EUROfusion

Benchmarks and validation of the LEONARDO HPC system

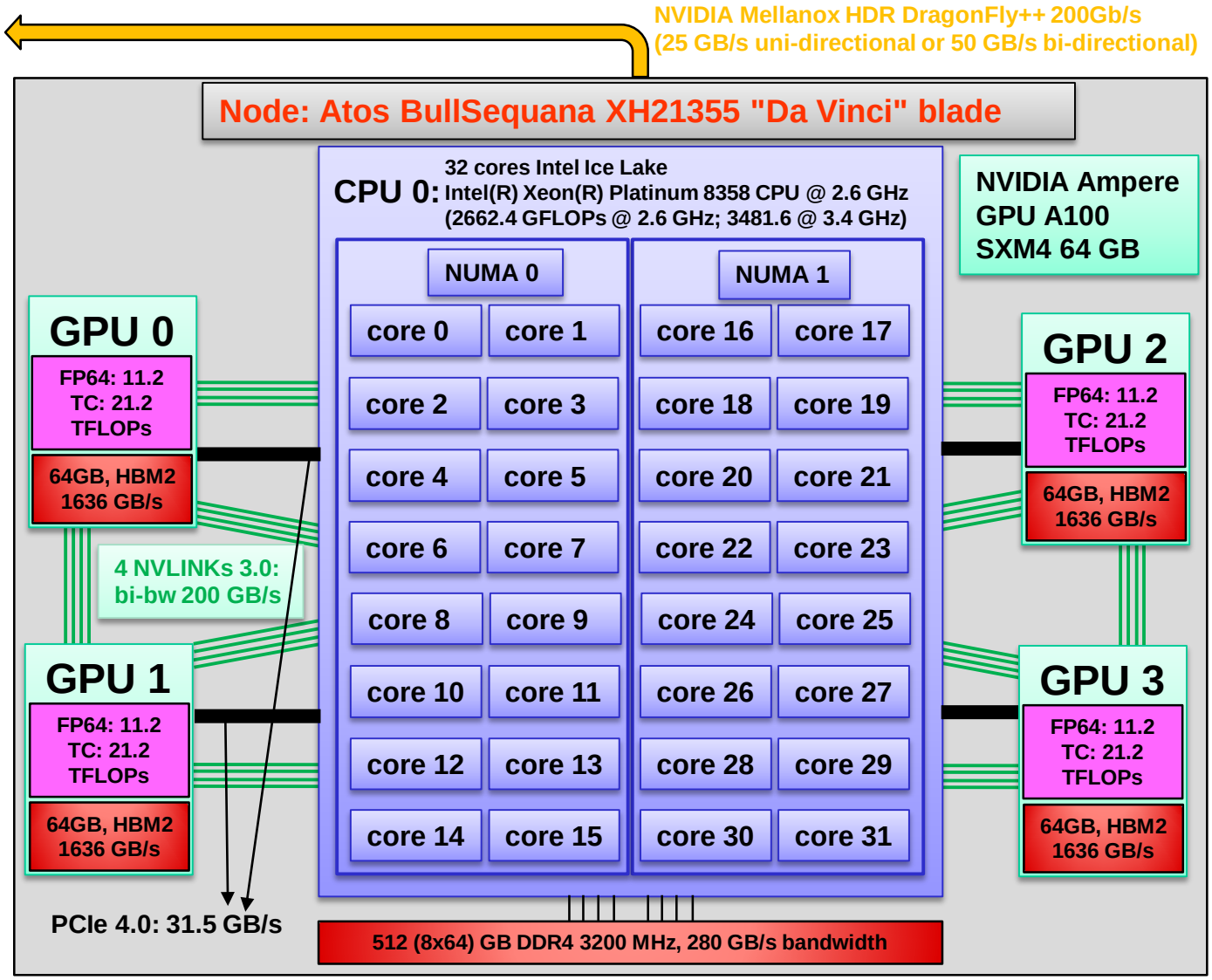
Serhiy Mochalskyy

Third IFERC workshop on the usage of GPU based system for fusion applications

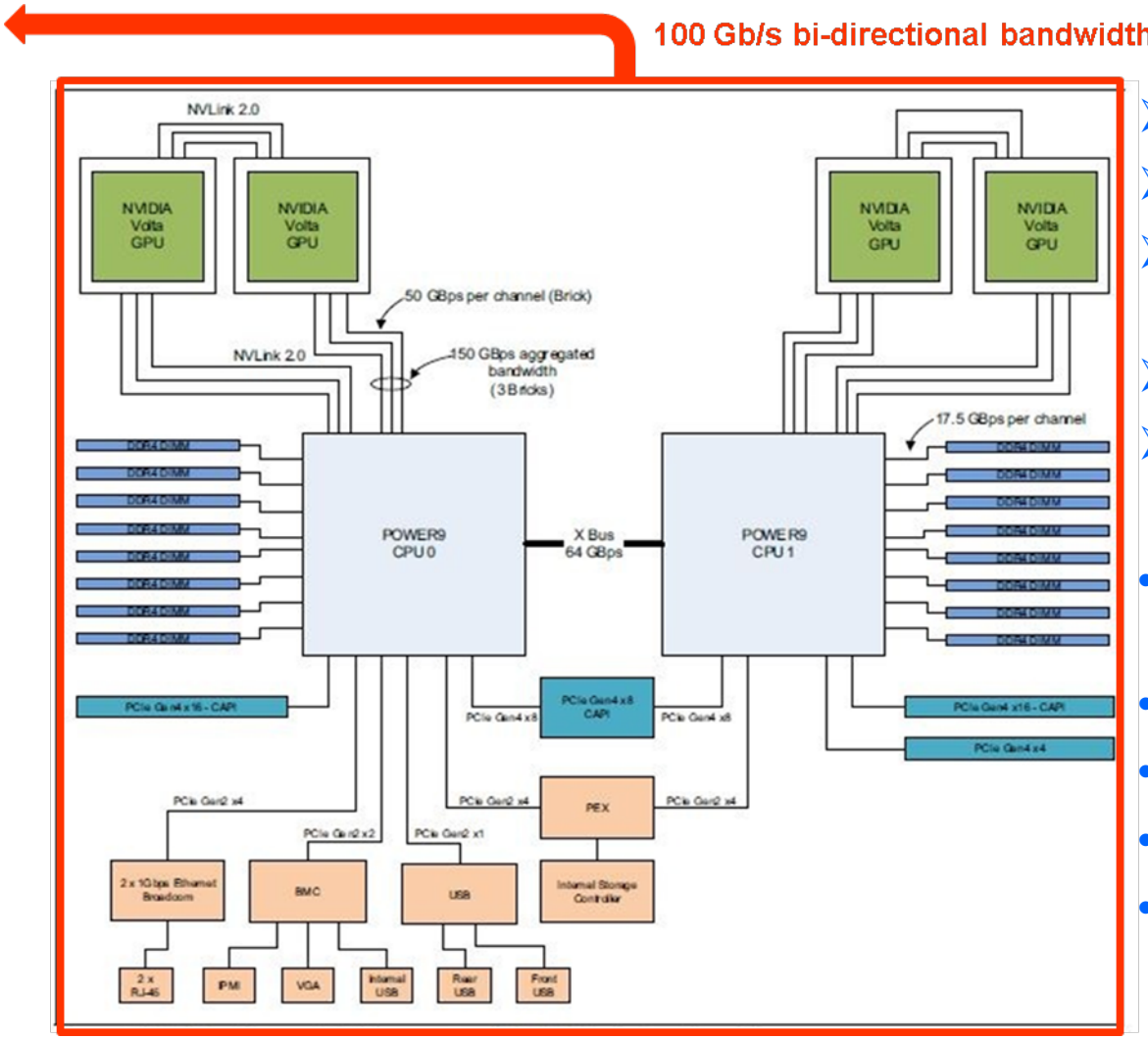
June 22, 2023

Advanced Computing Hub Garching
Max-Planck-Institut für Plasmaphysik
Boltzmannstr. 2, D-85748 Garching, Germany

LEONARDO Atos BullSequana XH21355 "Da Vinci" blade node



MARCONI100 Power9 node (AC922) architecture

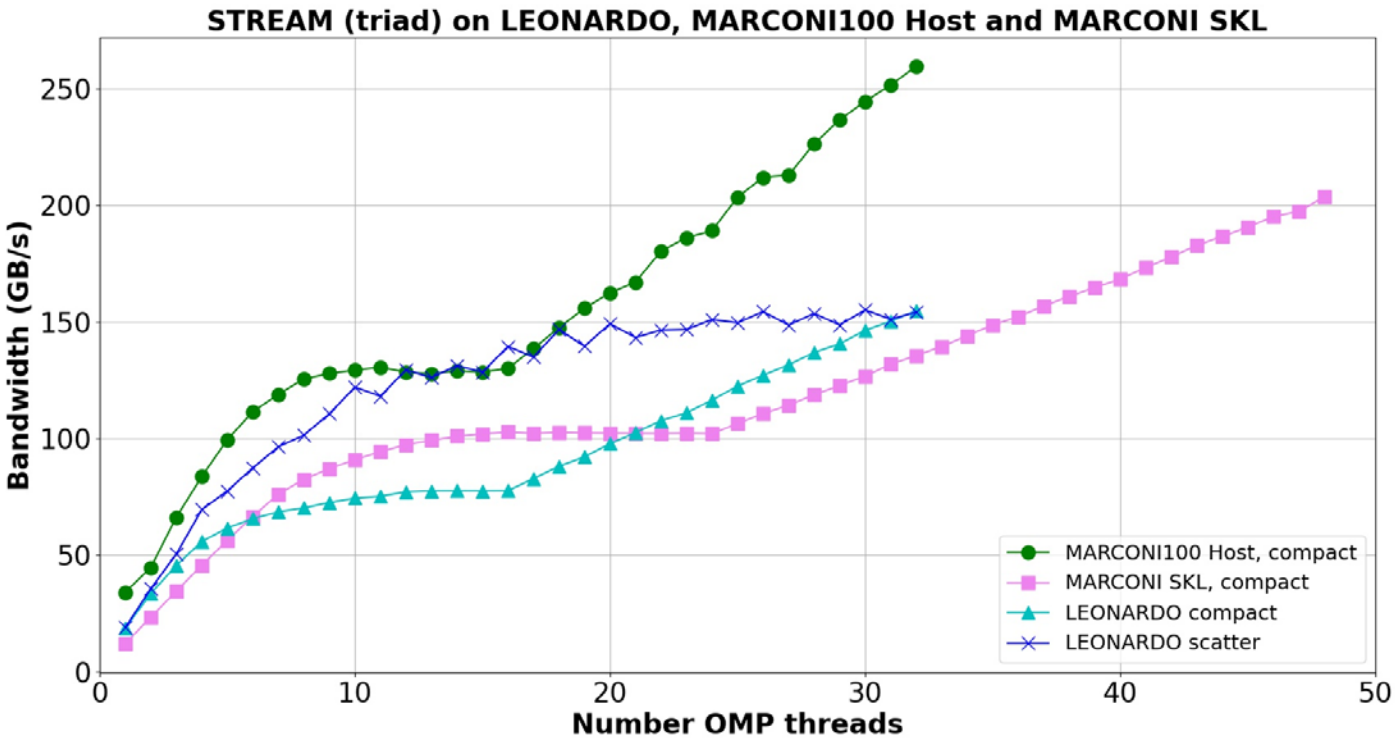


- 1 node
- 2 IBM Sockets (CPUs) / node
- 4 NVIDIA Volta V100 GPUs / node
- 16 cores / CPU
- 4 threads / core

- node – node: Mellanox IB EDR DragonFly+
- CPU – CPU: X Bus
- CPU – GPU: NVLink
- GPU – GPU: NVLink
- DDR4 – CPU: PCIe

STREAM on a single node

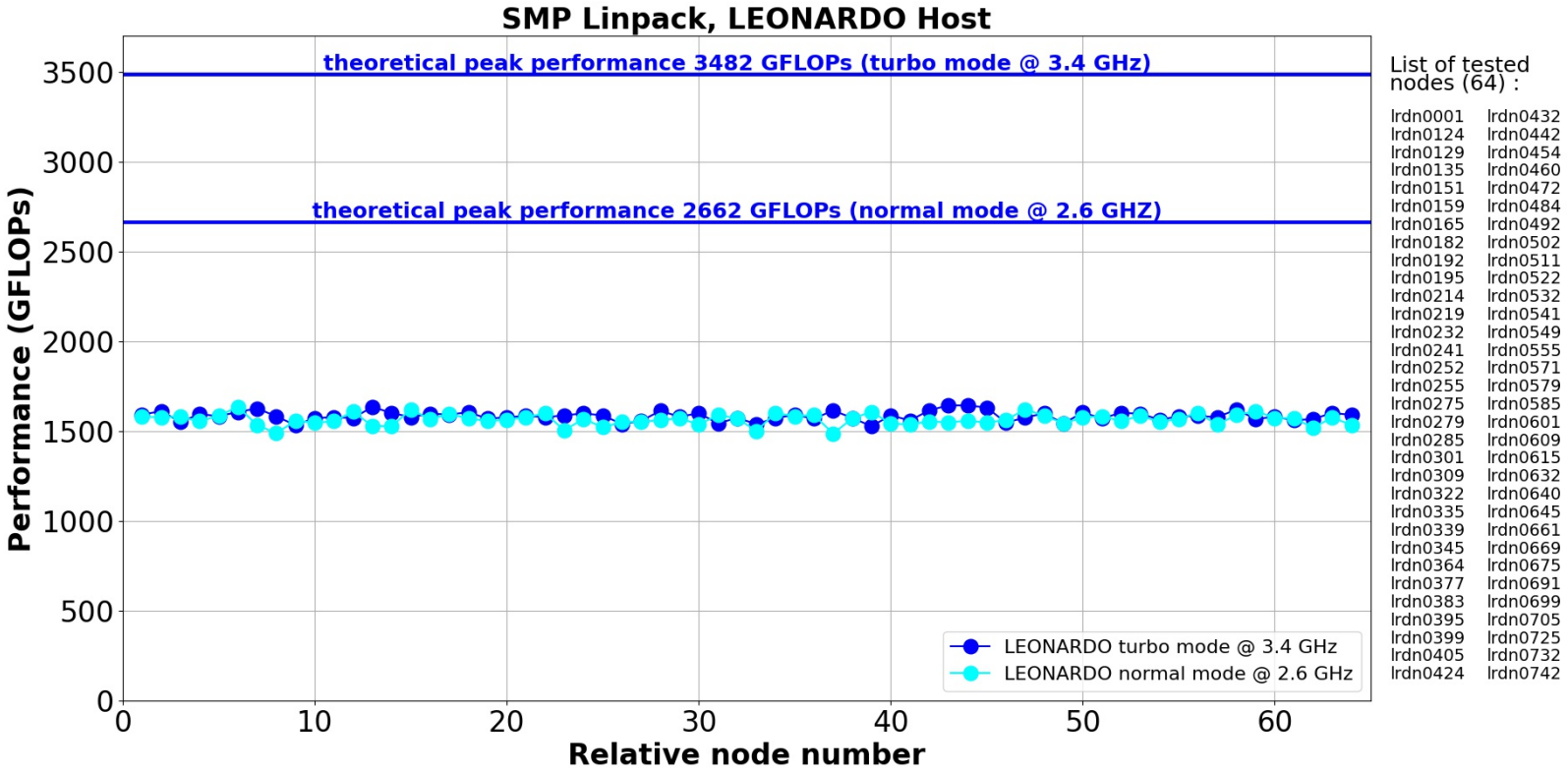
```
icx -O3 -qopenmp -mcmodel=medium -qopt-streaming-stores=always -mtune=icelake-client -xHost
-DSTREAM_ARRAY_SIZE=400000000 -DVERBOSE  -DNTIMES=100 stream.c -o stream_intel.x
```



- MARCONI SKL: 203 GB/s from 255.94 GB/s theoretical (80%).
- MARCONI100: 266 GB/s from 280 GB/s theoretical (95%).
- LEONARDO: 153 GB/s from 205 GB/s theoretical (75%).

April 20, 2023

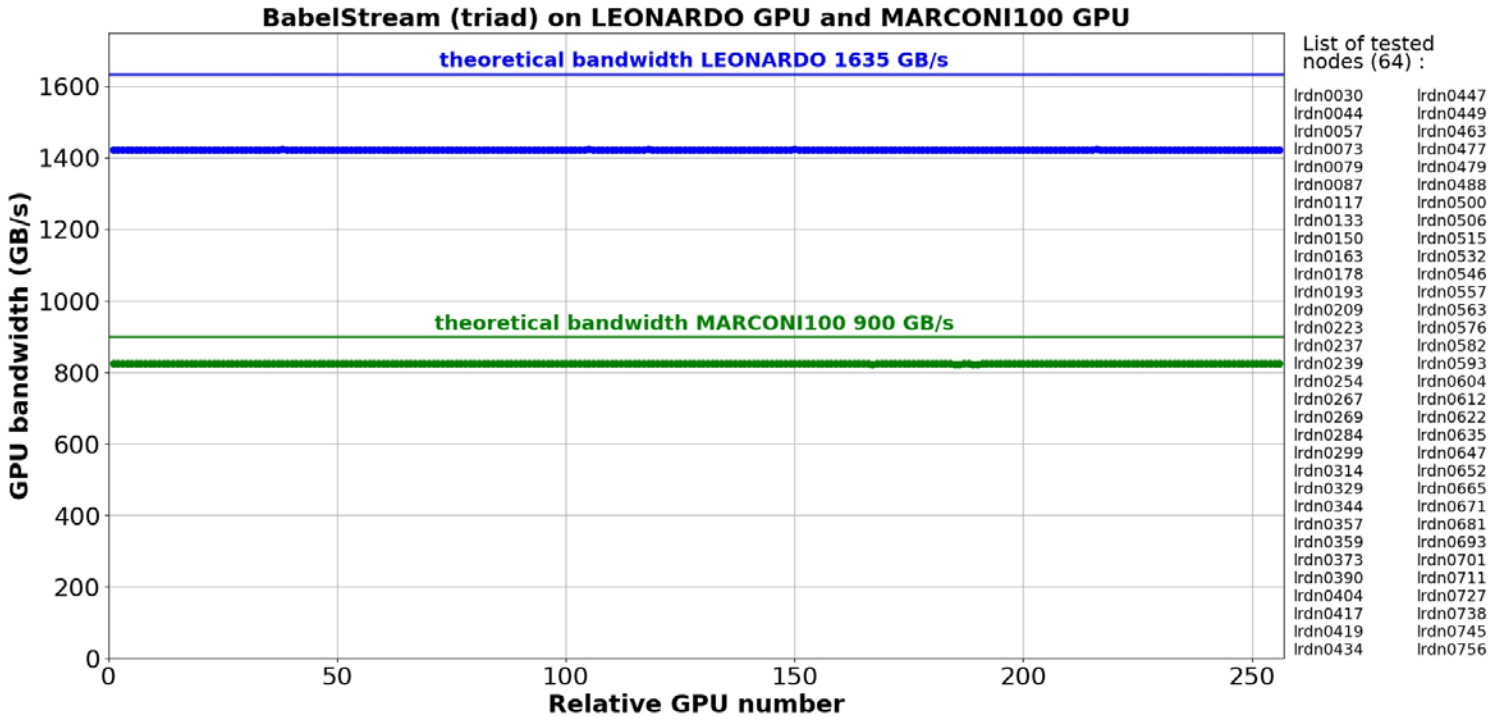
Performance on Host, stability test



- All nodes provide stable performance.
- The performance is relatively low:
  - turbo mode: mean=1584 GFLOPs (theoretical peak 3482 GFLOPs ~45.5%).
  - normal mode: mean=1563 GFLOPs (theoretical peak 2662.4 GFLOPs ~58.7%).
- MARCONI SKL = 2028.28 GFLOPs (theoretical peak 3211 GFLOPs ~63%).

May 22, 2023

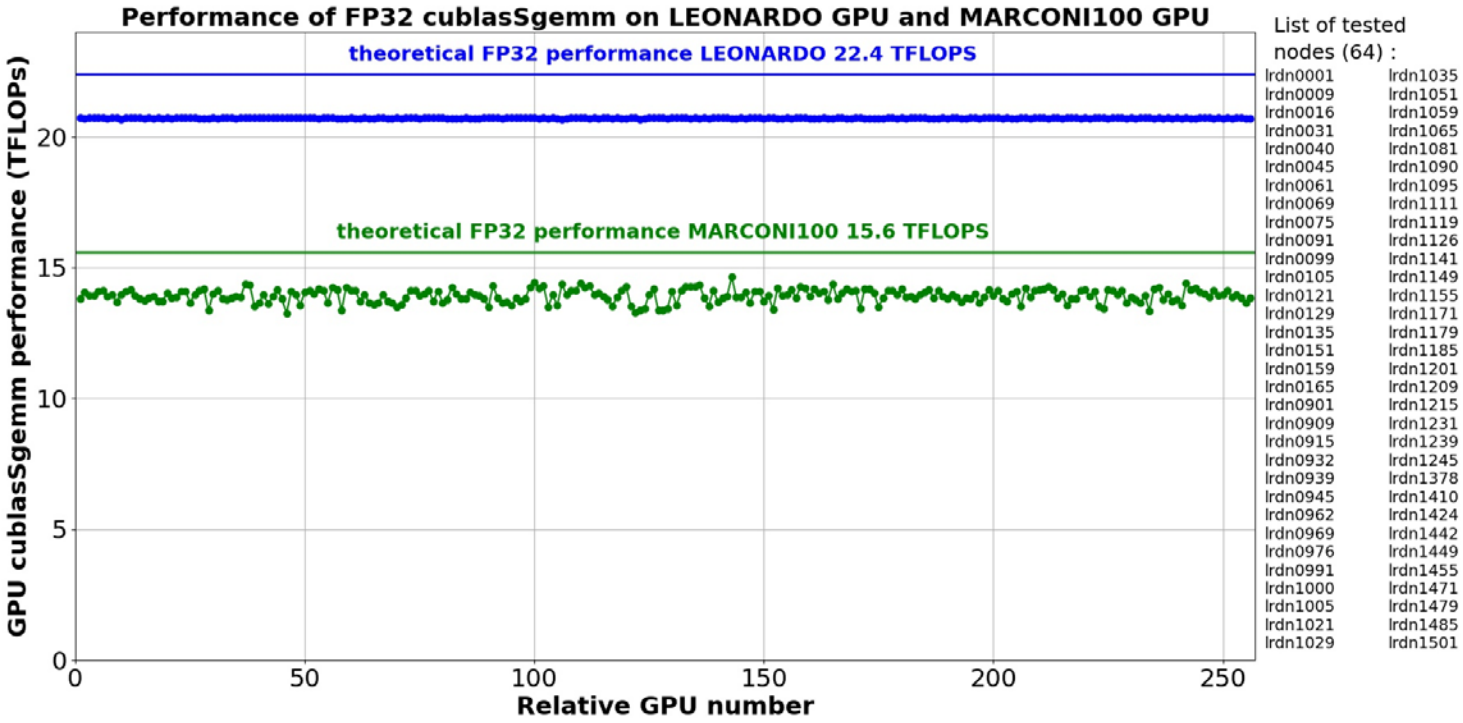
BabelStream benchmark on LEONARDO GPU



- All GPUs provide high, stable and symmetric bandwidth close to the theoretical value.
- No difference between GPUs on different nodes or GPUs inside one node.
- LEONARDO: 1423.5 GB/s from 1635 GB/s theoretical (87%).
- MARCONI100: 845 GB/s from 900 GB/s theoretical (94%).

April 26, 2023

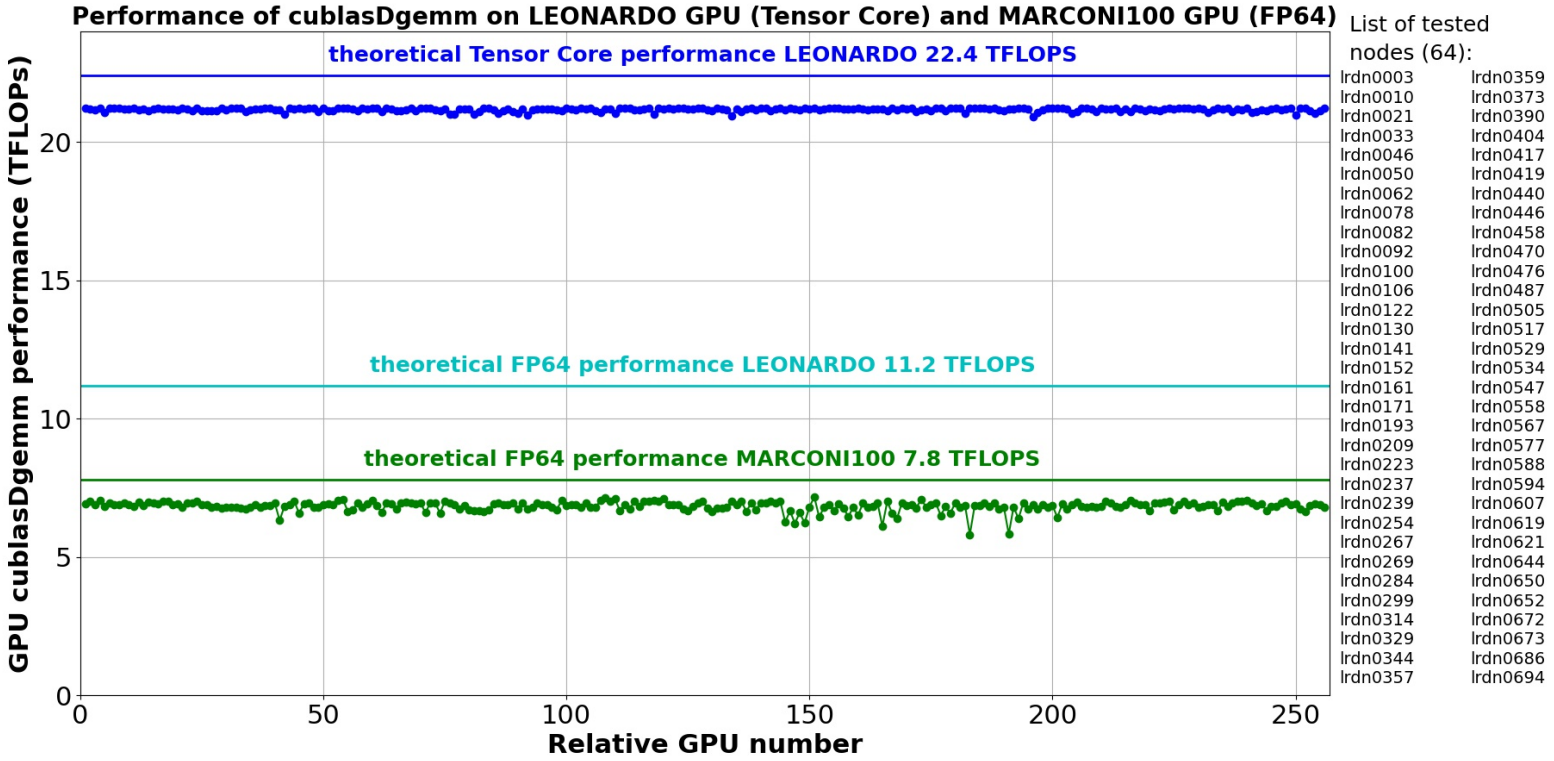
SGEMM (cublasSgemm) benchmark on LEONARDO GPU



- All GPUs provide high, stable and symmetric performance close to the theoretical value.
- No difference between GPUs on different nodes or GPUs inside one node.
- LEONARDO: 20.7 TFLOPs per GPU from 22.4 TFLOPs theoretical (92%).
- MARCONI100: 14 TFLOPs per GPU from 15.6 TFLOPs theoretical (90%).

April 27, 2023

DGEMM (cublasDgemm) benchmark on LEONARDO GPU

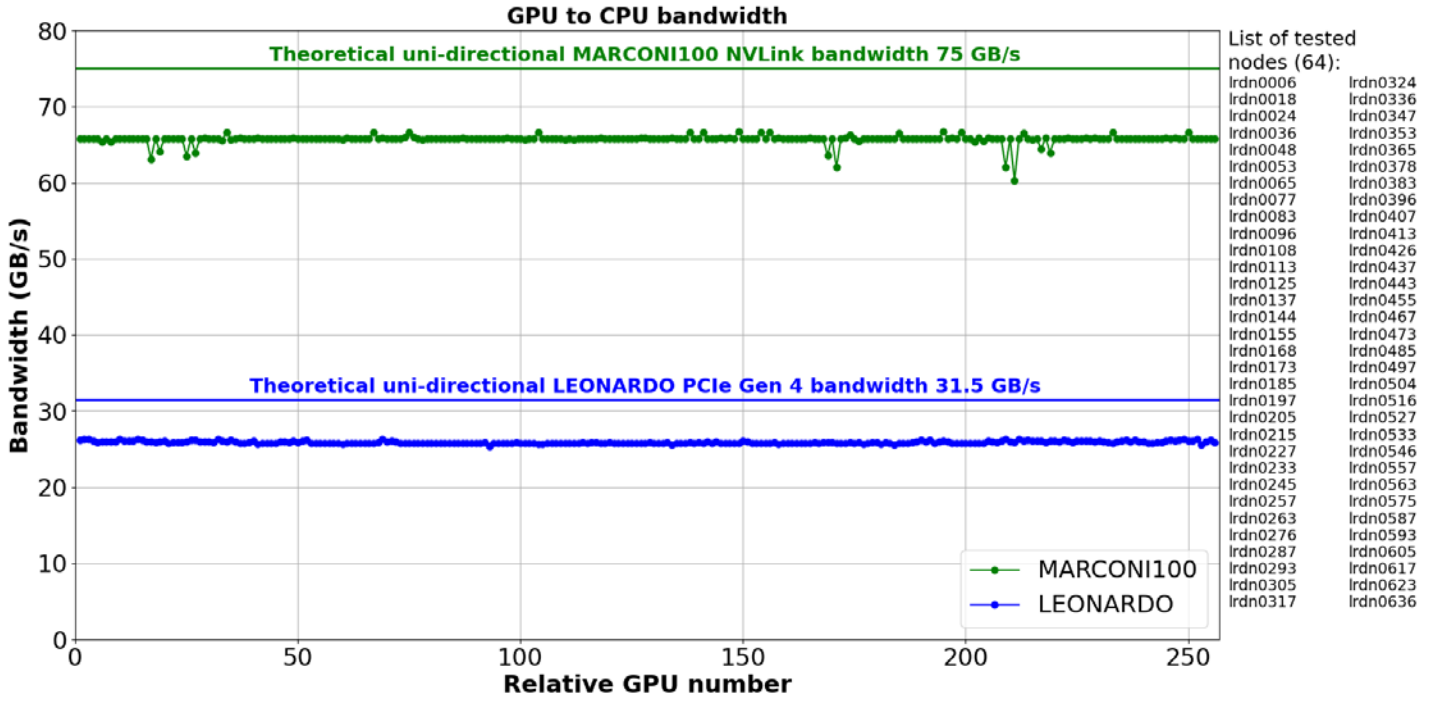


- All GPUs provide high, stable and symmetric performance close to the theoretical value.
- No difference between GPUs on different nodes or GPUs inside one node.
- LEONARDO FP64 Tensor Core: 21.2 TFLOPs per GPU from 22.4 TFLOPs theoretical peak (95%).
- LEONARDO FP64: 11.2 TFLOPs per GPU theoretical peak.
- MARCONI100 FP64: 6.8 TFLOPs per GPU from 7.8 TFLOPs theoretical peak (87%).

April 27, 2023

GPU to CPU bandwidth

using bandwidthTest benchmark from NVIDIA samples

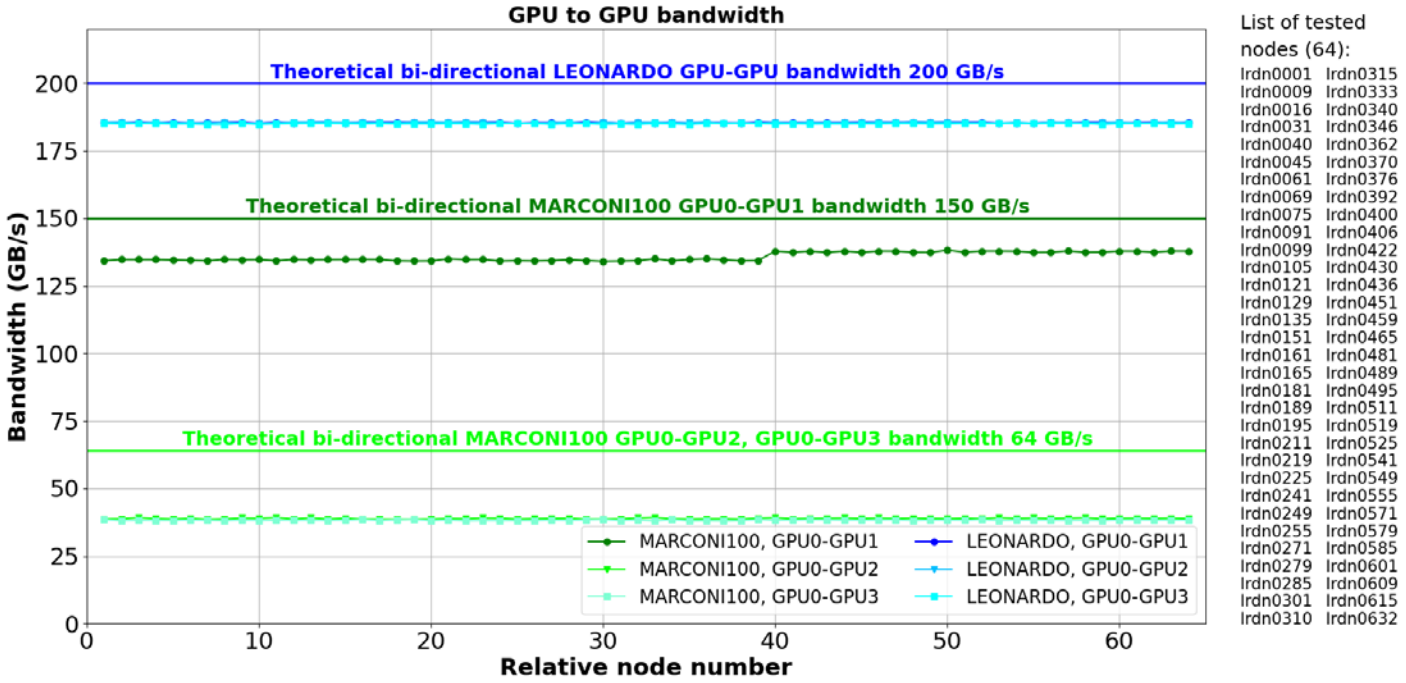


- The results are stable on LEONARDO and stable on MARCONI100.
- LEONARDO: the mean uni-directional bandwidth of ~26 GB/s from 31.5 GB/s of the theoretical value (83 %).
- MARCONI100: the mean uni-directional bandwidth of ~66 GB/s from 75 GB/s of the theoretical value (88 %).

May 2, 2023

GPU to GPU bandwidth

using p2pBandwidthLatencyTest benchmark from NVIDIA samples



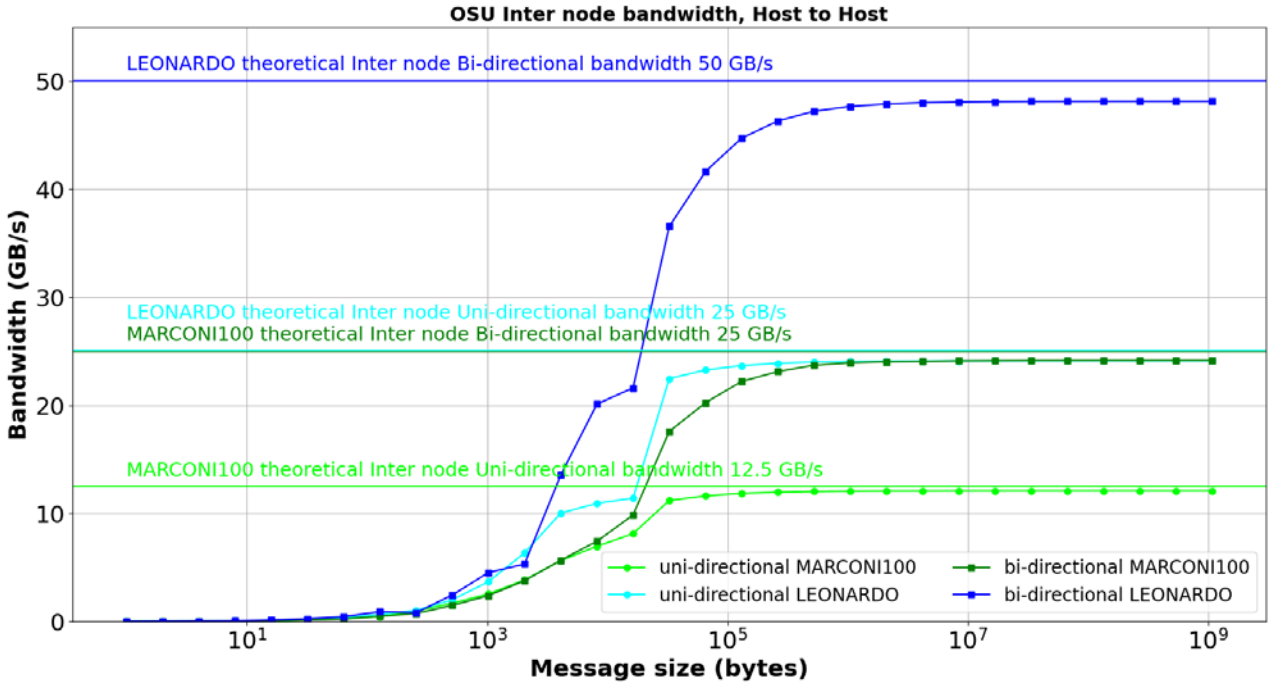
- The results are stable and symmetric.
- LEONARDO: the mean bi-directional bandwidth of all GPU pairs 185.5 GB/s from 200 GB/s of the theoretical value (93 %): 4 NVLINK with 50 GB/s each.
- MARCONI100: the mean bi-directional bandwidth of ~136 GB/s from 150 GB/s of the theoretical value (90 %) and ~39 GB/s from 60 GB/s of the theoretical value (60 %).

May 3, 2023

Inter node network bandwidth, Host to Host

NVIDIA Mellanox HDR DragonFly++ 200Gb/s (25 GB/s) uni-directional or 50 GB/s bi-directional

using osu_bw and osu_bibw benchmarks from OSU microbenchmark

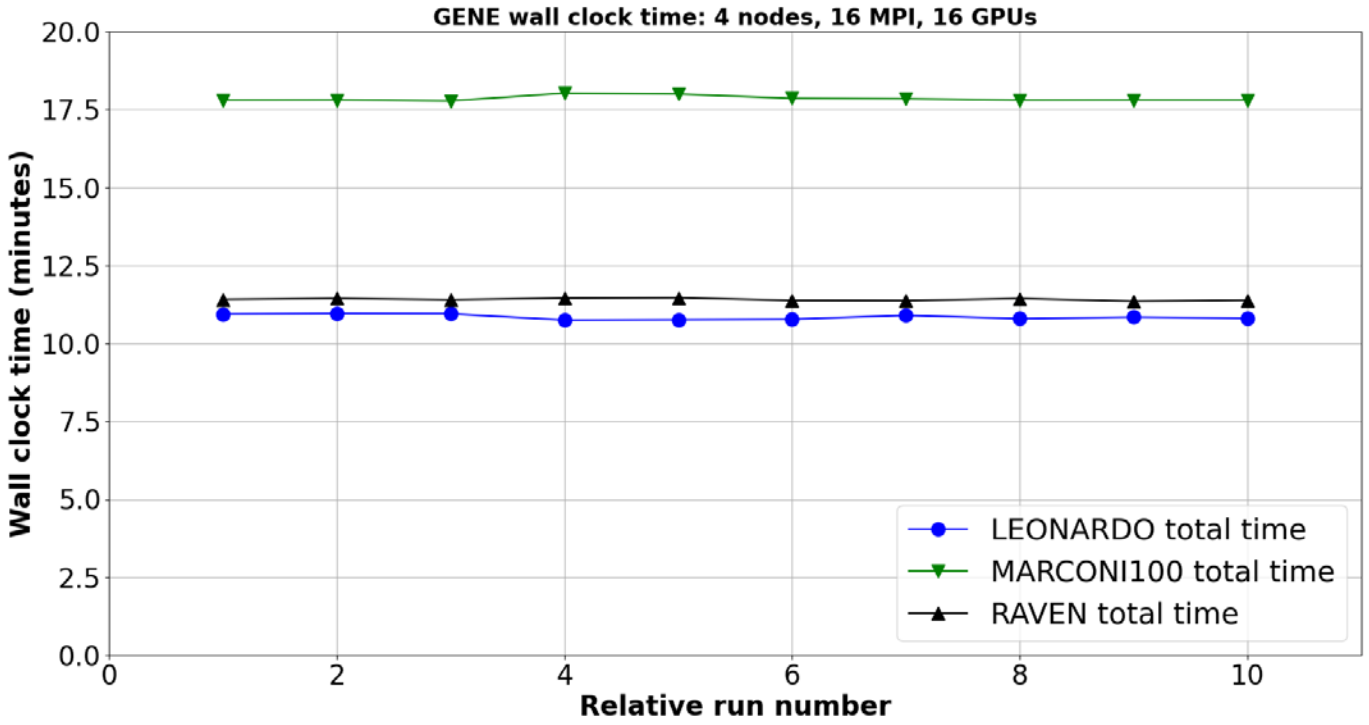


- **Stable and high bandwidth for uni- and bi-directional data transfer.**
- **LEONARDO: bi-directional bandwidth ~48 GB/s from 50 GB/s of the theoretical value (96 %).**
- **LEONARDO: uni-directional bandwidth ~24 GB/s from 25 GB/s of the theoretical value (96 %).**
- **MARCONI100: bi-directional bandwidth ~24.2 GB/s from 25 GB/s of the theoretical value (97 %).**
- **MARCONI100: uni-directional bandwidth ~12.1 GB/s from 12.5 GB/s of the theoretical value (99 %).**

May 3, 2023

GENE performance

4 nodes, 16 MPI, 16 GPUs

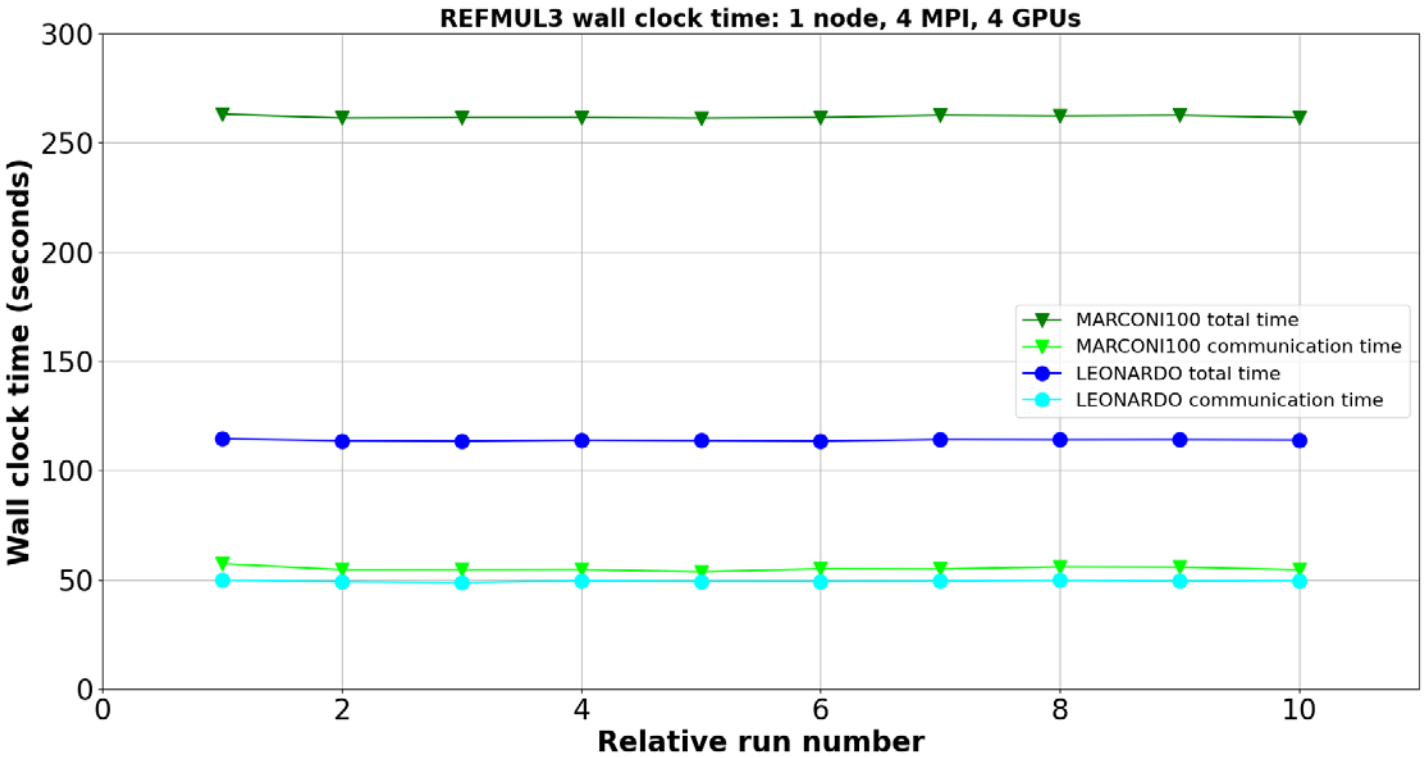


- The execution time is stable on all supercomputers.
- The code is faster on LEONARDO (factor of ~1.6) in comparison to MARCONI100.
- LEONARDO and RAVEN provide similar wall clock time.

May 30, 2023

REFMUL3 performance

1 nodes, 4 MPI, 4 GPUs

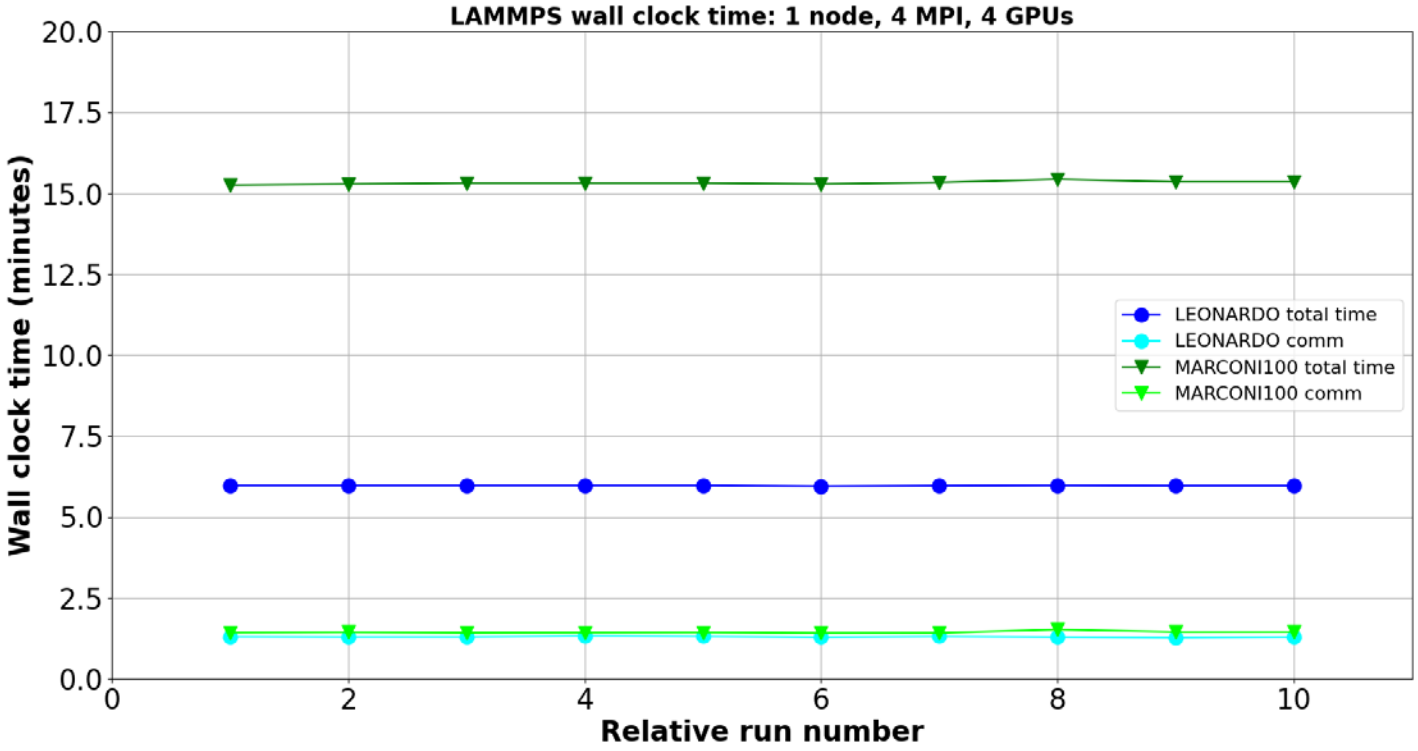


- Thanks to Tiago Ribeiro and Filipe da Silva for executing tests.
- The execution time is stable on both supercomputers.
- The code is faster on LEONARDO (factor of ~2.3).
- The communication time is similar (21 % of total time).

May 9, 2023

LAMMPS performance (1)

1 nodes, 4 MPI, 4 GPUs

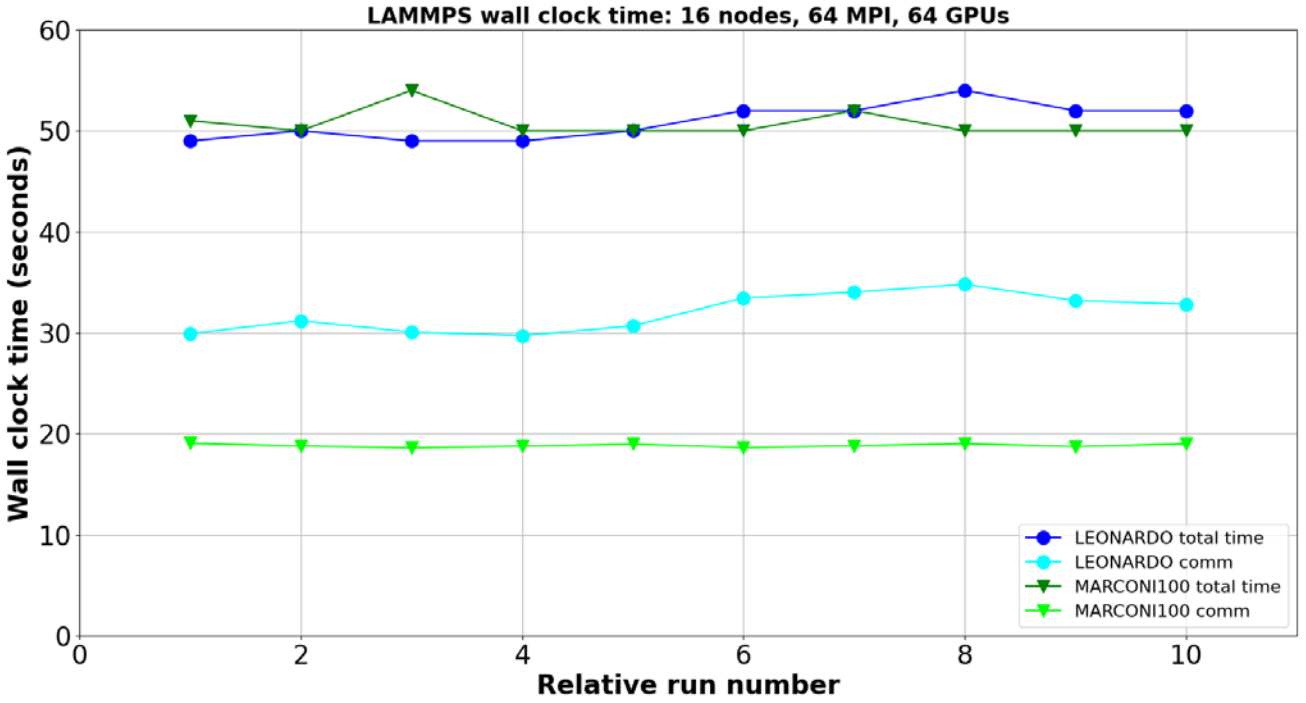


- The execution time is stable on both supercomputers.
- The code is faster on LEONARDO (factor of ~2.6).
- The communication time is similar.

May 8, 2023

LAMMPS performance (2)

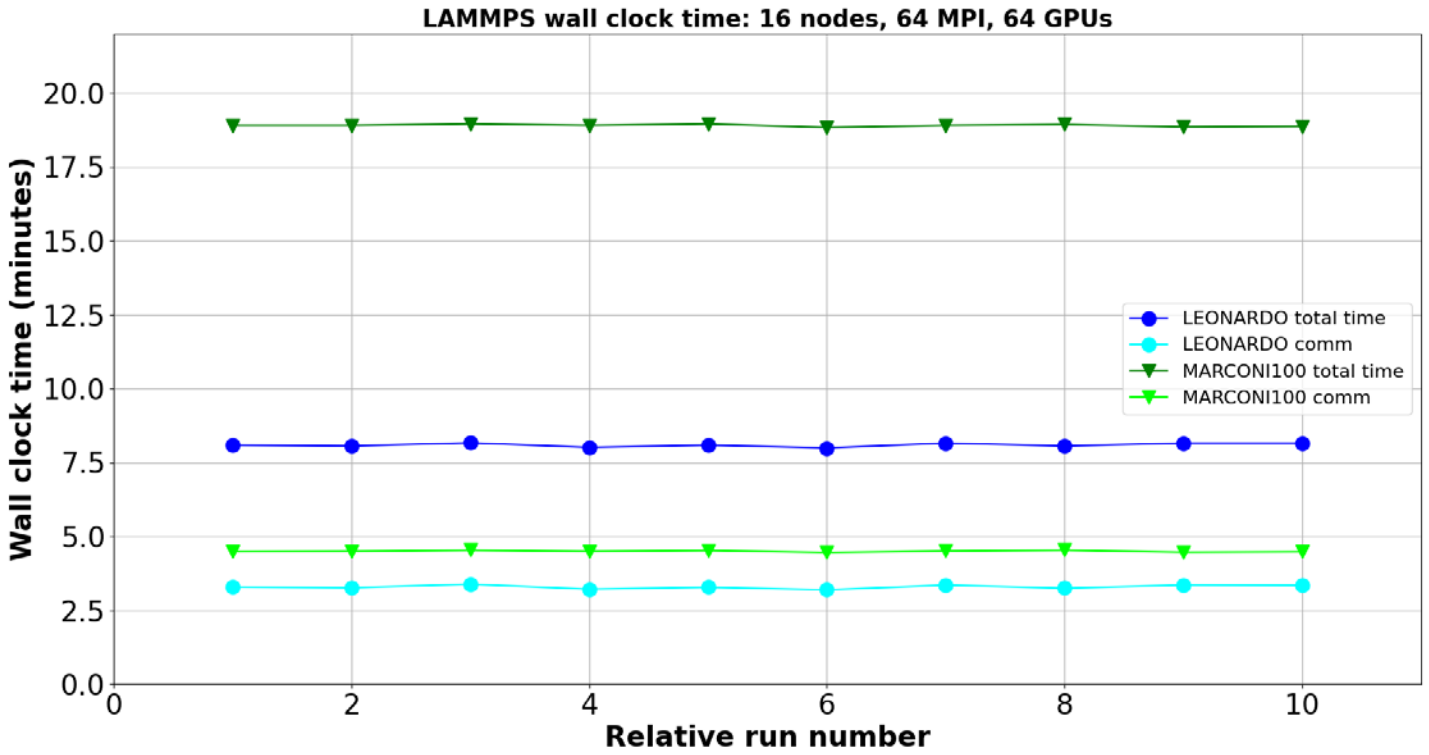
16 nodes, 64 MPI, 64 GPUs, strong scaling



- **When communication become dominant MARCONI100 is faster than LEONARDO.**

LAMMPS performance (3)

16 nodes, 64 MPI, 64 GPUs, weak scaling



- The execution time is stable on both supercomputers.
- The code is faster on LEONARDO (factor of ~2.3).
- The communication time is similar.

May 9, 2023

Conclusions

1. New node architecture: 1 CPU (1 socket), 2 NUMAs.
2. Host (CPU): LEONARDO has a lower memory bandwidth compared to M100 and/or SKL due to its single socket design (one socket vs 2 sockets).
3. Host (CPU): LEONARDO exhibits lower performance compared to SKL due to its single socket design (one socket vs 2 sockets), which offers fewer processing cores.
4. Device (GPU): LEONARDO has a higher memory bandwidth than a MARCONI100 GPU.
5. Device (GPU): LEONARDO demonstrates higher performance than a MARCONI100 GPU.
6. CPU to GPU or GPU to CPU communication: MARCONI100 offers better communication capabilities due to its use of NVLINK, a high-bandwidth interconnect technology, compared to the PCIe interface used in LEONARDO.
7. GPU to GPU communication: LEONARDO surpasses MARCONI100 in this aspect, as it utilizes NVLINK 3.0, which provides higher bandwidth than the NVLINK 2.0 used in MARCONI100.
8. Inter-node network bandwidth (node-to-node communication): LEONARDO exhibits superior performance in terms of network bandwidth with its NVIDIA Mellanox DragonFly++ 200Gb/s, outperforming the Mellanox DragonFly++ 100Gb/s on MARCONI100.
9. REFMUL3 real code performance: LEONARDO demonstrates better performance (factor 2.3-2.6).
10. LAMMPS real code performance: LEONADRO is better (factor 2.3-2.6).
11. In cases where communication between CPU and GPU dominates: MARCONI100 outperforms LEONARDO, likely due to its better communication capabilities using NVLINK.
12. GENE real code performance: LEONADRO is better (factor 1.6) in comparison to MARCONI100 and similar with RAVEN.

Thank you for our attention!