

Experience feedback on Fugaku supercomputer

Serhiy Mochalsky, Roman Hatzky

IFERC Annual Workshop on GPUs #2

June 14th, 2021

High Level Support Team

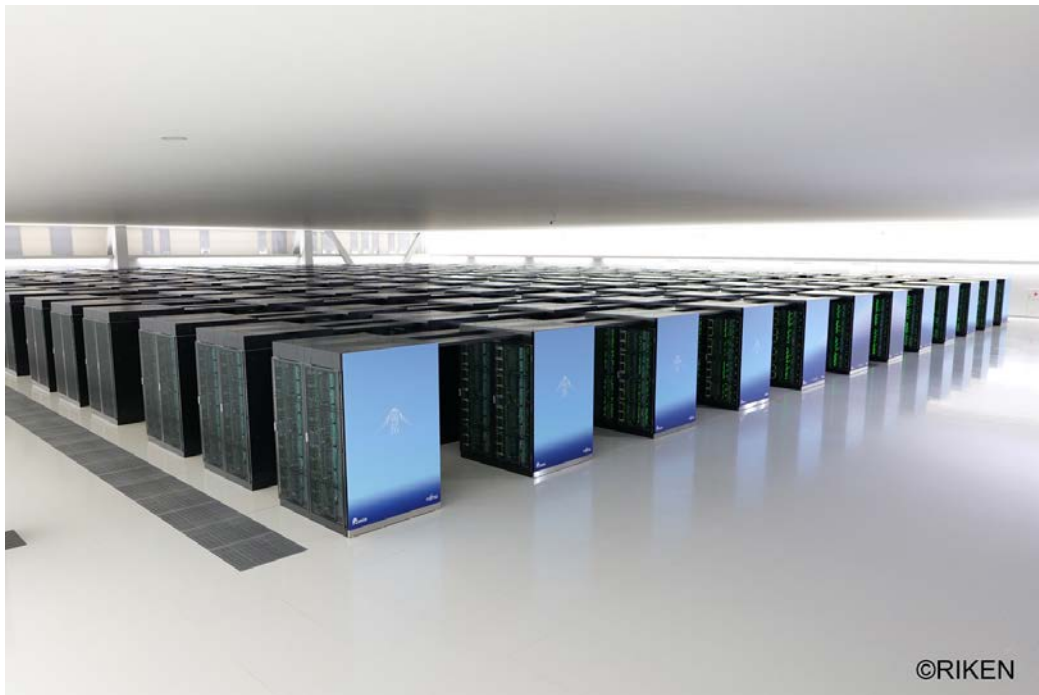
Max-Planck-Institut für Plasmaphysik

Boltzmannstr. 2, D-85748 Garching, Germany

Fugaku supercomputer

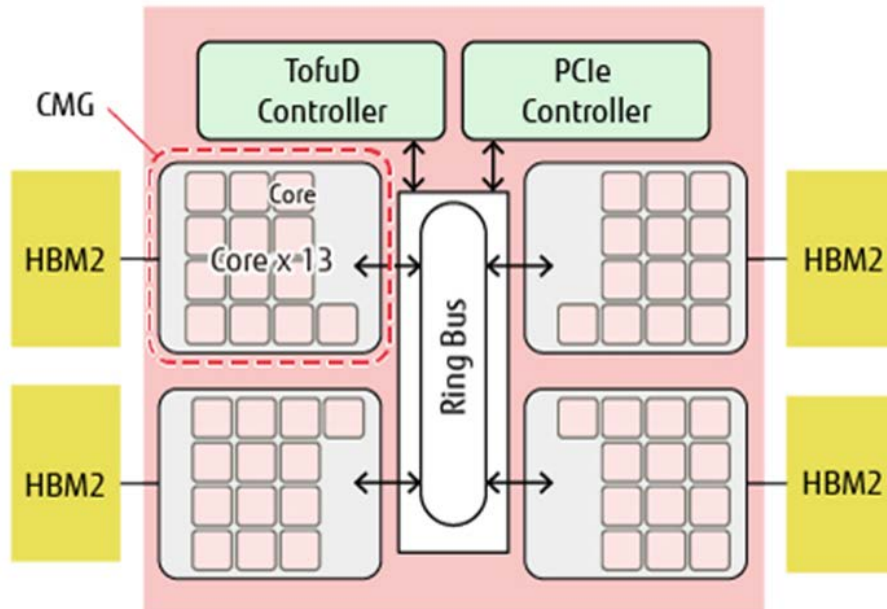


The **Fugaku supercomputer** is the **number one** system on the **Top500** list at the RIKEN Center for Computational Science in Kobe, Japan.



- Launched in **June 2020**.
- Production phase 2021.
- Theoretical peak performance **537 PFLOPS**.
- **442 PFLOPS** in the High Performance Linpack.

Fujitsu A64FX

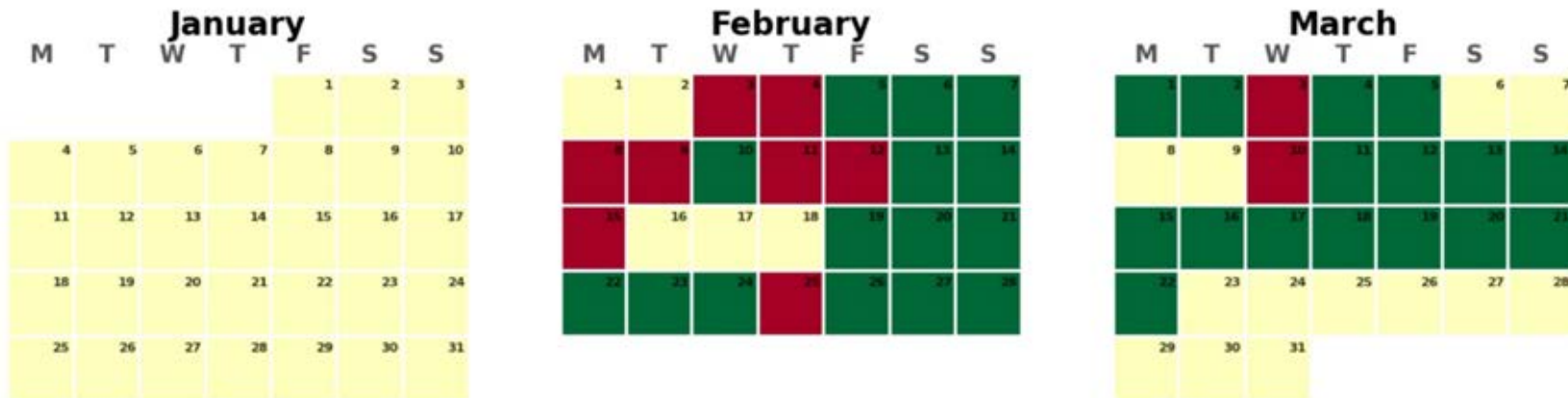





Marconi SKL vs Fugaku A64FX

Name	Marconi SKL	Fugaku
Processor name	Intel Xeon Platinum 8160	Fujitsu A64FX
Base CPU frequency (GHz)	2.1	2.0
Number of cores per node	48	48
Number of sockets per node	2	4
Peak performance per node (GFLOPS)	3225	3072
Interconnect	OPA	TofuD
Inter-node uni-directional bandwidth (GB/s)	12.5	$6.8 * 6 = 40.8$
Intra-node uni-directional bandwidth (GB/s)	not available	256
Main memory capacity (GB)	192	32
Main memory bandwidth (GB/s)	255.94	1024

- Similar performance (**~3 TFLOPS**) and number of compute cores (**48**).
- Different node architecture (**2 vs 4 sockets**).
- Different memory bandwidth (**1024 vs 255 GB/s**) and capacity (**32-5 vs 192-5 GB**).
- Different inter and intra node networks.

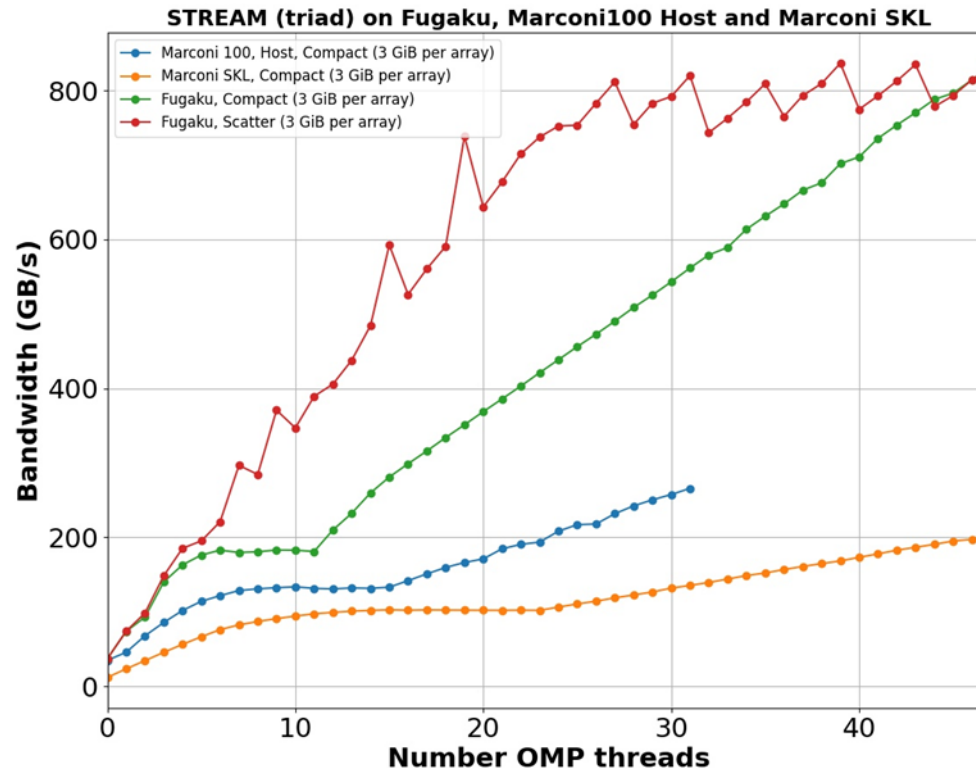
The procedure for opening an account and gaining access to the Fugaku supercomputer was relatively complicated and **took more than six weeks**. Access was granted on 12.01.2021.



-  Maintenance or large scale job execution (all nodes)
-  Issues and compute nodes unavailability
-  Regular operation

- Different **problems appeared regularly** during the whole project lifetime, which lasted until 31.03.2021.

STREAM on a single node



- Marconi SKL: **203 GB/s** from 255.94 GB/s theoretical (**80 %**).
- Marconi100: **266 GB/s** from 280 GB/s theoretical (**95 %**).
- Fugaku: **831 GB/s** from 1024 GB/s theoretical (**81 %**), with not default parameters.



The benchmark solves a dense (real*8) system of linear equations ($Ax=b$), measures the amount of time it takes to factor and solve the system, converts that time into a performance rate.

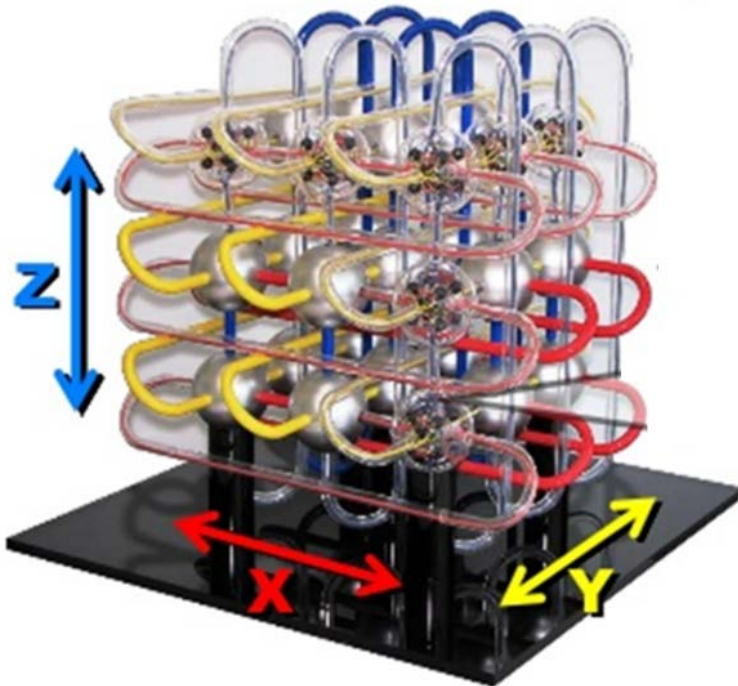
Marconi = **2.171** TFLOPS (theoretical peak 3.225 TFLOPS ~**67 %**)

Fugaku = **1.869** TFLOPS (theoretical peak 3072 TFLOPS ~**61 %**)

Fugaku, from support team = **2.152** TFLOPS (**70 %**)

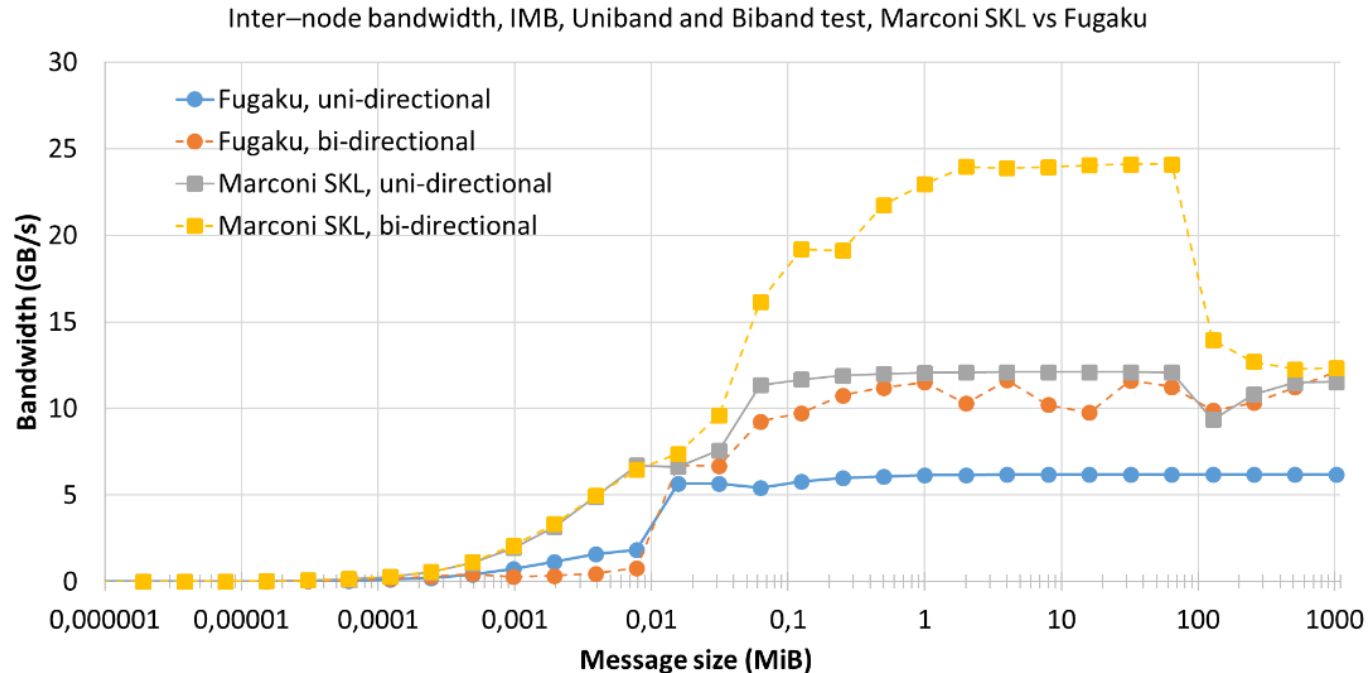
Fugaku in Top500 = **82 %** of theoretical peak

Fujitsu TofuD interconnect

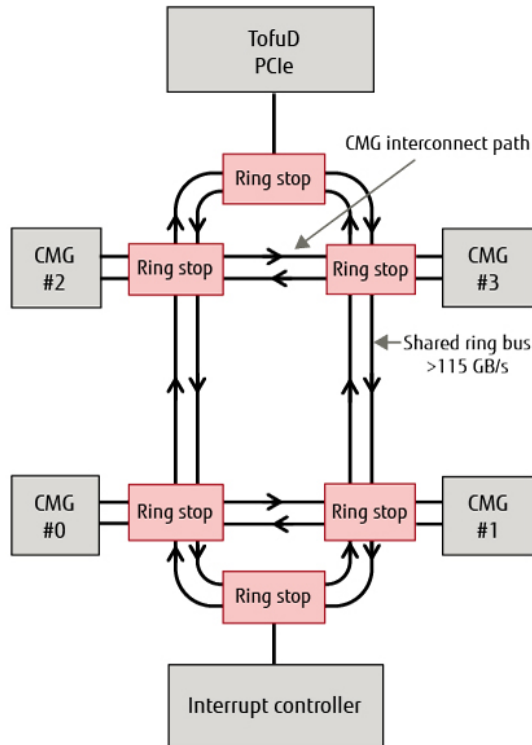


- Simultaneous data transfer in 6 directions
- 6 Tofu network interfaces (TNI)
- Peak memory bandwidth of **6.8 GB/s** per TNI.

- Theoretical bandwidth of **40.8 GB/s** (six TNI • 6.8 GB/s) can be achieved if a node communicates with six neighboring nodes.



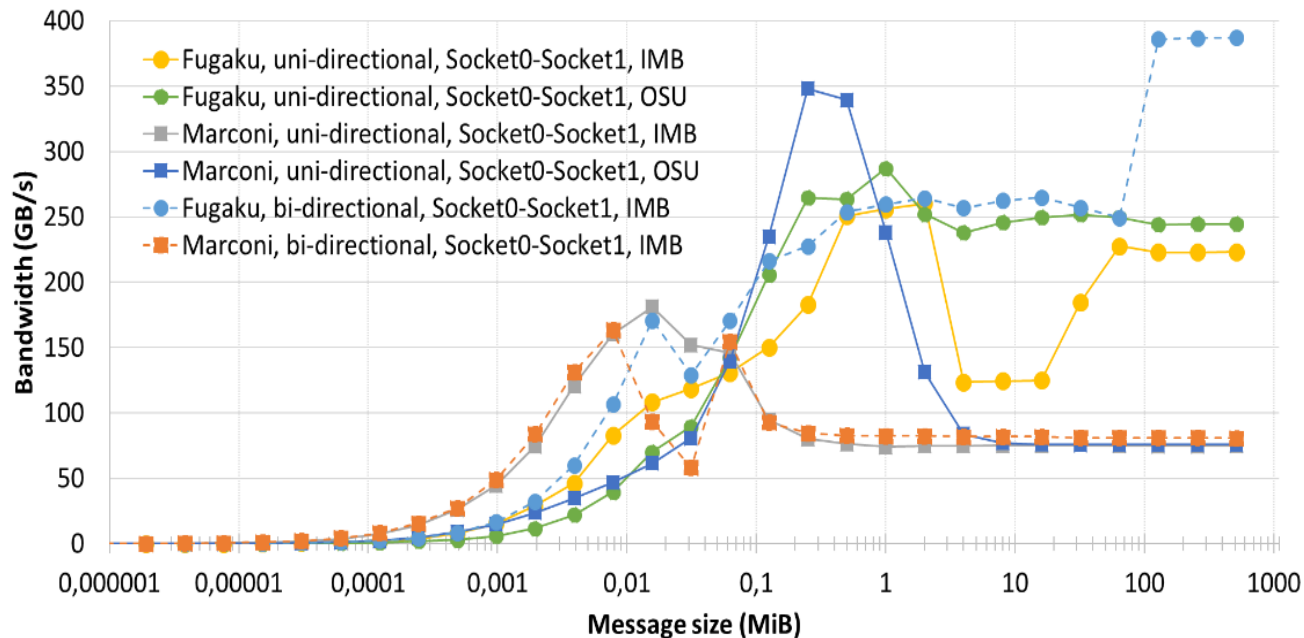
- A maximum *uni-directional* bandwidth of **6.2 GB/s (91.2 %)** and **12.1 GB/s (96.8 %)** reached on Fugaku and Marconi SKL using one TNI.
- A maximum *bi-directional* bandwidth of **12.1 GB/s (89 %)** and **24.1 GB/s (96.4 %)** reached on Fugaku and Marconi SKL using one TNI.
- Total uni-directional bandwidth on Fugaku using 6 TNI increases up to **38.1 GB/s** (results of Fugaku support team).



- **Connection between 4 sockets (CMG)**
- **Two ring bus network**

- **Theoretical uni-directional bandwidth of 128 GB/s (one ring) * 2 = 256 GB/s (512 GB/s of bi-directional bandwidth).**

Inter-socket bandwidth, IMB and OSU benchmarks, Marconi SKL vs Fugaku

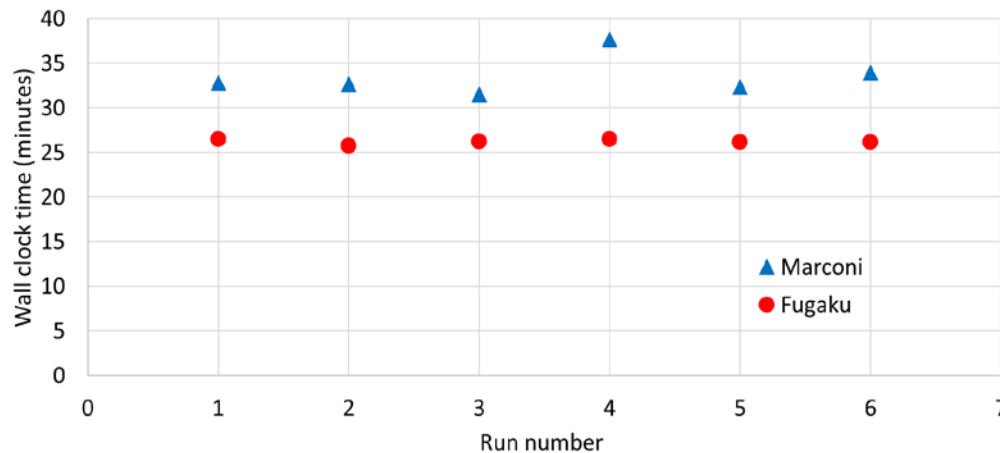


- Fugaku *uni-directional* bandwidth: **223–244 GB/s (87.1–95.3 %)**.
- Marconi *uni-directional* bandwidth: **75 GB/s**.
- Marconi *bi-directional* bandwidth: **81 GB/s**.
- Fugaku *bi-directional* bandwidth: **386 GB/s (75.4 %)**.

ONIX code

real execution time

ONIX code, 96 nodes, 4608 MPI tasks



profiling

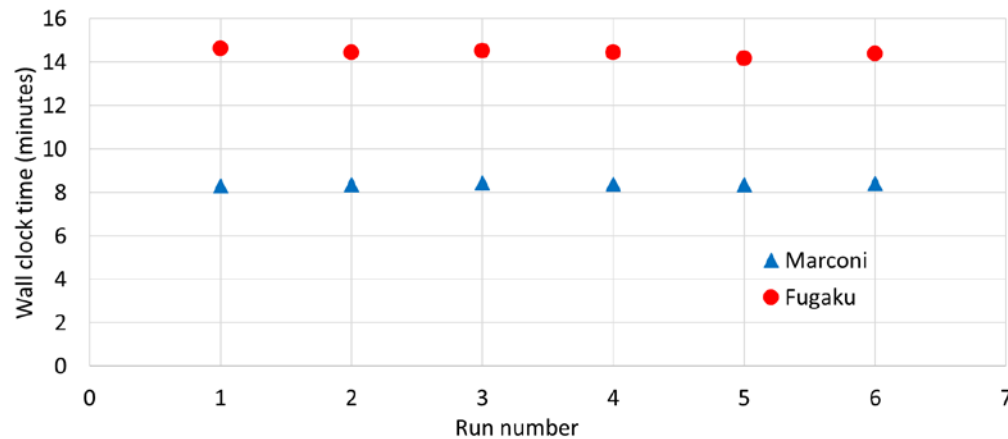
Subroutine name	Fugaku (sec)	Marconi (sec)	Speed-up (Fugaku/Marconi)
reset_variables	3,26564161	13,04882455	0,250
ChargeProjection_Second_orderAll	319,744771	107,6539345	2,970
ChargeExchange_3planes	10,1824197	6,323091269	1,610
Solve_Potential_bias	416,454874	1074,635107	0,388
Send_Plane	5,01059208	3,011766672	1,664
E_field_calculation15points	42,4912262	26,8415668	1,583
movepartAll	440,535539	209,9441543	2,098
Reflection2	167,272553	47,98361921	3,486
electron_thermalization	72,2587613	25,30626822	2,855
Particle_Exchange	1,53754144	4,13376236	0,372
Heapsort	2,02E-03	1,71E-03	1,183
DeleteParticle	1,12E-03	1,21E-03	0,923
AddParticle	29,4634382	431,4100542	0,068
Total	1615,98634	1950,295066	0,829

- **Memory bound subroutines: Fugaku faster.**
- **Compute bound subroutines : Marconi faster.**

STARWALL code

real execution time

STARWALL code, 16 nodes, 64 MPI



➤ **STARWALL code is mostly compute bound: Marconi faster.**

profiling

Subroutine name	Fugaku (sec)	Marconi (sec)	Speed-up Small testcase (Fugaku/Marconi)	Speed-up Large testcase (Fugaku/Marconi)
All Input	0,012	0,004	2,934	3,785
All surface_wall	0,741	1,223	0,605	0,900
matrix_pp	267,786	154,000	1,739	1,776
matrix_wp	287,796	157,253	1,830	1,892
matrix_ww	251,117	142,324	1,764	1,744
matrix_rw	0,016	0,017	0,915	0,700
matrix_cc	0,000	0,000	1,462	1,577
matrix_pe	0,007	0,011	0,690	0,513
matrix_ep	1,384	0,888	1,559	1,118
matrix_ew	1,277	1,156	1,105	1,189
matrix_ec	0,000	0,000	1,826	1,005
a_pe_transpose_sca	0,067	0,014	4,827	1,456
cholesky_solver	5,860	3,813	1,537	1,325
a_pwe_s_computing	0,887	0,475	1,867	2,002
a_ee_computing	0,039	0,023	1,684	0,860
a_ew_computing	0,178	0,148	1,204	1,054
a_we_computing	0,139	0,164	0,848	0,840
matrix_multiplication	1,982	3,549	0,559	0,546
simil_trafo	31,499	26,423	1,192	0,836
a_ye_computing	0,174	0,131	1,328	1,228
a_ey_computing	0,198	0,155	1,281	1,283
d_ee_computing	0,037	0,023	1,608	0,797
computing_s_ww_inverse	6,316	4,991	1,266	1,286
Total	857,521	496,796	1,726	1,642



EUTERPE code:

- ❖ *netcdf* (01.02.2021 – open; 25.02.2021 – resolved)
- ❖ *netcdf*
- ❖ *hdf5*
- ❖ *fftw*
- ❖ *PETSc* (25.02.2021 – open; 25.03.2021 – resolved)

GENE code:

- ❖ *hdf5*
- ❖ *fftw*
- ❖ *PETSc* (25.02.2021 – open; 25.03.2021 – resolved)

- **We ran out of time to complete these compilation (project ended 31.03.2021).**

Thank you for your attention