



# NEMO5

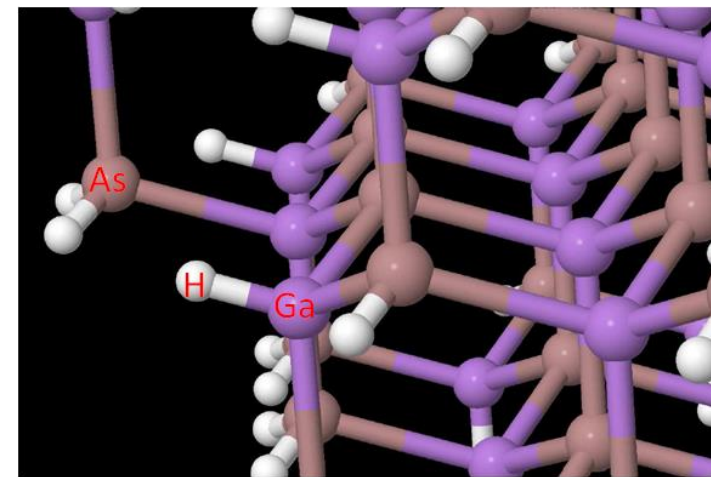
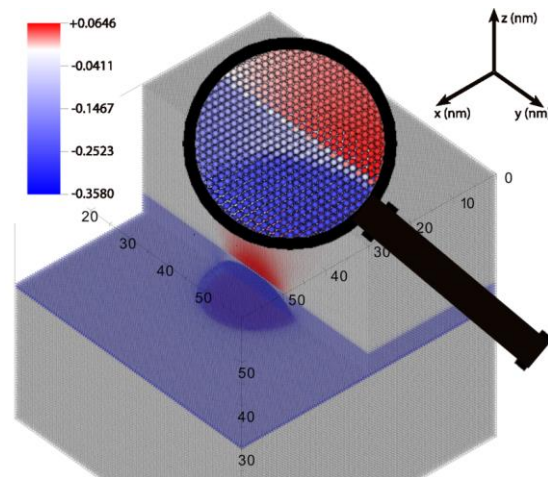
## Performance Benchmarking

July 2017



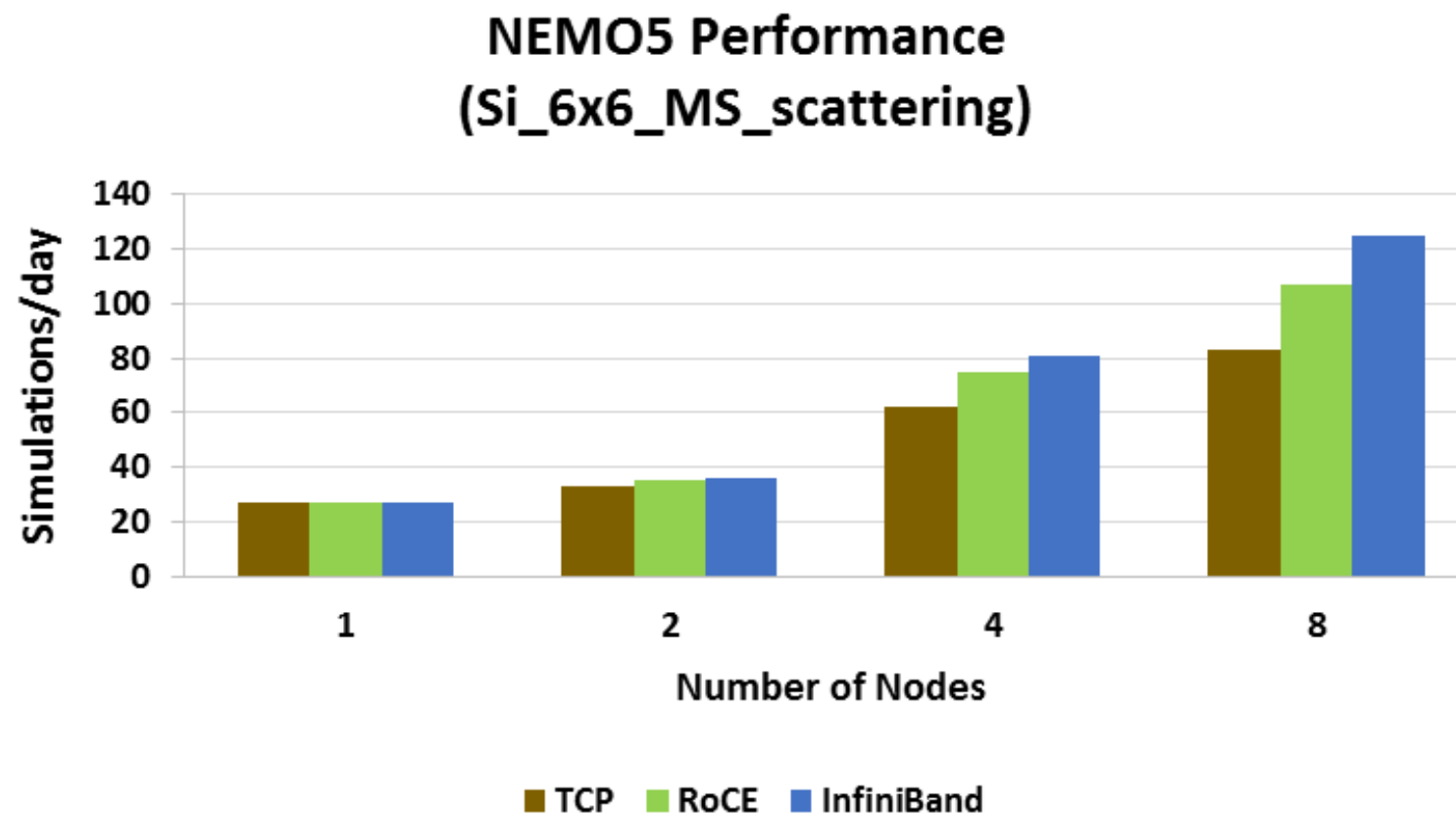
- **NEMO5**

- the 5th edition of NanoElectronics MOdeling (NEMO) Tools of the Klimeck group
- incorporates the core concepts and insights gained from 15 years of development of:
  - NEMO-1D, NEMO-3D, NEMO-3D-Peta and OMEN
- Licensing agreements for academic and commercial use are available; Free with restrictions for academic use
- Software distribution and support is handled through the NEMO5 distribution and support group on nanoHUB
- **The core capabilities lie in the atomic-resolution calculation of nanostructure properties**
  - strain relaxation, phonon modes, electronic structure using the tight-binding model, self consistent Schroedinger-Poisson calculations, and quantum transport.



- **Dell PowerEdge R720 32-node (640-core) “Jupiter” cluster**
  - Dual-Socket 10-Core Intel E5-2680v2 @ 2.80 GHz CPUs
  - Memory: 64GB memory, DDR4 1600 MHz
  - BIOS: Maximum Performance; Memory Snoop Mode: Home Snoop
  - OS: RHEL 6.5
  - MLNX\_OFED\_LINUX-4.0-2.0.0.1 InfiniBand SW stack
  - Mellanox ConnectX-4 EDR 100Gb/s InfiniBand Adapters
  - Mellanox Switch-IB SB7800 36-port EDR 100Gb/s InfiniBand Switch
  - Mellanox Spectrum SN2700 32-port 100Gb/s Ethernet Switch
  - Dell InfiniBand-Based Lustre Storage based on Dell PowerVault MD3460 and MD3420
  - Compilers: Intel Compilers 2016.4.258
  - MPI: Intel MPI 5.1.3
  - Python: 2.7.13
  - MPI Profiler: IPM

- InfiniBand provides 17% higher performance versus Ethernet RoCE
- InfiniBand provides 51% higher performance versus Ethernet TCP
- Ethernet RoCE provides 29% higher performance versus Ethernet TCP

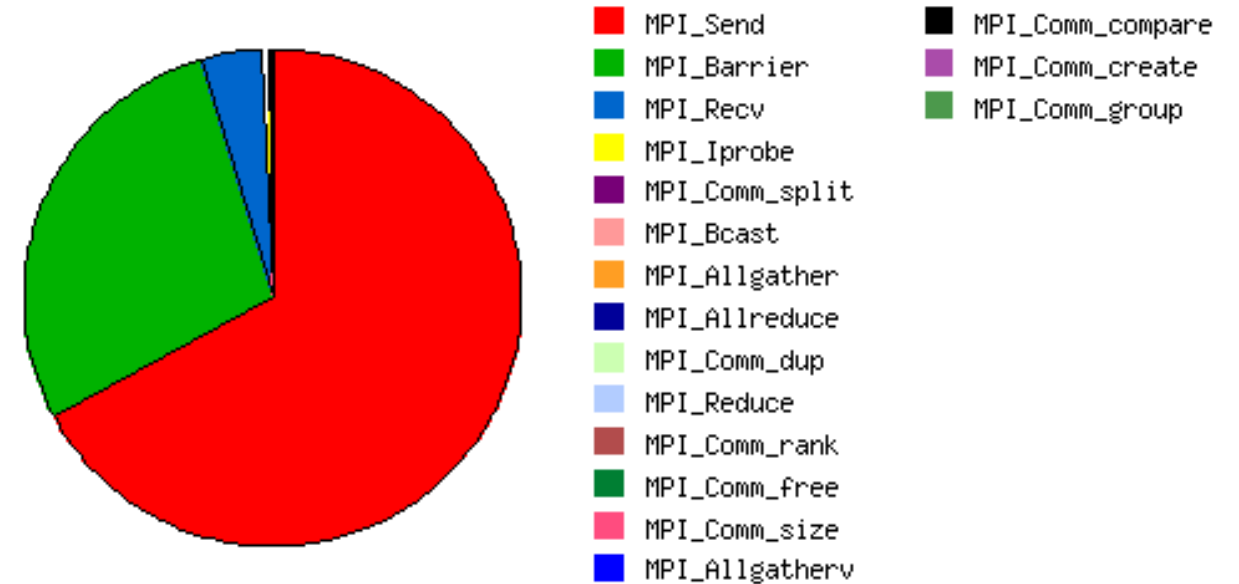
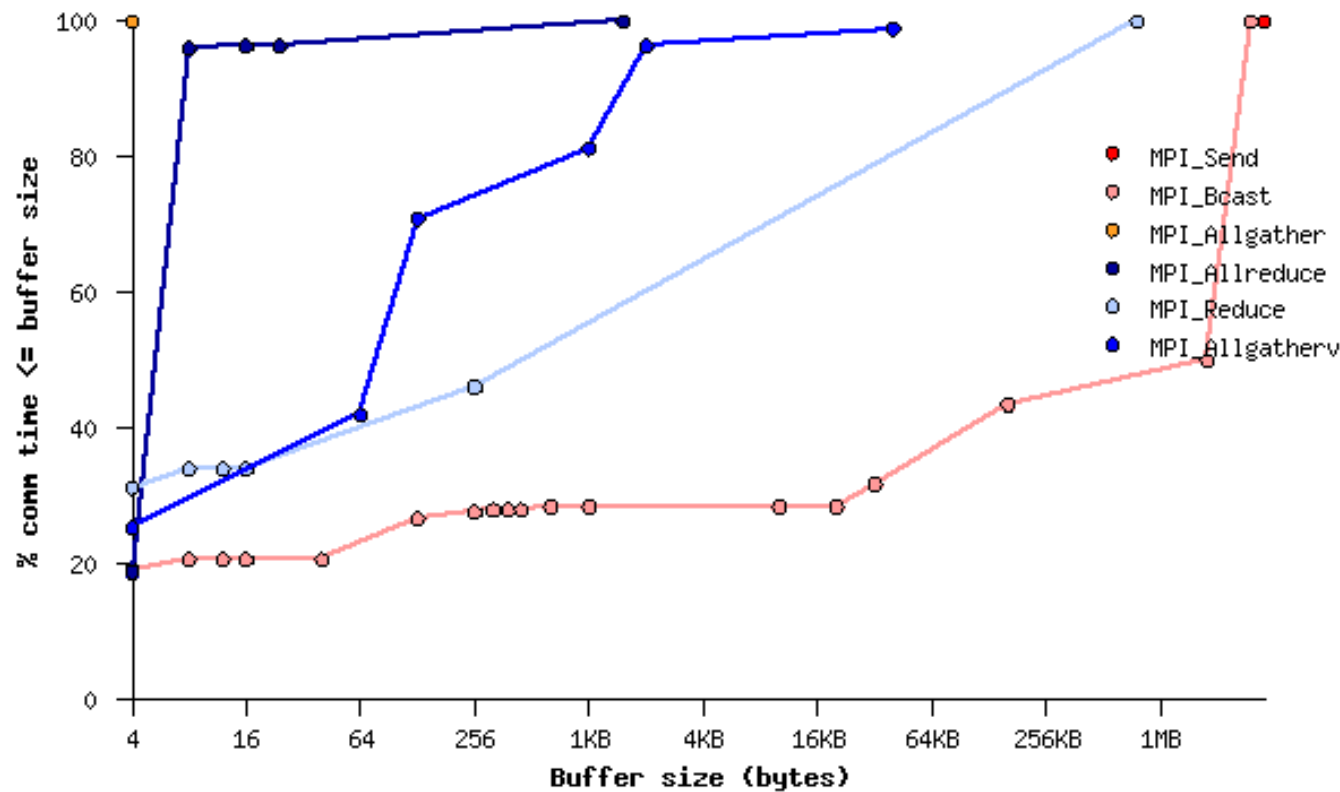


*Higher is better*

*20 Processes Per Node*

# NEMO5 Profiling – Time Spent by MPI Calls

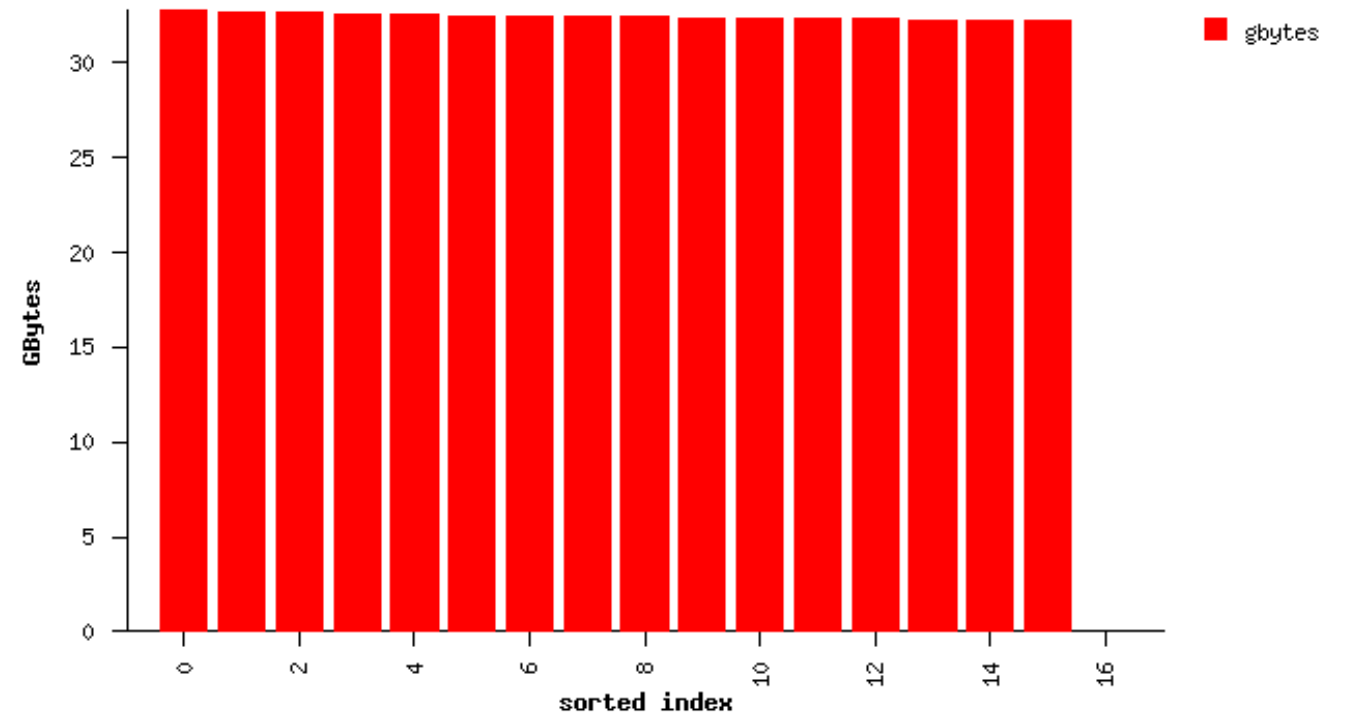
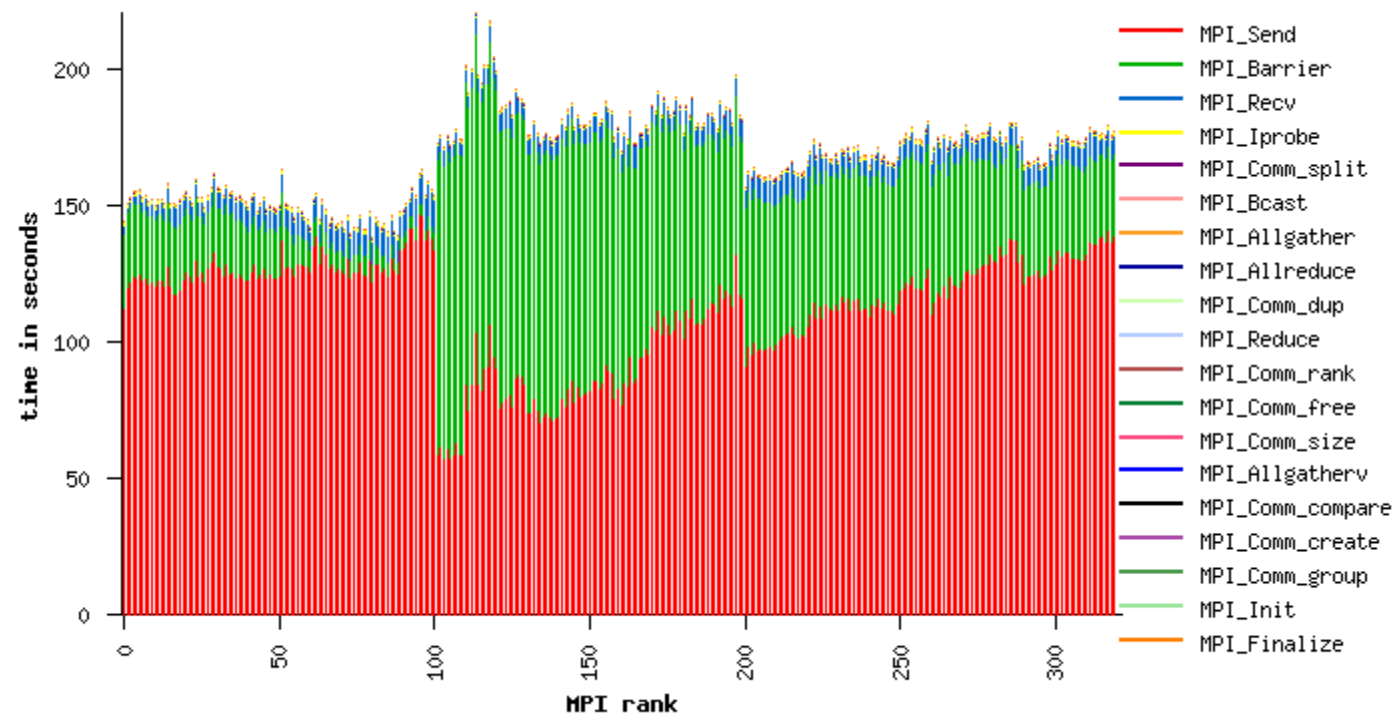
- For the most time consuming MPI calls (as % of MPI time):
  - MPI\_Send (67% of MPI time, 18% overall)
  - MPI\_Barrier (29% of MPI time, 8% overall)
  - MPI\_Recv (4% of MPI time, 1% overall)



16 Nodes / 320 Processes

# NEMO5 Profiling – Communication Balance by Task

- For the most time consuming MPI calls is MPI\_Send
  - Some load imbalance observed in the time spent by the MPI tasks

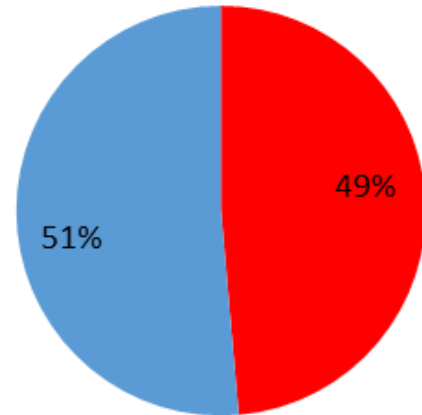


16 Nodes / 320 Processes

# NEMO5 Profiling – % Time in Communications

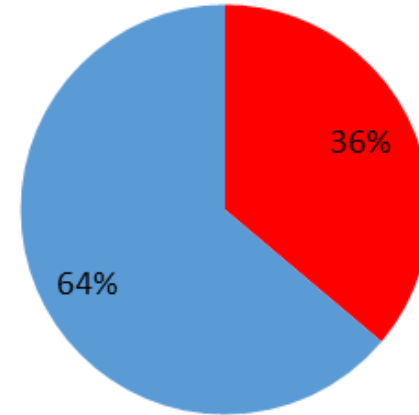
- **InfiniBand consumed the least percentage time in overall runtime**
  - InfiniBand consumed around 24% of overall runtime, while TCP consumes up to 49%

NEMO5 Profiling  
(TCP)



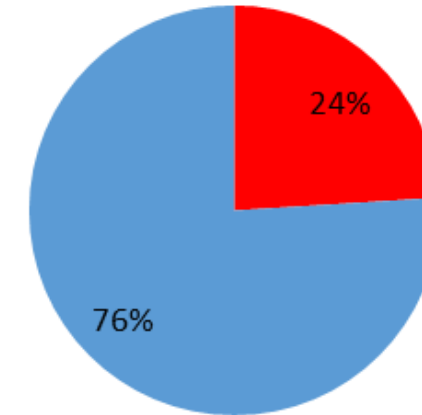
■ % Communications ■ % Computation

NEMO5 Profiling  
(RoCE)



■ % Communications ■ % Computation

NEMO5 Profiling  
(InfiniBand)



■ % Communications ■ % Computation

16 Nodes / 320 Processes

- **NEMO5 is the 5th edition of NanoElectronics MOdeling (NEMO) Tools**
- **InfiniBand provides 17% higher performance versus Ethernet RoCE**
- **InfiniBand provides 51% higher performance versus Ethernet TCP**
- **Ethernet RoCE provides 29% higher performance versus Ethernet TCP**
- **RDMA over Ethernet improved scalability on Ethernet network for NEMO5**
- **For the most time consuming MPI calls (as % of MPI time):**
  - MPI\_Send (67% of MPI time, 18% overall)
  - MPI\_Barrier (29% of MPI time, 8% overall)
  - MPI\_Recv (4% of MPI time, 1% overall)
- **For the most time consuming MPI calls is MPI\_Send**
  - Some load imbalance observed in the time spent by the MPI tasks





# Thank You



All trademarks are property of their respective owners. All information is provided "As-Is" without any kind of warranty. The HPC Advisory Council makes no representation to the accuracy and completeness of the information contained herein. HPC Advisory Council undertakes no duty and assumes no obligation to update or correct any information presented herein