SCALABLE LATTICE BOLTZMANN LEAPS TO EXASCALE

J. Badwaik (JSC), R. Vavřík, O. Vysocký, K. Kadlubiak, L. Říha (IT4I), P. Suffa (FAU), G. Staffelbach, M. Holzer (CERFACS), R. Cuidard, D. Ricot (CS GROUP)

SC23 Denver, CO | I am hpc.

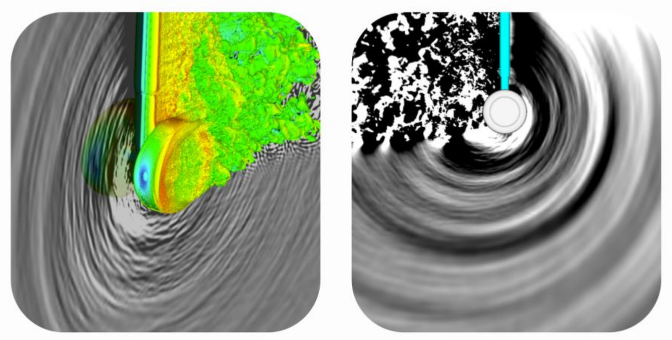
SCALABLE

INTRODUCTION

Lattice Boltzmann methods (LBM) are trustworthy alternatives to conventional CFD, showing roughly an order of magnitude performance advantage over Navier-Stokes approaches in comparable scenarios. The SCALABLE project brings together industrial and academic partners to create a new framework for an industrial LBM-based computational fluid dynamics (CFD) solver which can achieve high performance, scalability, and energy efficiency.

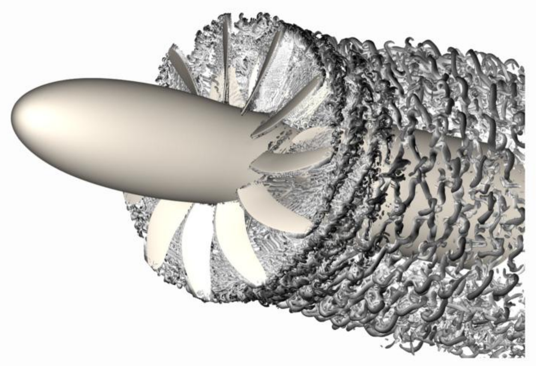
In the context of EuroHPC, LBM is especially well suited to exploit advanced supercomputers through vectorization, accelerators, and massive parallelization. To achieve its goals, SCALABLE aims to transfer leading-edge performance technology between waLBerla and LaBS, thus breaking the silos between the worlds of scientific computing and physical flow modelling.

TEST CASES AND INDUSTRIAL GRADE BENCHMARKS



SCALABLE has chosen a suite of test cases reflecting the interest of the industrial partners. One of these test cases is the Lagoon landing gear. During the landing phase of an aircraft, the landing gear is an important source of noise and produces a non-trivial turbulent flow as illustrated by the visualizations of turbulence patterns and acoustic waves on the left.

The simple geometry and an existing large experiment database make the test case attractive for both validation as well as performance evaluation.



As a final application, SCALABLE is focusing on a novel contra-rotating open-rotor propulsion system. This concept is one of the possible designs being studied to reduce emissions and improve efficiency in a post-carbon aviation future. The advancements from SCALABLE will be used to perform the LBM simulation of the Contra rotative open rotor Z08-AIPX7 case.

ENERGY EFFICIENCY OPTIMIZATION

Energy consumption can be reduced by exploiting the dynamic behavior of the different phases of the applications. In SCALABLE, the MERIC run time system developed as part of H2020 READEX project is used to scale CPU core and uncore frequencies or GPU SM frequency to best fit the needs of the executed phase of the applications.

Dynamic Tuning of LaBS for Lagoon.

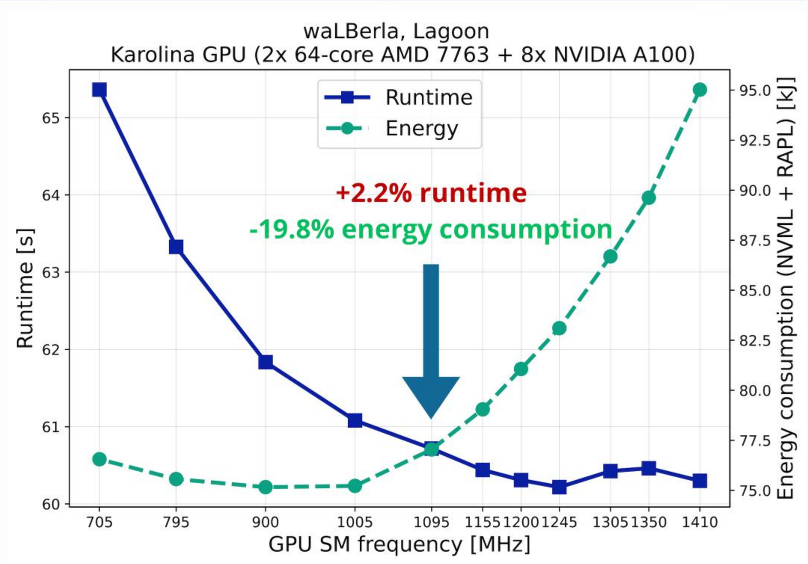
Effect of MERIC on LaBS when executing the Lagoon test case. Significant energy savings of the compute nodes can be obtained with negligible performance loss thanks to use of dynamic tuning.

	Default HW settings	Static tuning	Dynamic tuning without runtime penalty	Dynamic tuning for maximum energy savings
Runtime [s]	1797	1942 (+8.1%)	1807 (+0.5%)	1871 (+4.1%)
Energy consumption [kJ]	3102	2631 (-15.1%)	2726 (-12.1%)	2496 (-19.5%)
Solver energy-efficiency [MLUPs/W]	0.054	0.059	0.056	0.056

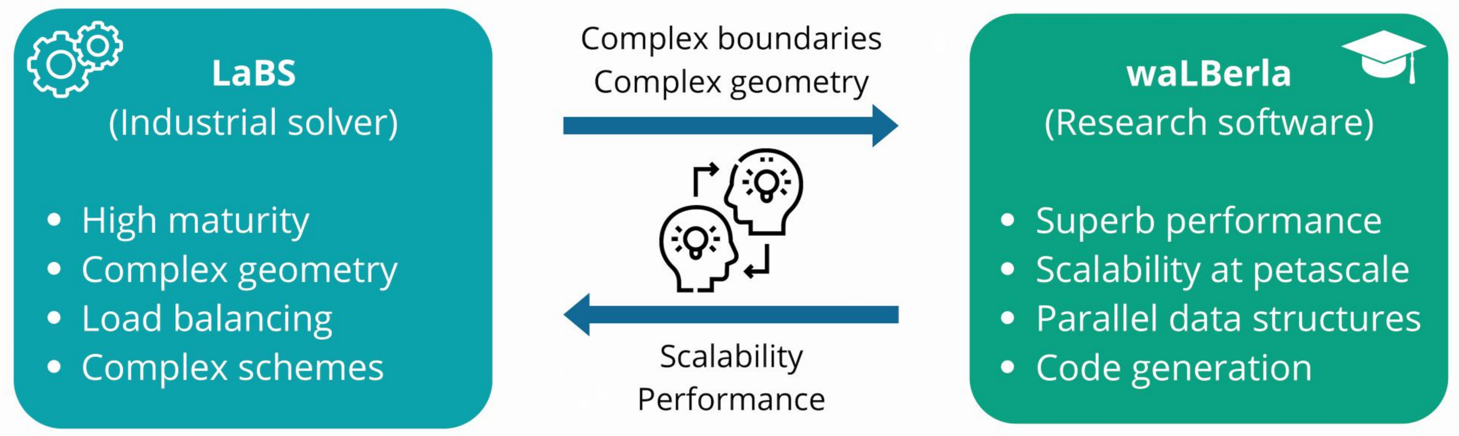
Static Tuning of GPU version of waLBerla

GPU version of waLBerla was evaluated using the Lagoon test case. Energy optimization was done by tuning the GPU streaming multiprocessor (SM) frequency.

In this case, tuning the CPU parameters did not bring any additional savings since the GPU is the main energy consumer.



EXCHANGE OF TECHNOLOGIES BETWEEN LABS AND WALBERLA



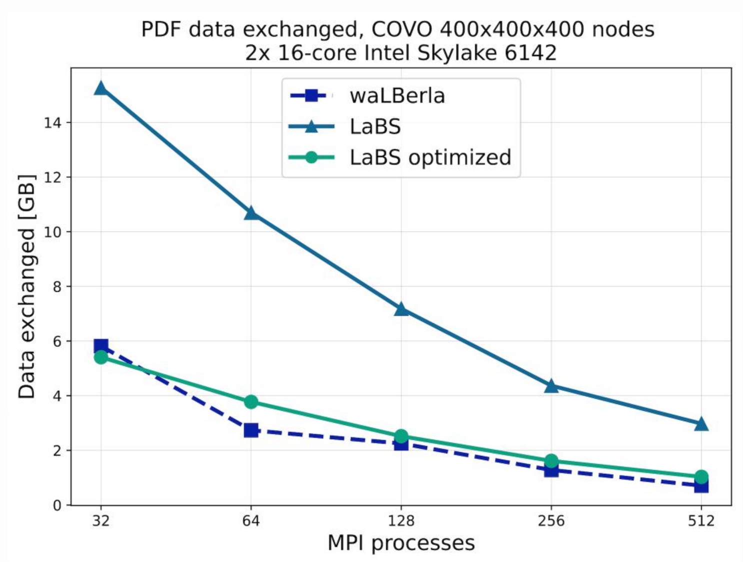
LaBS decided to develop a GPU version due to good performance results achieved by waLBerla. During this porting, with the experience of waLBerla we made two major optimizations.

The **merging of functions**. We observed that the less one calls the device during a time step, the more computations are accelerated on GPU. So, we defined a simplified scheme for the purpose of the GPU version which uses mainly two functions: a physics computation function, which sequentially calls two GPU kernels (*Propagate* + *Macroscopic*) and (*Gradient* + *Collide*).

The **reduction of memory footprint**. Among the quantities analyzed using NVIDIA profiler tools, one can denote the size of memory and number of registers used by each GPU thread, the occupancy percentage of the GPU card and the size of the grid used by the GPU kernels. We have observed that the grid size is automatically adjusted by NVC++ at runtime such that at least two parallel launches can be made for each kernel call. The parameter of which optimization improved the performances, and which depends on the kernel implementation was the size of memory used within the GPU threads.

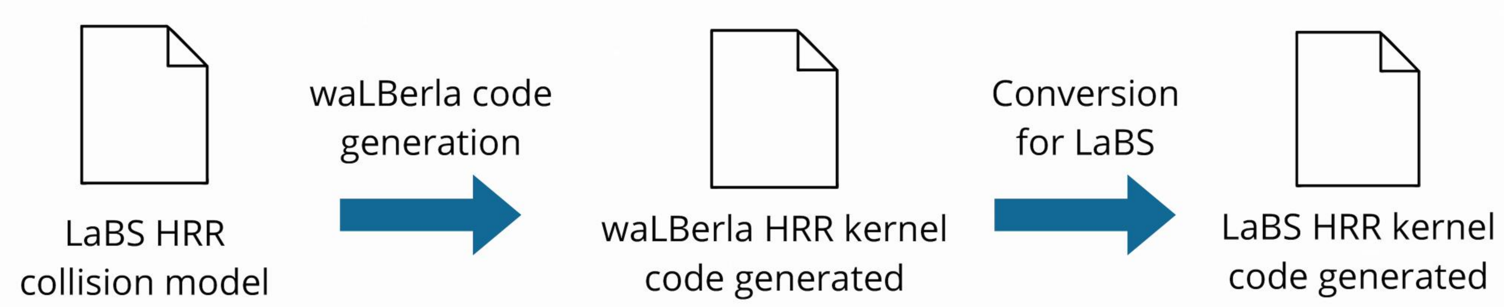
Mesh number of fluid nodes	Solver step performances in MLUPs on GPU			
	Merging functions		Memory footprint optimization	
	with	without	with	without
100x100x100 (1M)	27.8	204.06 (+734%)	290.89	558.04 (+192%)
300x300x300 (27M)	28.04	206.43 (+736%)	309.74	609.31 (+197%)

We also have deeply compared performances between LaBS and waLBerla on a simple test case and we noticed that LaBS exchange more data among MPI processes than waLBerla.



We have investigated the differences in data exchange amount, and it appears that LaBS was not **optimizing the particle distribution function (PDF)** data exchange. So, these optimizations have been added to LaBS and the figure on the left shows that the amount of data exchanged is of the same order of magnitude in LaBS and in WaLBerla now.

Moreover, we are currently trying to import some waLBerla optimizations through the code generation by injecting the LaBS collision model in the waLBerla code generation algorithm. The process is still in progress.

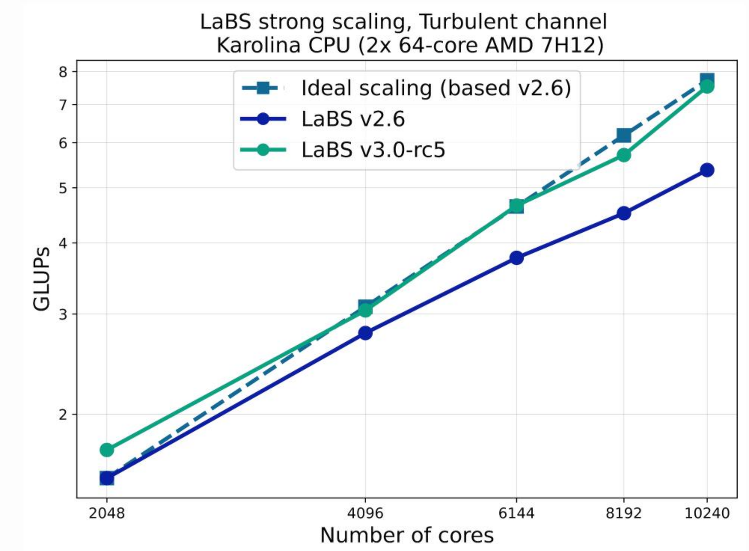


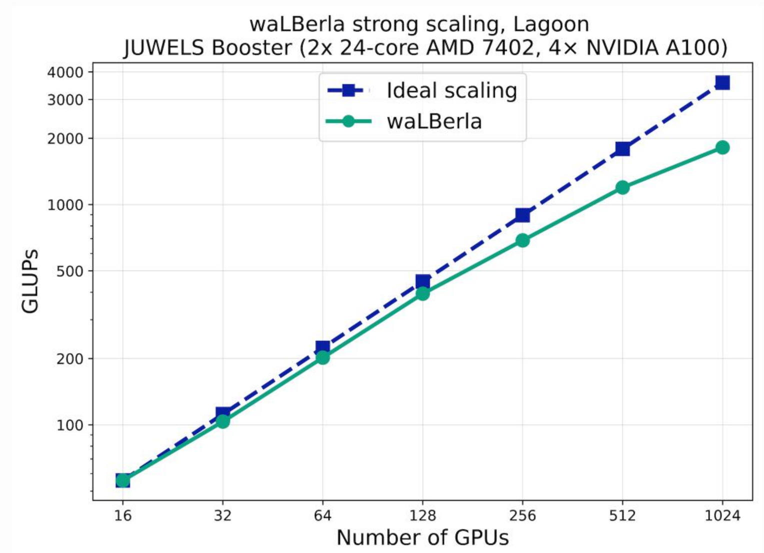
PERFORMANCE AND SCALABILITY RESULTS

LaBS strong scaling on KAROLINA

LaBS running the turbulent channel test case on up to 10,240 cores of the CPU partition of the KAROLINA supercomputer at IT4I.

On 10,240 CPU cores we are able to achieve 7500 MLUPs of performance.





waLBerla strong scaling on JUWELS Booster

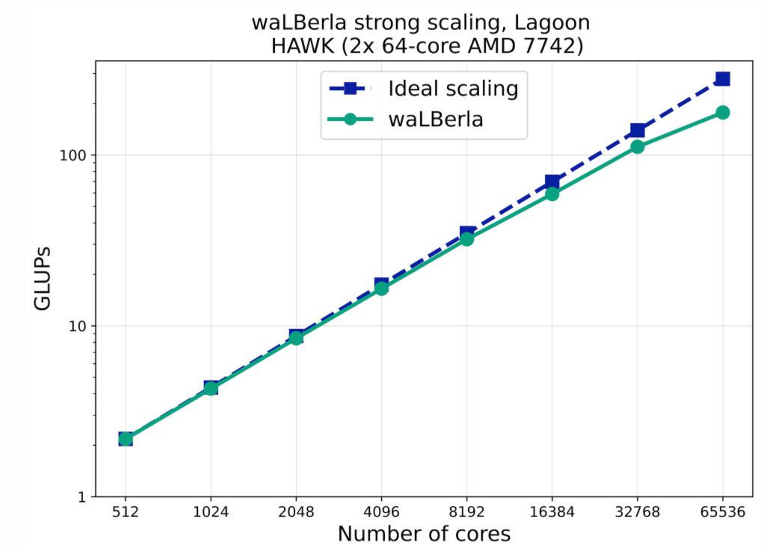
GPU version of waLBerla running the Lagoon use case on the Booster module of the JUWELS machine at JSC.

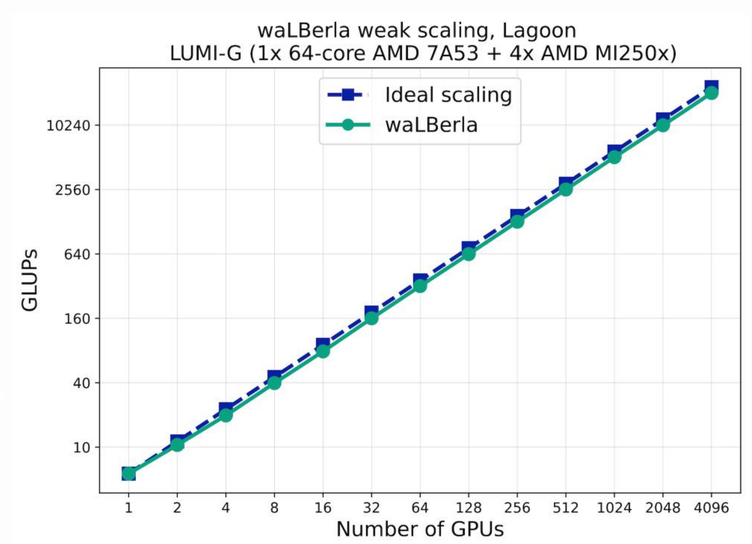
The test case contains approximately 277 million fluid nodes and can scale acceptably up to 1024 A100 GPUs, corresponding to approximately 200,000 fluid nodes per GPU.

waLBerla strong scaling on HAWK

CPU version of waLBerla executed Lagoon strong scaling benchmark on up to 512 compute nodes and 65,536 cores on the HAWK supercomputer at HLRS.

On 65,536 CPU cores, we are able to achieve 176,000 MLUPs of performance and a strong scalability of about 64%. This allows us to update 176 billion lattice cells per second.





In the SCALABLE project, waLBerla was ported to support AMD GPUs using the ROCm framework and the code generation pipeline for highly efficient compute kernels. This effort can be demonstrated firstly on the LUMI supercomputer using up to 4096 MI250x GPUs (2x GCD per GPU).

The results of a weak-scaling experiment on the Lagoon test case are shown in figure on the left using a uniform mesh with 512^3 cells on each GCD. On 4096 GPUs we reach a scaling efficiency of 88% using the AMD ROCm RDMA (GPU direct).

CONCLUSIONS

SCALABLE illustrates the synergy between academic and industrial application developers. Using academic codes as lighthouse applications to single out successful porting and optimization strategies leads to concrete day-to-day improvements for applied tools used in industry.

waLBerla continues to pave the way with the first results on the EuroHPC JU system LUMI-G using AMD MI250x GPUs showing excellent scaling up to 4096 GPUs for now.

This work will continue until the end of the project with a showcase application leveraging all improvements on these codes. Finally, using hardware tuning, we have improved the performance/energy consumption ratio by 10 to 20 % on CPU and GPU.

CS GROUP · AIRBUS · CERFACS · FAU Friedrich-Alexander-Universität Erlangen-Nürnberg · GROUPE RENAULT · JÜLICH Forschungszentrum · VSB Technical University of Ostrava IT4Innovations National Supercomputing Center · Neovia Innovation

EuroHPC Joint Undertaking · anr · Federal Ministry of Education and Research · Ministry of Education, Youth and Sports

The SCALABLE project (2021-2023) has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956000. The JU receives support from the European Union's Horizon 2020 research and innovation program and France, Germany, Czech Republic. This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (ID: MC2103) and through the e-INFRA CZ (ID:90254).