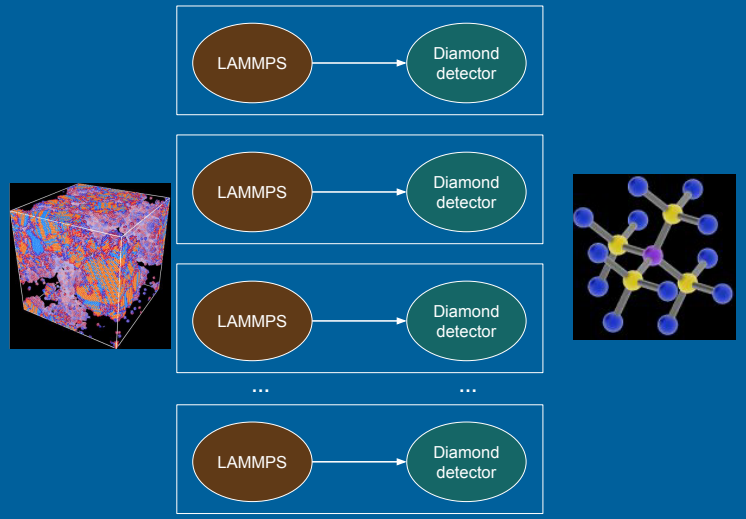
github.com/diatomic/LowFive
github.com/orcunyildiz/wilkins

BERKELEY LAB

Argonne NATIONAL LABORATORY

High-Performance Workflows: In Situ Data Transport and Workflow Management

Tom Peterka,	ANL
Orcun Yildiz	ANL
Bogdan Nicolae	ANL
Dmitriy Morozov	LBNL
Arnur Nigmetov	LBNL



"Somewhere, something incredible is waiting to be known."

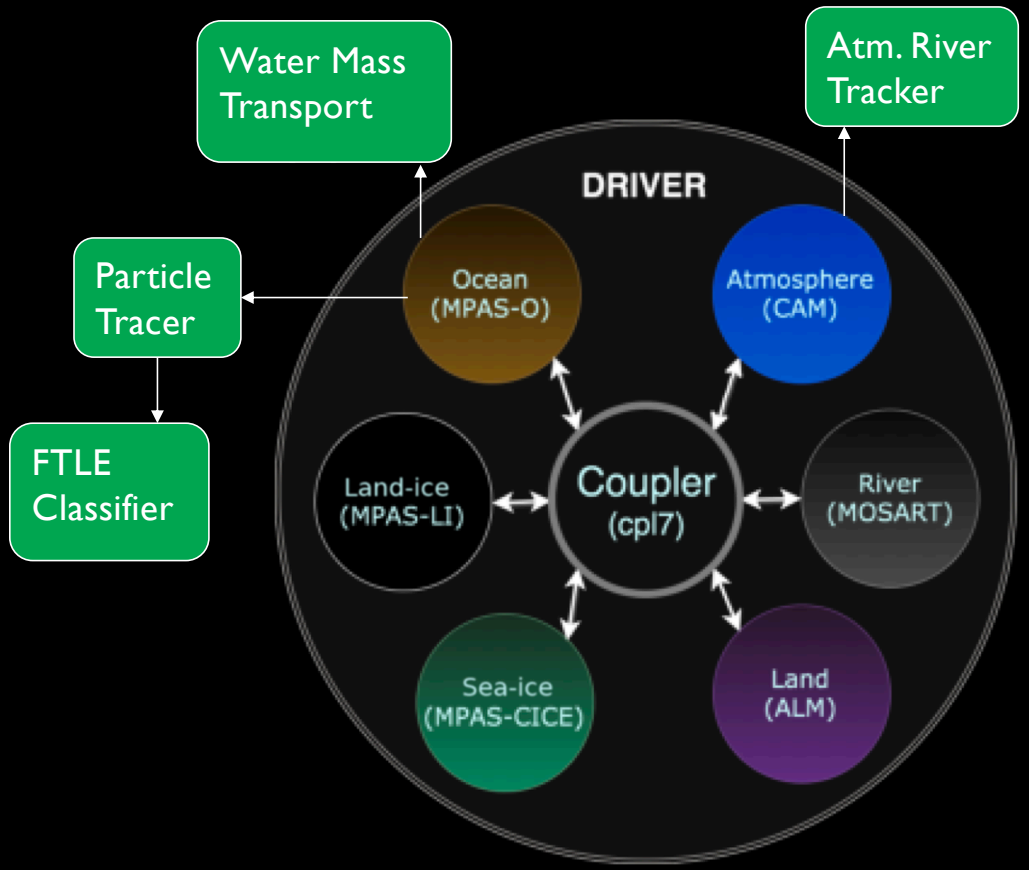
–Carl Sagan

E3SM Webinar
July 17 2025

Tom Peterka. Orcun Yildiz
tpeterka@mcs.anl.gov, oyildiz@anl.gov
Mathematics and Computer Science Division

Driving Problem

- Attach external analysis codes (diagnostics, AI, ML, visualization) to simulation codes
- Develop and build codes separately
- Run codes together in a single HPC job
- Communicate data efficiently between codes

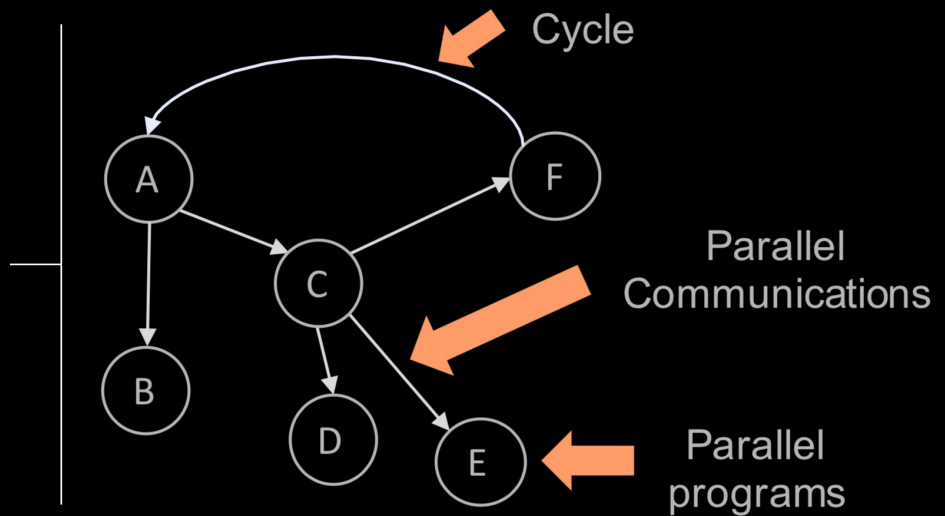


Original image from "Improving climate model coupling through a complete mesh representation: a case study with E3SM (v1) and MOAB (v5.x"). Vijay S. Mahadevan, Iulian Grindeanu, Robert Jacob, Jason Sarich. 2020

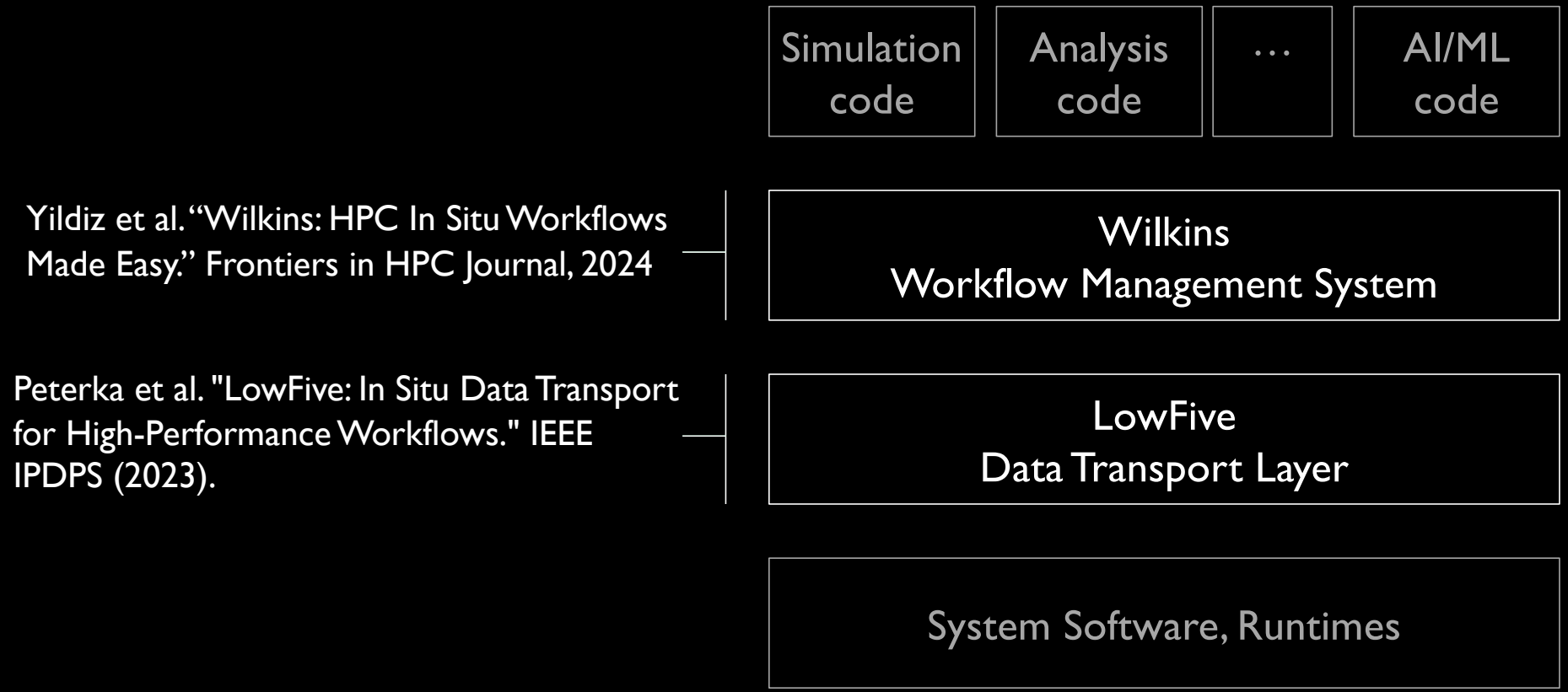
Definitions

- Workflow: concurrent computing tasks coordinating for a common objective.
- In situ HPC workflow: a workflow executed in a single job launch of an HPC system.
- Mental model: directed graph of vertices and edges, where vertices are tasks (programs) and edges are data dependencies (communication).

Logically, an in situ HPC workflow is modeled as a directed graph of tasks (parallel programs) and data dependencies (parallel communications). The graph does not need to be acyclic.



Software Stack



LowFive
Data Transport Library

github.com/diatomic/LowFive

Executive Summary

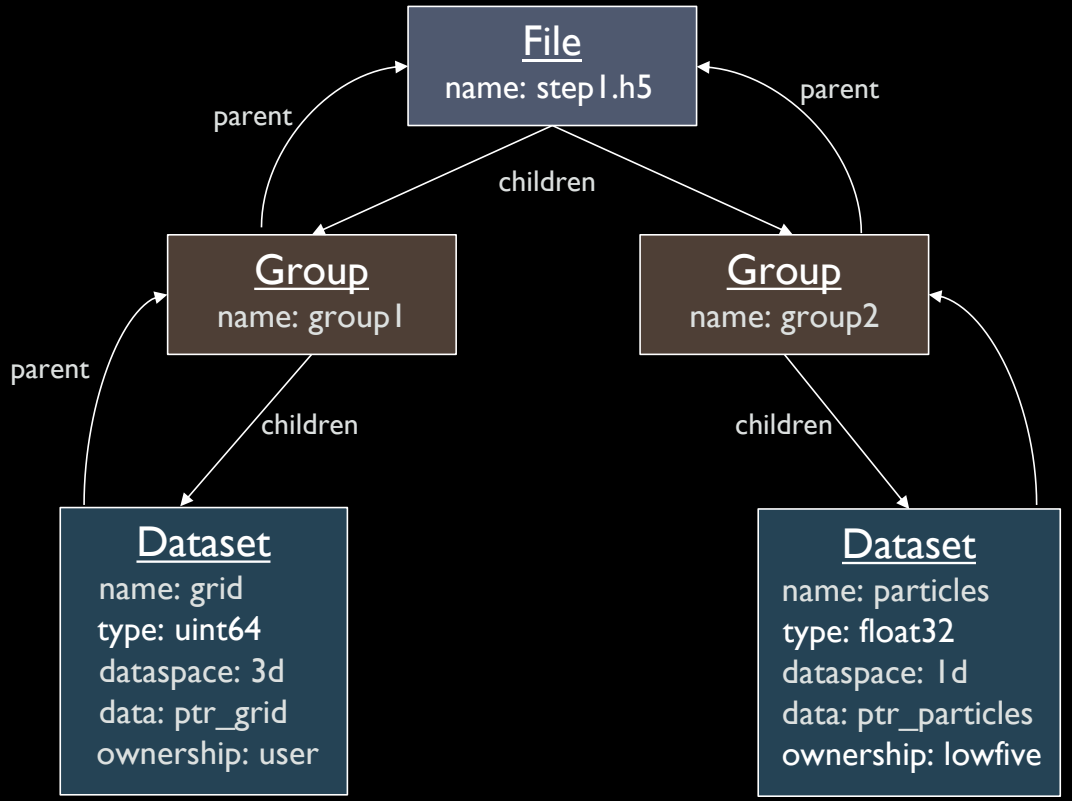
Objectives

- Move data efficiently in parallel between tasks (codes)
- No modification to task codes
- Redistribute data among different decompositions of tasks

LowFive

- In situ data transport layer for workflows
- HDF5 data model
- Built as an HDF5 VOL plugin
- Allows bypassing storage and sending data over MPI
- Redistributes data between producer and consumer tasks
- Standalone software library that workflow systems can use

LowFive Metadata Tree



Our own LowFive in-memory replica of HDF5 data model.

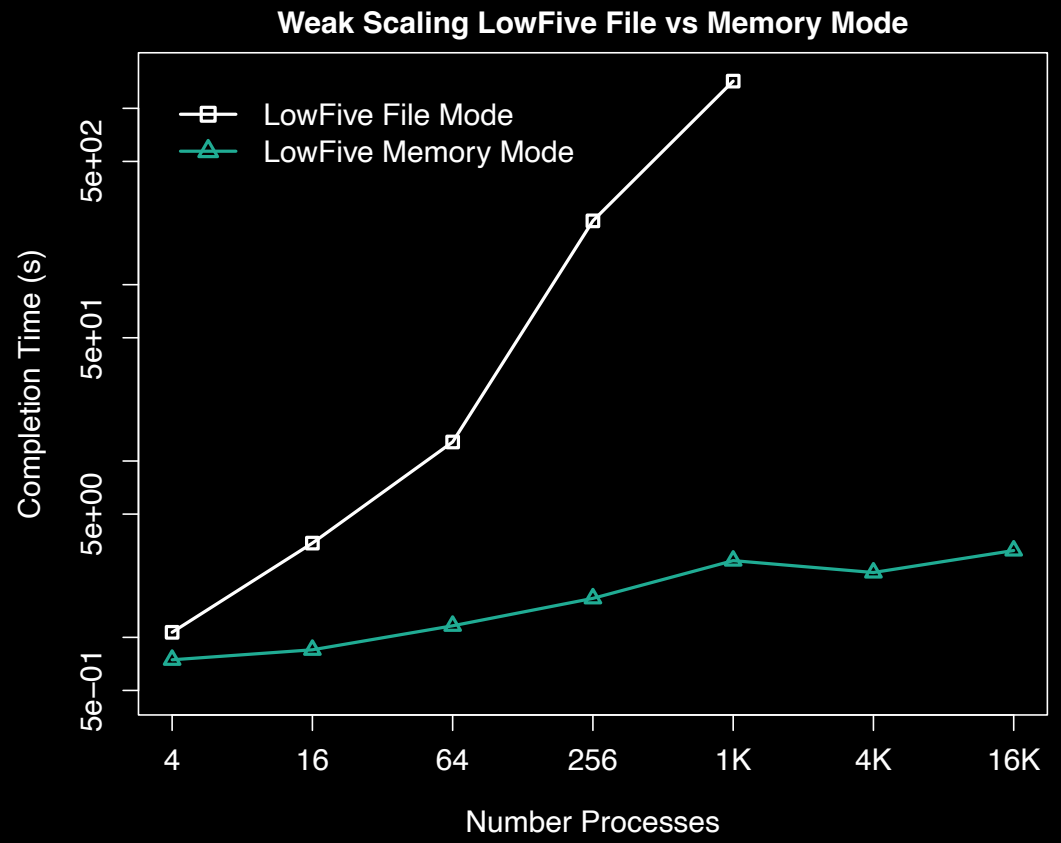
HDF5 Data Model

- Hierarchical data model much like a UNIX file system
- Root is the file
- Internal nodes are groups
- Leaves are datasets or other objects (e.g., attributes)

LowFive Data Model

- Our in-memory replica of HDF5 metadata
- One object for every HDF5 object
- Shallow or deep data pointer or copy

Synthetic Benchmarks: In Situ vs. Storage



Time to write/read grid and particles between 1 producer task and 1 consumer task, comparing LowFive file and memory modes, in a weak scaling regime.

Wilkins
Workflow Management System

github.com/orcunyildiz/wilkins

Executive Summary

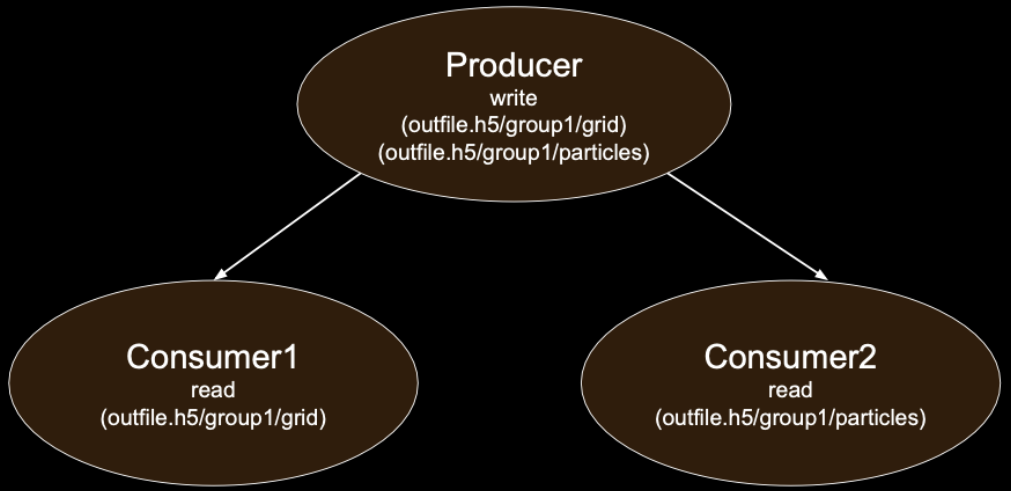
Objectives

- Easy to use workflow definition interface
- No modification to task codes
- Support various task graph topologies, fan-in, fan-out, ensembles, cycles
- Support stateful and stateless tasks
- Manage disparate execution rates of tasks

Wilkins

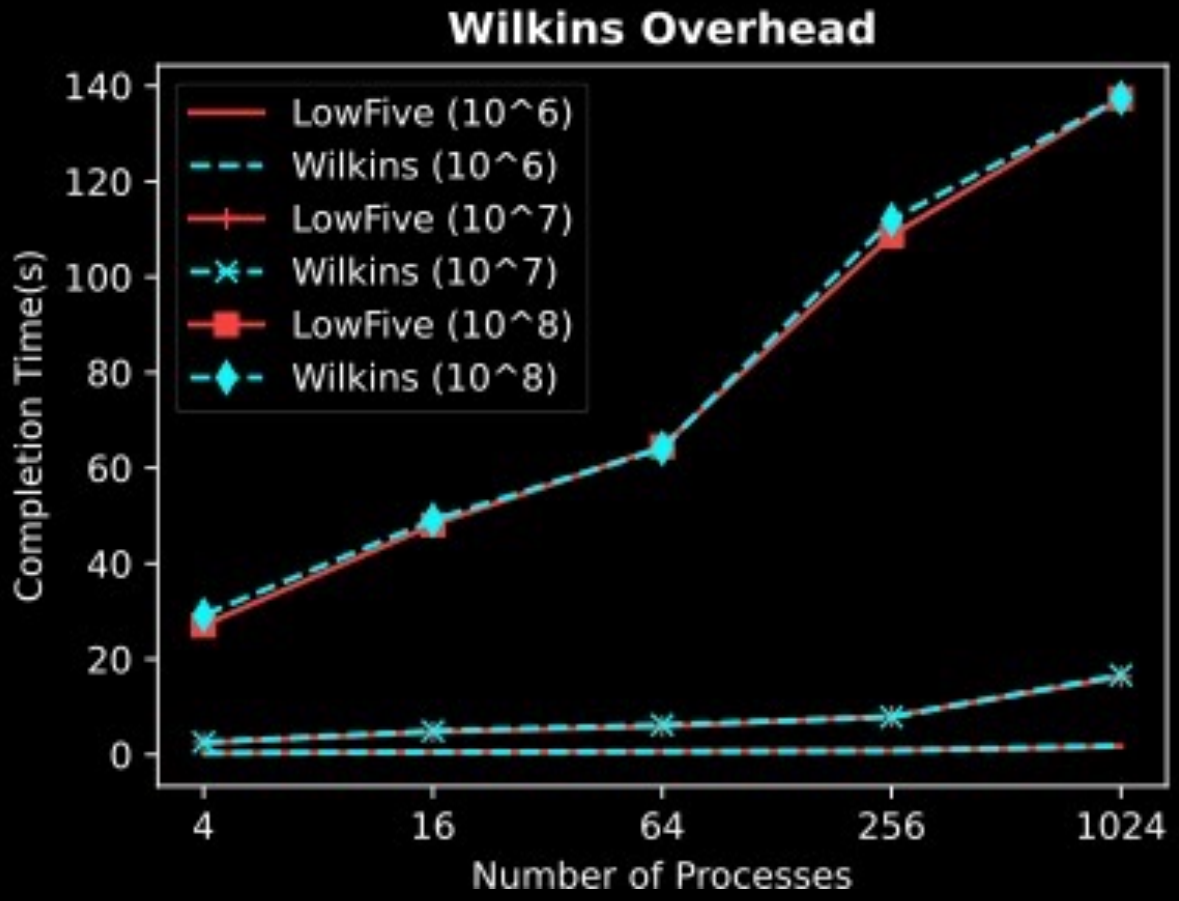
- In situ HPC workflow management system
- Built on LowFive
- Data-centric workflow definition
- Ensemble definition
- Flow control

Data-Centric Workflow Definition

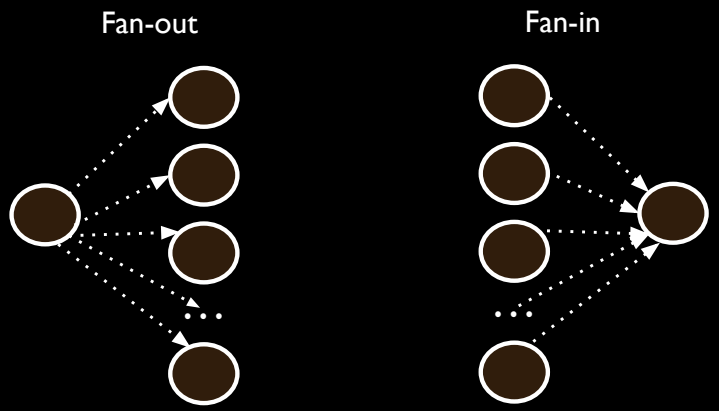


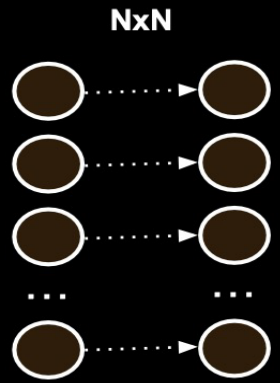

```
tasks:
  - func: producer
    nprocs: 3
    outports:
      - filename: outfile.h5
        dsets:
          - name: /group1/grid
            file: 0
            memory: 1
          - name: /group1/particles
            file: 0
            memory: 1
  - func: consumer1
    nprocs: 5
    inports:
      - filename: outfile.h5
        dsets:
          - name: /group1/grid
            file: 0
            memory: 1
  - func: consumer2
    nprocs: 2
    inports:
      - filename: outfile.h5
        dsets:
          - name: /group1/particles
            file: 0
            memory: 1
```

Synthetic Benchmarks: Overhead



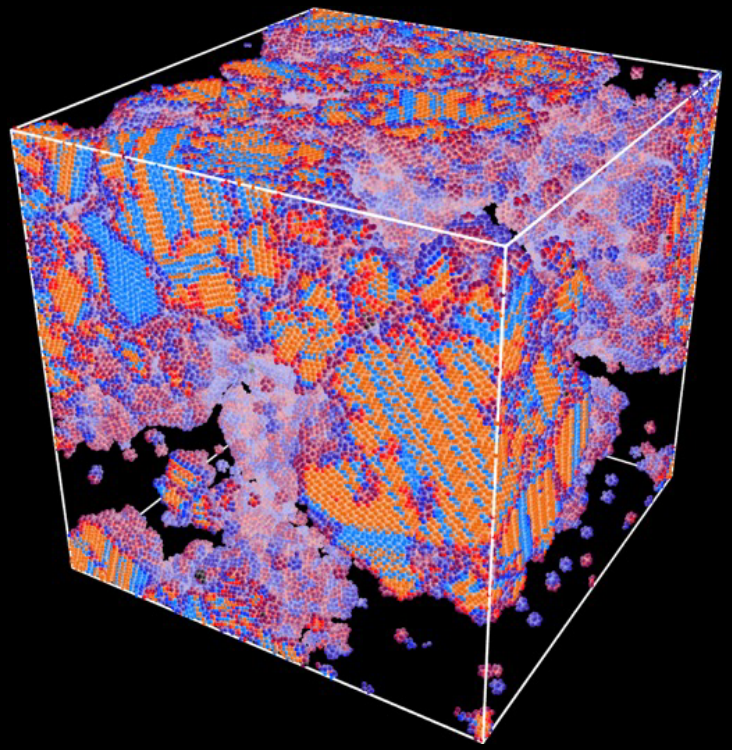
Ensemble Workflows

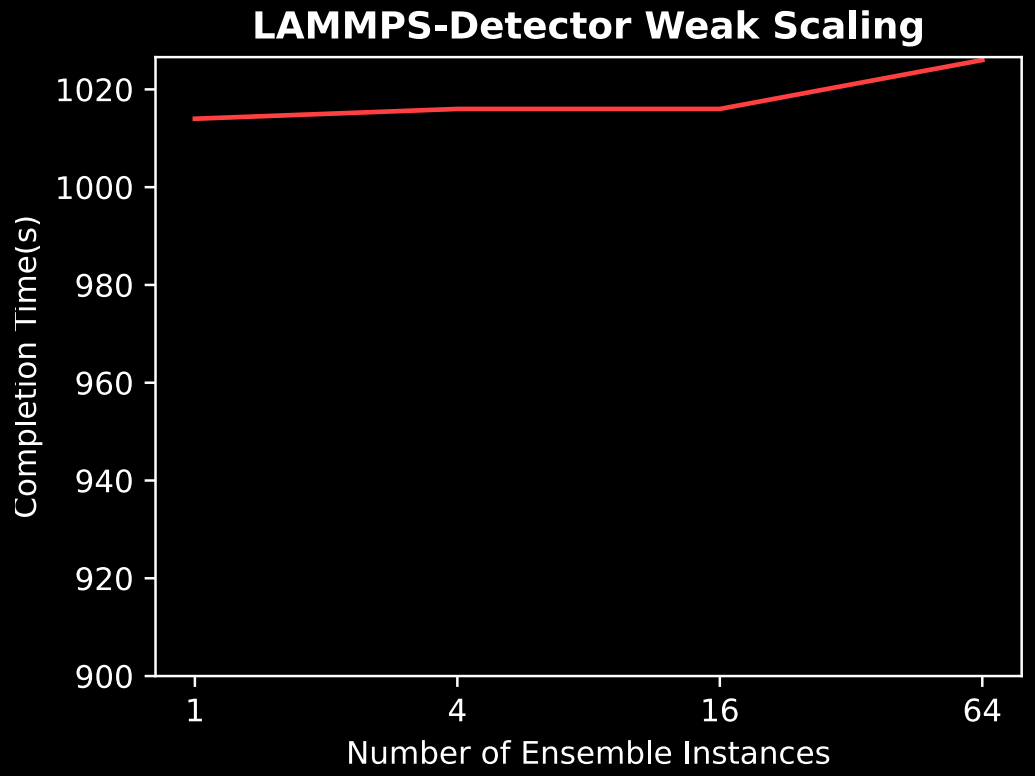





```
tasks:
  - func: producer
    taskCount: 4 #Only change needed to define ensembles
    nprocs: 3
    outports:
      - filename: outfile.h5
        dsets:
          - name: /group1/grid
            file: 0
            memory: 1
  - func: consumer
    taskCount: 2 #Only change needed to define ensembles
    nprocs: 5
    inports:
      - filename: outfile.h5
        dsets:
          - name: /group1/grid
            file: 0
            memory: 1
```

Science Use Case: Ensemble Instances in Molecular Dynamics





Application to E3SM

SOMA Test Case

- Understand sensitivity of pathline placement on large-scale ocean features in MPAS-Ocean
  - Input parameters: simulation duration, vector field output frequency, pathline seed locations
  - Output: double-gyre circulation in SOMA test case
- Workflow infrastructure for ocean modeling problem
  - Functioning Wilkins workflow including MPAS-O and FTK particle tracing

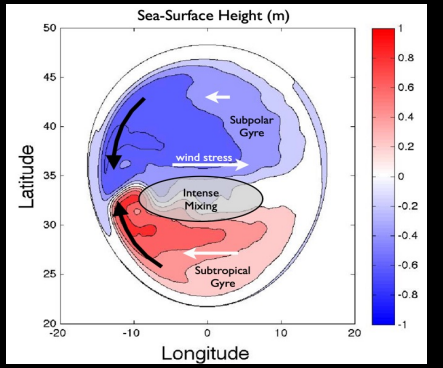
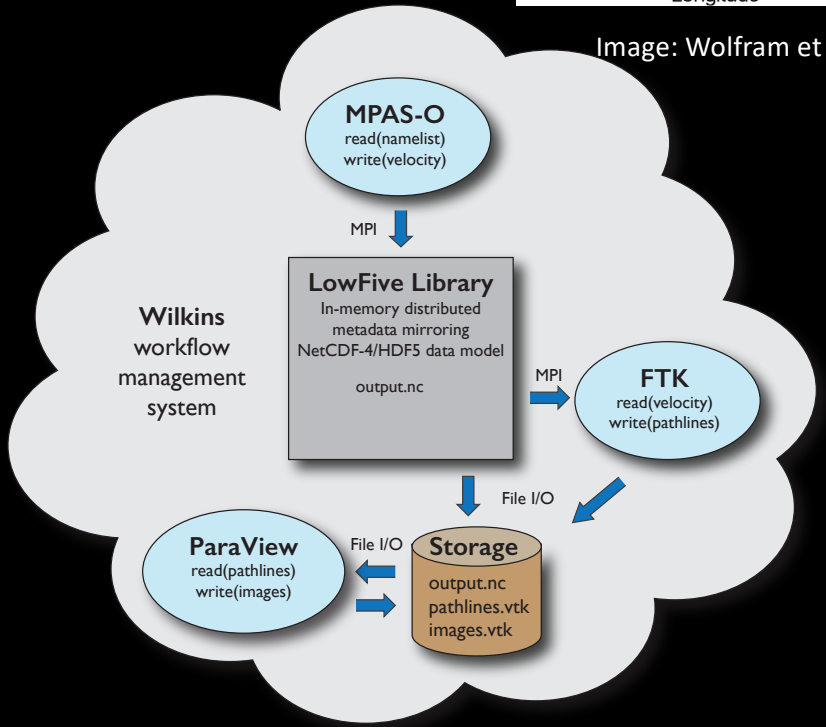
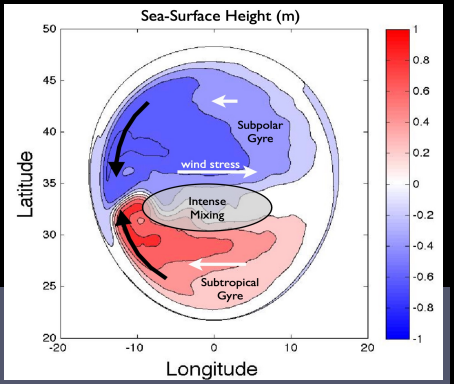


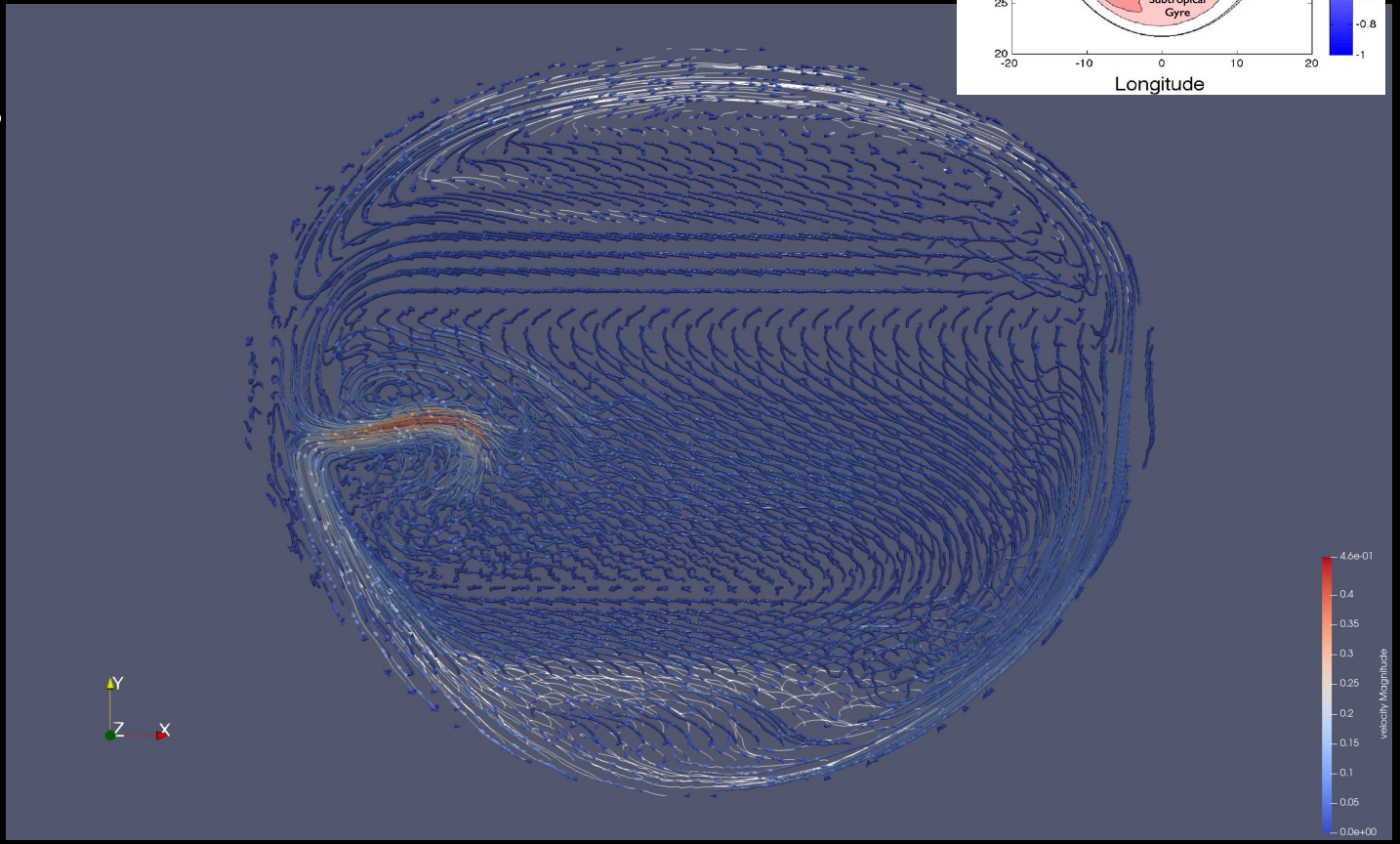
Image: Wolfram et al. 2015



SOMA Test Case

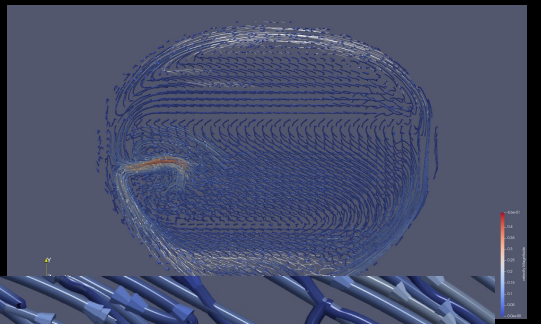
- Platform: Perlmutter
- 32 km grid resolution
- 1 year simulation time, outputs every 2 days
- Also ran 6 months duration output every day, 2 years output every 2 days, and 4 years output every 15 days
- Transfer data successfully through files, memory, or both
- Entire run takes ~2 hours for long pathlines as shown, shorter pathlines take ~few minutes
- 2500 pathlines, 50x50 grid of seeds

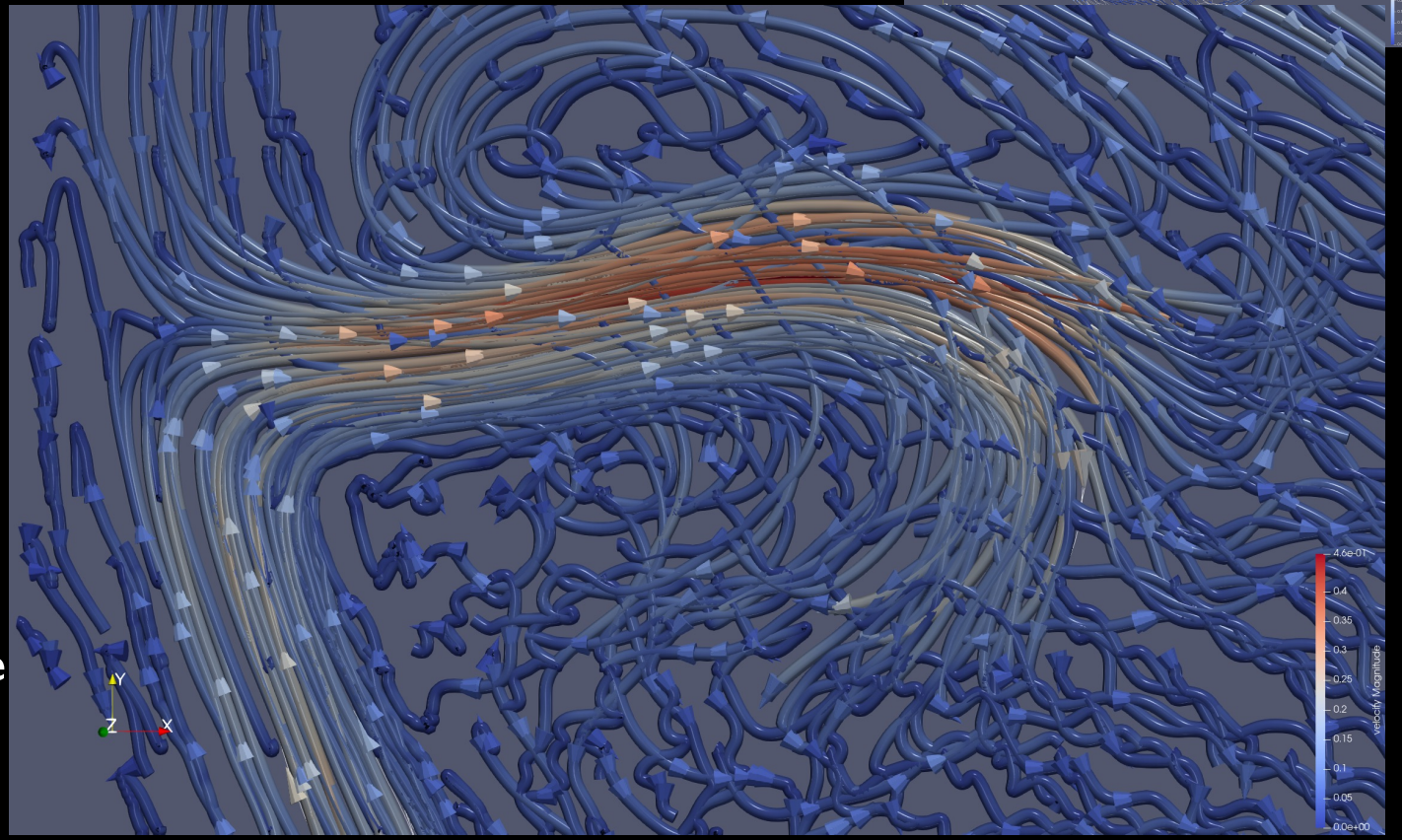




SOMA Test Case

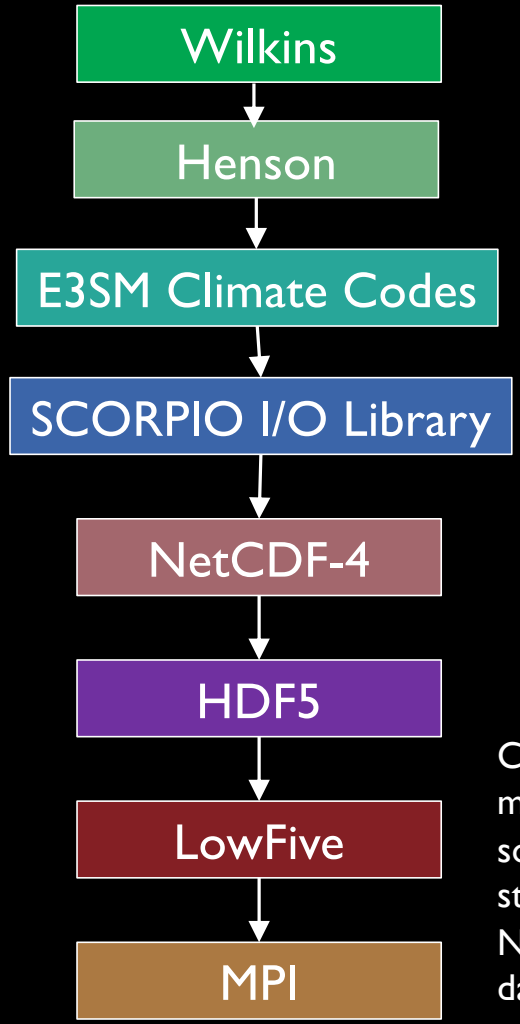
- Visualization parameters
  - Stream tubes
  - Lighting
  - Cone glyphs
  - Color by vel. mag.
- Science questions
  - Run duration
  - Output frequency
  - Seed placement
- Next steps
  - Scale up
  - Measure performance



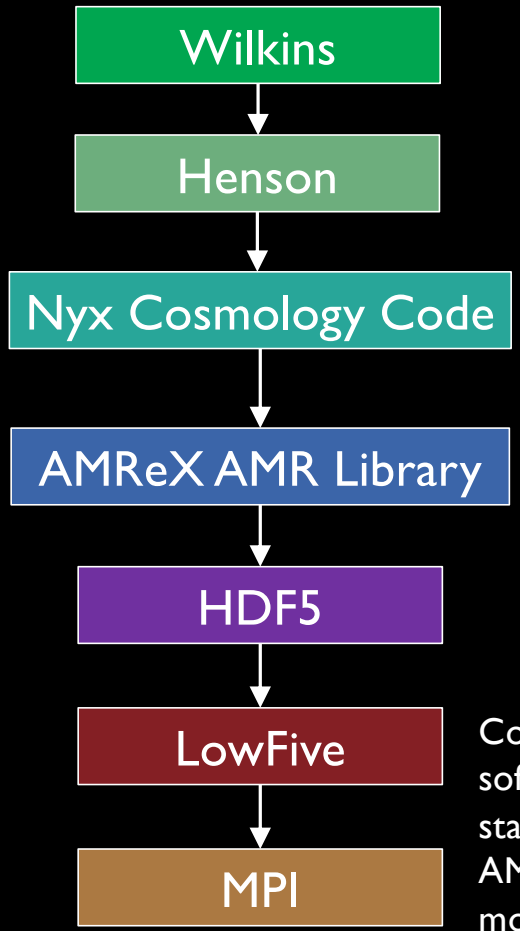


Summary

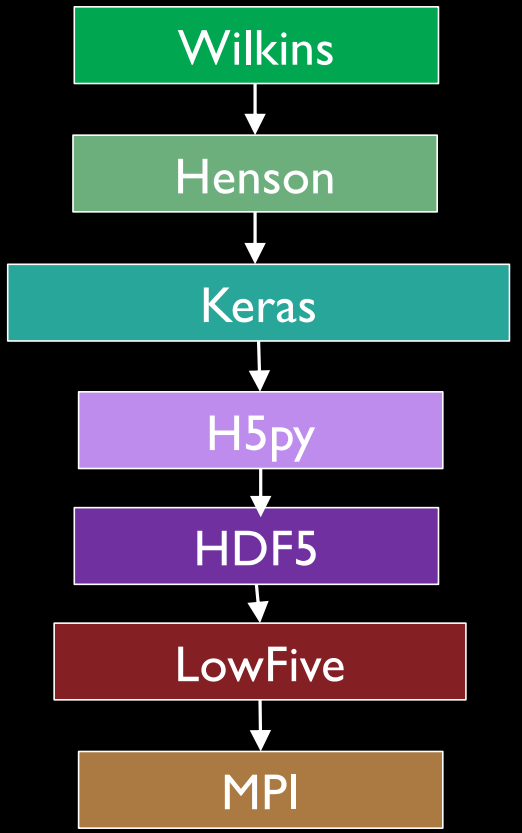
Use Cases and Deeper Software Stacks



Climate modeling software stack using NetCDF data model



Cosmology software stack using AMR data model



AI software stack using tensor data model

Recap

LowFive

- In situ data transport layer for workflows
- HDF5 data model
- Built as an HDF5 VOL plugin
- Allows bypassing storage and sending data over MPI
- Redistributes data between producer and consumer tasks
- Standalone software library that workflow systems can use

Wilkins

- In situ HPC workflow management system
- High-performance data transport from LowFive
- Data-centric workflow definition
- Ensemble definition
- Flow control
- Custom user-defined actions triggered by data operations

Next Steps

- Building entire E3SM software stack
- Adding more analysis codes (water mass transport)
- Profiling and benchmarking
- Eventual production use by science teams

github.com/diatomic/LowFive
github.com/orcunyildiz/wilkins

Acknowledgments

Facilities

Argonne Leadership Computing Facility (ALCF)
Argonne Laboratory Computing Resource Center (LCRC)
Oak Ridge Leadership Computing Facility (OLCF)
National Energy Research Scientific Computing Center (NERSC)

Funding

DOE ASCR Research Program
Margaret Lentz

People

Tom Peterka, Dmitriy Morozov, Arnur Nigmetov, Orcun Yildiz, Bogdan Nicolae

Tom Peterka. Orcun Yildiz
tpeterka@mcs.anl.gov, oyildiz@anl.gov
Mathematics and Computer Science Division