

HIGH PERFORMANCE LINEAR ALGEBRA ON GPUS

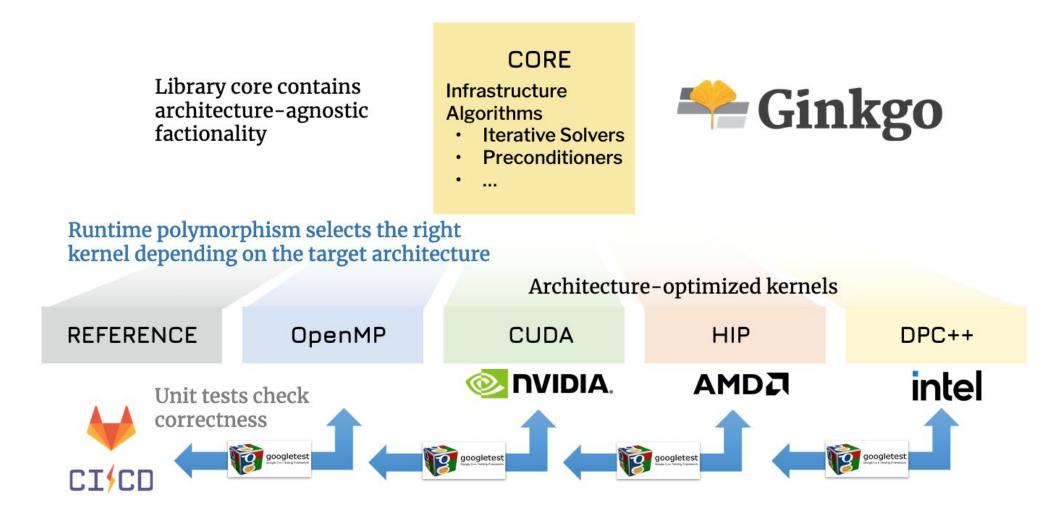
I. Aggarwal, H. Anzt, N. Beams, T. Cojean, Y. Chen, T. Grützmacher, F. Göbel, M. Koch, G. Olenik, P. Nayak, P. Nguyen, T. Ribizel, Y. Tsai

University of Tennessee, USA
Karlsruhe Institute of Technology, Germany

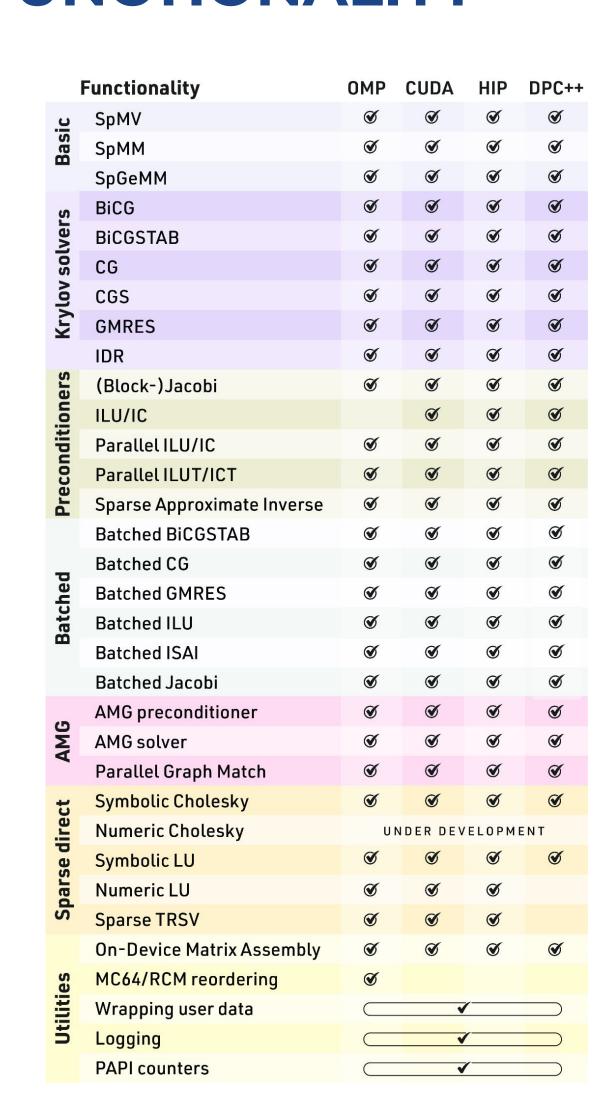


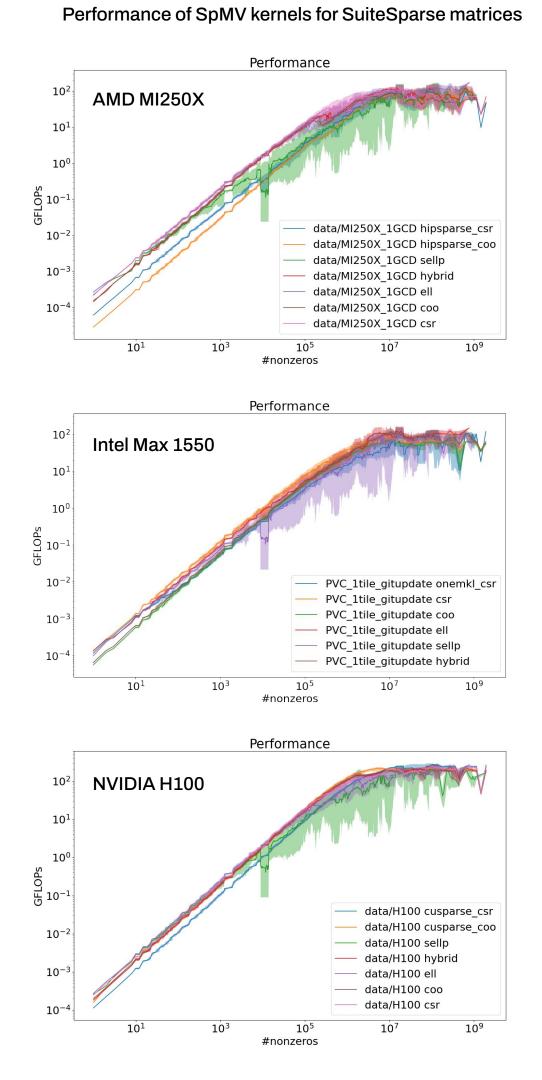
OVERVIEW

Ginkgo is a C++ framework for sparse linear algebra. Using a universal linear operator abstraction, Ginkgo provides basic building blocks like the sparse matrix vector product for a variety of matrix formats, iterative solvers, preconditioners, and batched routines. Ginkgo targets multi- and many-core systems, and currently features back-ends for AMD GPUs, Intel GPUs, NVIDIA GPUs, and OpenMP-supporting architectures. Core functionality is separated from hardware-specific kernels for easy extension to other architectures, with runtime polymorphism selecting the proper kernels.



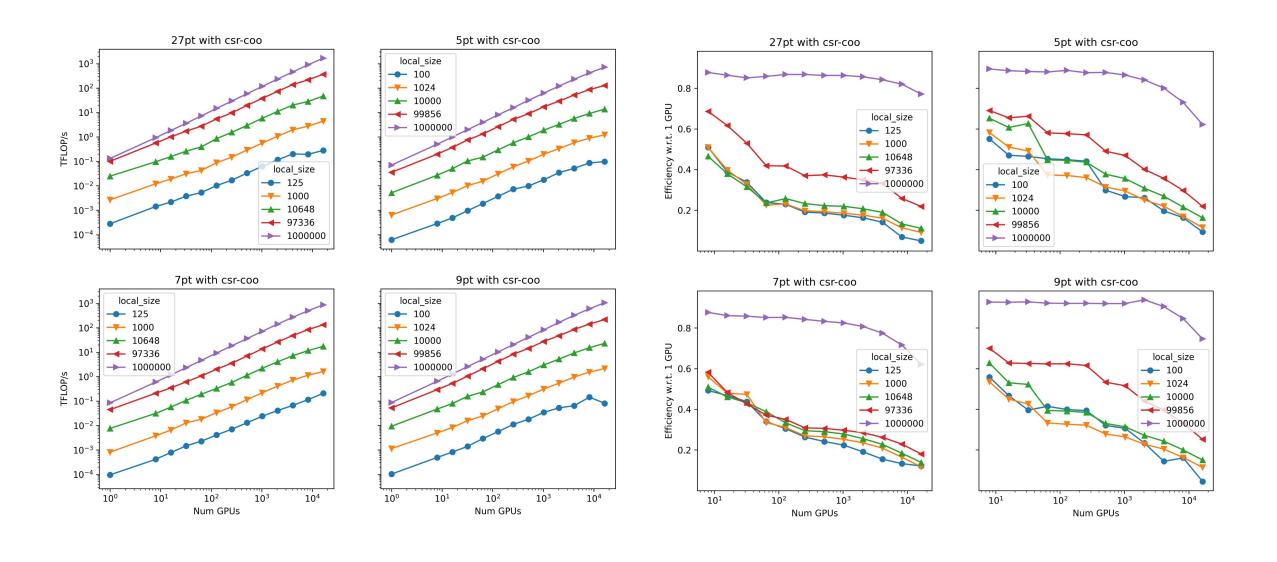
FUNCTIONALITY





FRONTIER SCALING

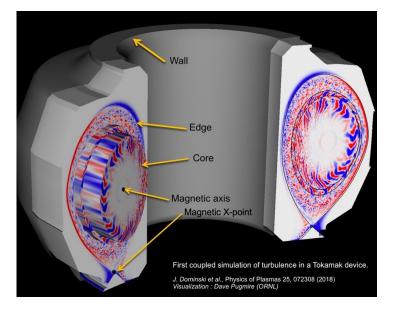
Ginkgo was able to run on the ECP Frontier Exascale machine (current TOP500 #1) and reach **up** to 2 PFlop/s on 20.000 GPUs with weak scaling, therefore reaching an efficiency of over 80% with respect to using 1 GPU.

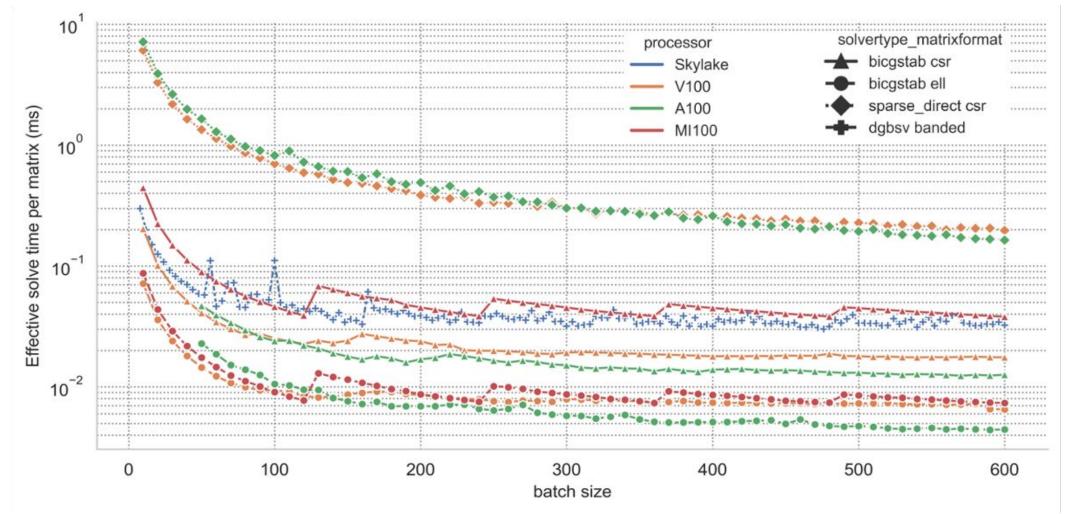


APPLICATION INTEGRATION

Batched Iterative Solvers in Plasma Simulations

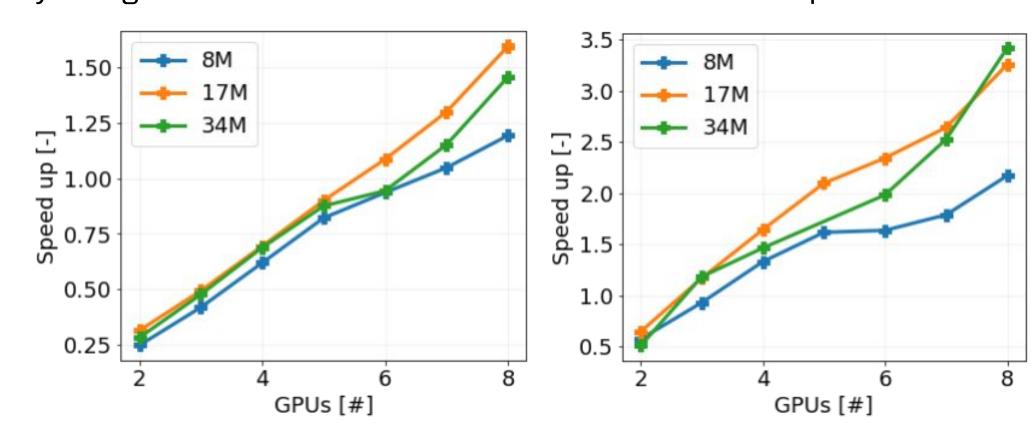
XGC is a gyrokinetic particle-in-cell code, which specializes in the simulation of the edge region of magnetically confined thermonuclear fusion plasma. Ginkgo is used to solve the velocity problems at different mesh nodes in a batched manner. https://xgc.pppl.gov/





Accelerating OpenFOAM Simulations

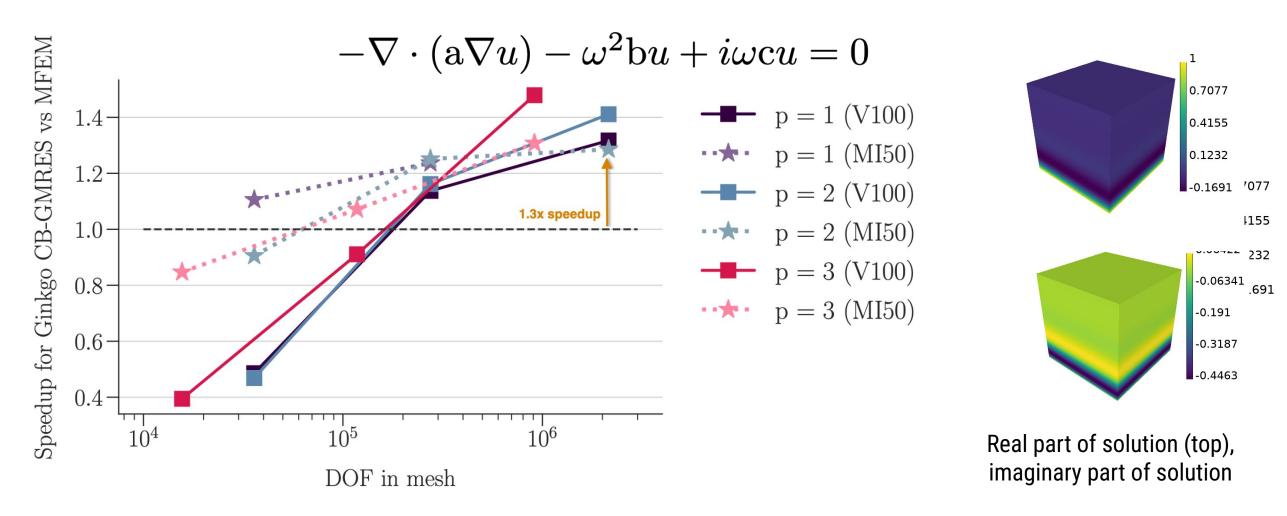
OpenFOAM is an open source CFD code heavily used for research and development in academia and industry. Ginkgo can be used as numerical backend to accelerate OpenFOAM simulations.



Speedup of a CG solver using a distributed Schwarz preconditioner with ISAI (left), and Multigrid (right) as local inner solver on multiple MI100s in comparison to the execution on 36 MPI ranks on an AMD EPYC 7302 with an IC preconditioner.

Speeding up MFEM's "example 22" on GPUs

Example 22 of the MFEM finite element library solves harmonic oscillation problems, with a forced oscillation imposed at the boundary. In this test, we use variant 1:



Speedup of Ginkgo's Compressed Basis-GMRES solver vs MFEM's GMRES solver for three different orders of basis functions (p), using MFEM matrix-free operators and the Ginkgo-MFEM integration wrappers in MFEM. CUDA 10.1/V100 and ROCm 4.0/MI50.

Stakeholder

















