# Kokkos: State of Experimental Backends

Rahul Gayatri, NERSC, LBNL Seyong Lee, ORNL

Kokkos User Group Meeting 2023

December 14, 2023

This manuscript has been authored by UT-Battelle, LLC, under contract DE-ACOS- 000R22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

#### Experimental backends

- OpenMPTarget and OpenACC are the current experimental backends supported.
- Not all features are implemented.
- Not all supported compilers and versions are tested and supported.
- Latest compiler versions might lead to feature or performance regressions. This might be less frequent for non-experimental backends.

December 14, 2023 2/14

- Uses target directives from OpenMP5.0 and above to offload Kokkos parallel patterns onto GPUs.
- Backend is supported on NVIDIA, AMD and Intel architectures
- Supports multiple compilers on a single architecture whenever possible
- Vendor compilers and clang on NVIDIA and AMD architecture
- Intel compiler on Intel architectures.

December 14, 2023 3/14

### Hierarchical Parallelism in OpenMPTarget backend

- ► No 3 level parallelism
- ► No SIMD support inside *target* region

December 14, 2023 4/14

### Unsupported features in OpenMPTarget backend

- No true L0 scratch support.
- ▶ Ilvm extensions that allow requesting scratch memory is now in *develop*
- No abort from inside kernel
- Minimum team size of 32 threads

December 14, 2023 5/1

#### Future Work for OpenMPTarget

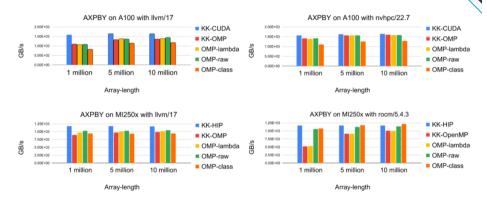
- Actively working with IIvm extensions to make OpenMP performant inside a target region.
- ▶ PR for using L0 on NVIDIA and AMD GPUs with OpenMPTarget is out
- Soon to be extend this to Intel GPUs.

December 14, 2023 6/14

#### OpenMPTarget Performance - AXPBY



#### **AXPBY Results (Higher is Better)**

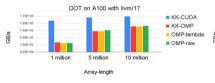


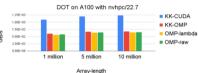
December 14, 2023 7/14

#### OpenMPTarget Performance - DOT

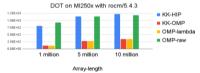


#### **DOT Results (Higher is Better)**









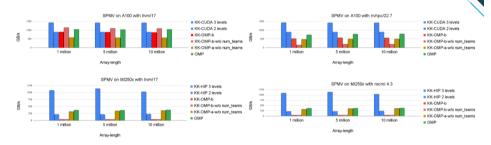
13

December 14, 2023 8/14

#### OpenMPTarget Performance - SPMV



#### **SPMV Results**



December 14, 2023 9/14

#### OpenACC backend

- Uses OpenACC parallel directives to offload Kokkos parallel patterns onto GPUs.
- Backend is supported on NVIDIA GPUs, AMD GPUs, and Intel/AMD/IBM CPUs.
- ▶ NVIDIA NVHPC compiler (nvc++) and CLACC compiler¹ for NVIDIA GPUs
- CLACC compiler for AMD GPUs
- ▶ NVIDIA NVHPC compiler (nvc++) compiler for INTEL/AMD/IBM CPUs

December 14, 2023 10/14

<sup>&</sup>lt;sup>1</sup>Open-source OpenACC compiler built on LLVM/OpenMP; available in the LLVM-DOE fork (https://github.com/llvm-doe-org/llvm-project/wiki)

#### Unsupported features in OpenACC backend

- No atomic operation support (available in the *develop* branch; will be added in the next release (V4.3))
- ► No scratch pad memory support
- Not-supported team-level APIs: team\_barrier(), team\_broadcast(), team\_reduce(), team\_scan(), etc.
- No abort from inside kernel
- No custom reduction support
- No SIMD support inside parallel region

December 14, 2023 11/14

## Future Work for OpenACC

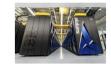
- ► Add atomic operation support
- Add custom reduction support
- Fully support hierarchial parallelism using CLACC as an OpenACC backend compiler

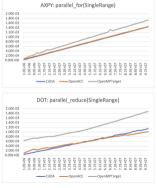
December 14, 2023 12/14

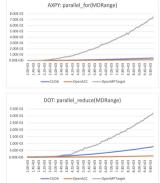
#### Performance Evaluation on GPUs

#### ORNL SUMMIT

- 1x NVIDIA Volta V100 GPU (16 GB)
- CUDA backend (CUDA 11.5.2)
- OpenMP Target backend (LLVM 17.0.0) OpenACC backend (NVHPC 22.11)









——CUDA ——OpenACC

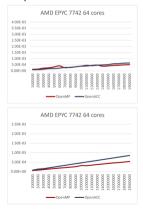
AXPY: parallel\_for(Team)

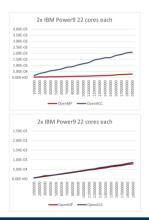
December 14, 2023 13/14

#### Performance Evaluation on CPUs

- · Intel, AMD, and IBM CPUs
- Mini-benchmarks (AXPY and DOT)







December 14, 2023