Six-fold Speedup on Ice Calving Detection Achieved by AMR-aware Parallel Connected Component Labeling

Xiaocheng (Chris) Zou, Surendra Byna, Hans Johansen, Daniel Martin, Nagiza F. Samatova, Arie Shoshani, John Wu
North Carolina State University and Lawrence Berkeley National Laboratory

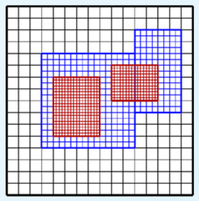
ABSTRACT

Ice calving event is a process of producing free-floating icebergs and ice fracture. Studying this event helps scientists to project global climate change. In this work, we present a **parallel *in situ* AMR-aware connected-component labeling** algorithm, which efficiently **detects real time ice calving event** in the AMR-based BISICLES simulation.

MOTIVATION

❖ Adaptive Mesh Refinement (AMR)

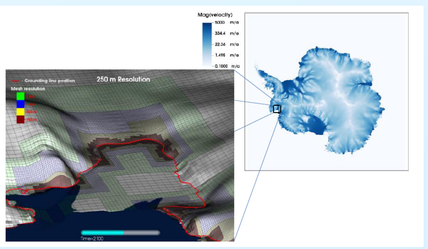
- Dynamically **refines logically-rectangular patches** in time and space dimensions
- **Improve efficiency** of computational resources while meeting desirable error levels
- **Hierarchical and complex** data structure



Sample AMR Meshes with 3 levels

❖ AMR-based BISICLES Simulation

- A **scalable AMR ice sheet modeling** application built on Chombo framework
- **Resolve dynamic features** like grounding lines and ice streams using very fine resolution
- Solves a **nonlinear coupled elliptic system** for the ice velocity field over the entire ice sheet/ice shelf system



Schematic showing computed ice velocity for Antarctic continent (right), and meshing for the Pine Island Glacier (left). The grounding line location is shown as red line.

❖ Ice Calving Event

- A process of producing **free-floating icebergs** and **ice fracture**
- **Large-scale calving events** (i.e., iceberg twice size of Atlanta breaks off of Antarctica) are highly interesting to scientists
- Impact global climate change

❖ Impact of Ice Calving Event

- Occasional ice calving events can result in disconnected portions of floating ice shelves, leading to an **ill-posed system** which causes the **solvers to diverge**

PROBLEM STATEMENT : How to efficiently detect ice calving events in an extreme scale simulation run

CHALLENGES

- The **hierarchical, multiresolution structure of AMR** renders existing connected component labeling techniques ineffective
- **Complex *in situ* data distribution** poses challenges of scalability

OBJECTIVES

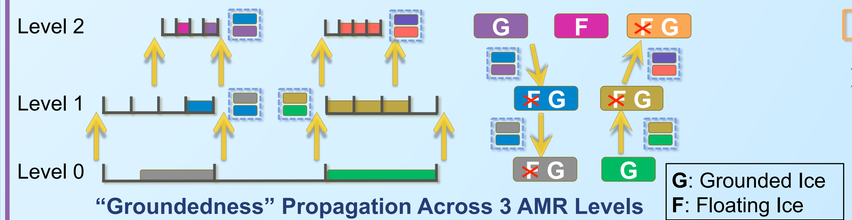
- Detect the **connected components** from AMR structure
- Identify "**Groundedness**" of floating ice (connectedness to land)
- Achieve **high scalability** in the **real simulation run**

TECHNICAL APPROACH

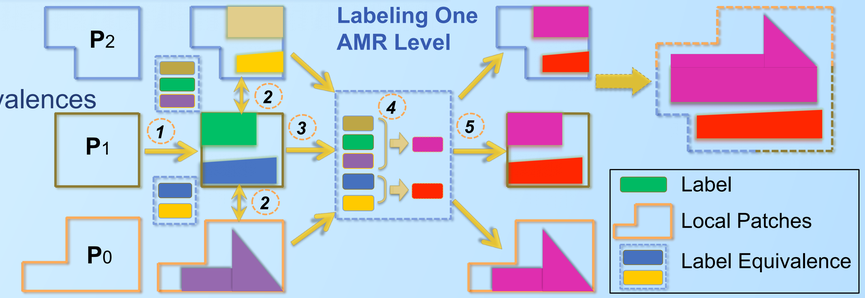
Part A: AMR-aware Connected Component Labeling

A.1 Multiple stages of labeling one AMR level

1. Discover local connectivity: assign local labels, record label equivalences
2. Exchange labels on boundary and collect expanded equivalence
3. Aggregate label equivalences to the selected processor
4. Determine final labels using aggregated label equivalences
5. Disseminate final labels to each processor



"Groundedness" Propagation Across 3 AMR Levels

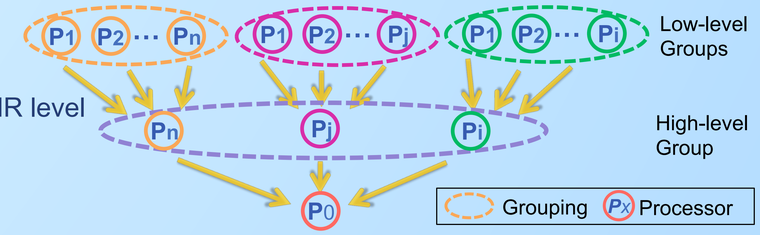


A.2 Propagate component's "Groundedness" across AMR levels

1. Collect label equivalence between AMR levels using Chombo
2. Propagate "Groundedness" along the label equivalence chain to determine the "Groundedness" of the whole component

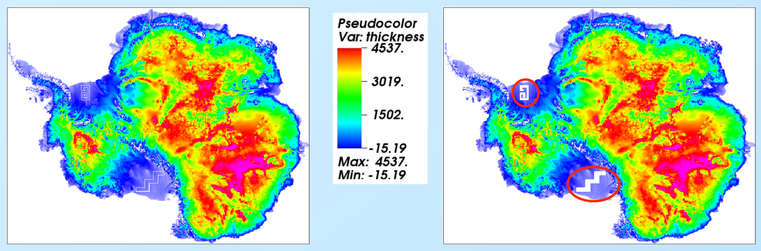
Part B: Hierarchical Data Aggregation

- Divide entire processors into two-level computing groups:
  Low-Level group: a group of processors with data from the same AMR level
  High-level group: selected processors from low-level groups
- Collect label equivalences and "Groundedness" info in parallel
- Avoid expensive global data communication overhead



PRELIMINARY RESULTS

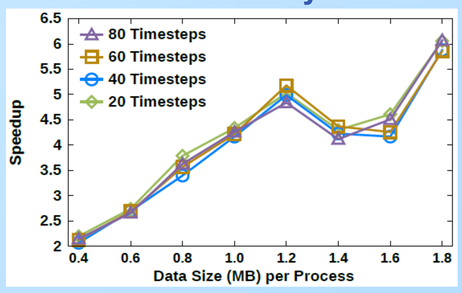
❖ Detected Disconnected Floating Ice



Ice thickness in the Antarctica AMR computation with artificially-cut ice (spiral and stair-like shape)

Ice thickness in the Antarctica AMR computation with detected ice (red circles)

❖ Performance Analysis



1. As the data size per process increases, the relative speedup of our method increases.
2. Slight dip: AMR boxes are distributed in a more scattered fashion, leading to slightly increased communication.
3. at a fixed data size per process, PCCL yields steady speedups (up to 6x) for more consecutive timesteps.

Contact: Arie Shoshani < shoshani@lbl.gov >

Zou, et al. Parallel In Situ Detection of Connected Components in Adaptive Mesh Refinement Data. CCGrid 2015. DOI:10.1109/CCGrid2015.154

SDAV Scalable Data Management, Analysis, and Visualization

NC STATE UNIVERSITY

BERKELEY LAB Lawrence Berkeley National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science