13
MPI_Recv	0	5120	25536107	4.951769e+02	1.939125e-05	0.000000e+00	1.575500e-01	0.78	0.13
MPI_Recv	0	12288	11392000	4.430782e+02	3.889380e-05	0.000000e+00	1.479600e-02	0.70	0.11

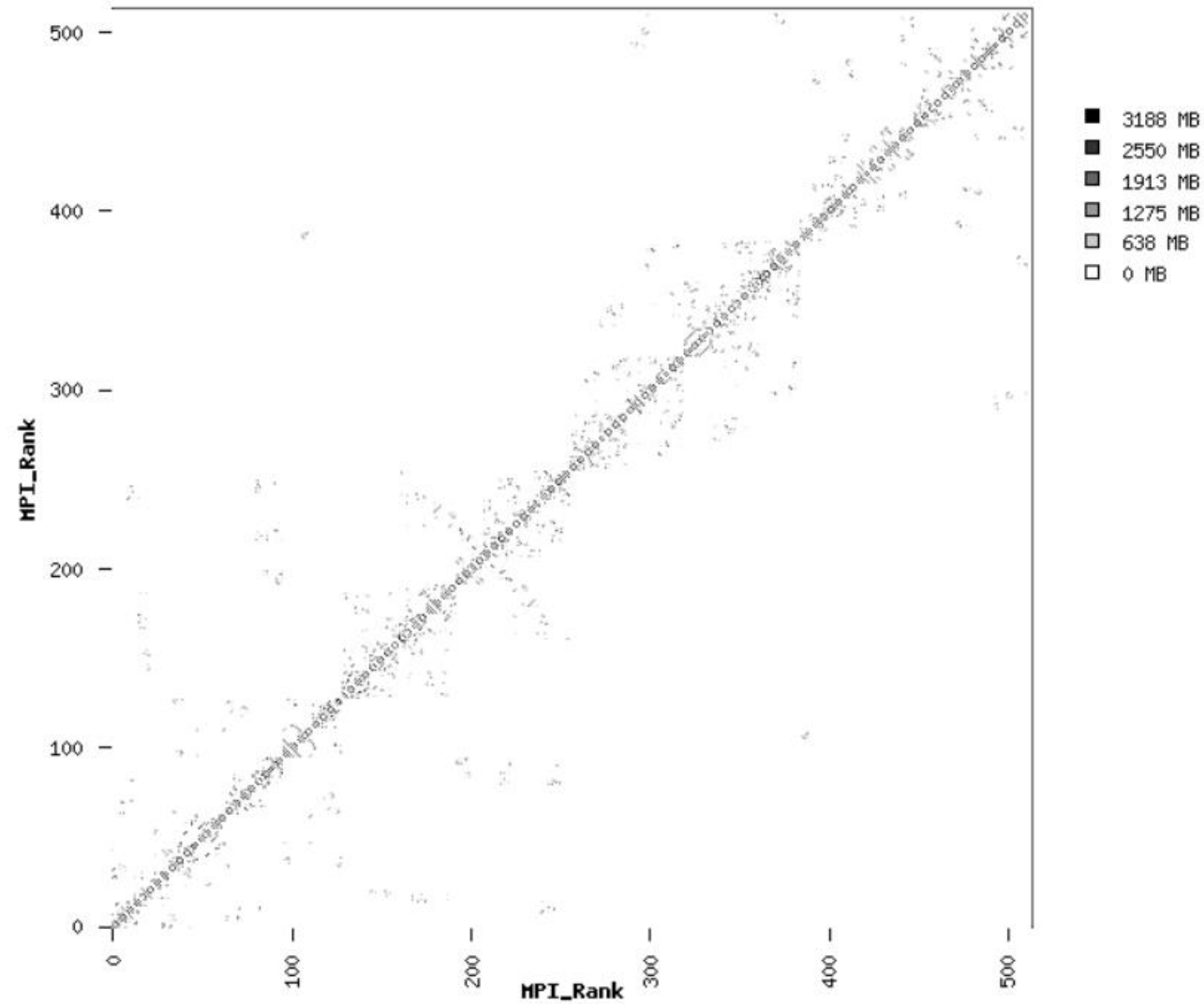
MetaComp CDF++ Application Profile (8 nodes AMD EYPC)

- **Memory usage: ~10GB per node**



MetaComp ICFD++ Application Profile (8 nodes AMD EYPC)

- **Communication is done mainly between near ranks**



- **MetaComp performance testing over AMD EYPC based platform**
 - An average of 63% scaling was achieved from 4 to 8 nodes
- **MetaComp profiling on “3D_CHANNEL”**
 - MPI communication accounts for 16.14% of overall wall clock time at 8 nodes
 - MPI_Recv is 66% of MPI, MPI_Barrier is 28% of MPI and MPI_Allreduce is 6% of MPI
 - Most communication is done between ranks that are close to each other

Thank You

