Scaling Geoscience Applications on Sunway Supercomputer

Lin Gan

Assistant Professor, Tsinghua University, Beijing

Assistant Director, NSCC-Wuxi, Jiangsu

Sep. 10, 2019, PPAM 2019 @ Bialystok, Poland

Location of Wuxi

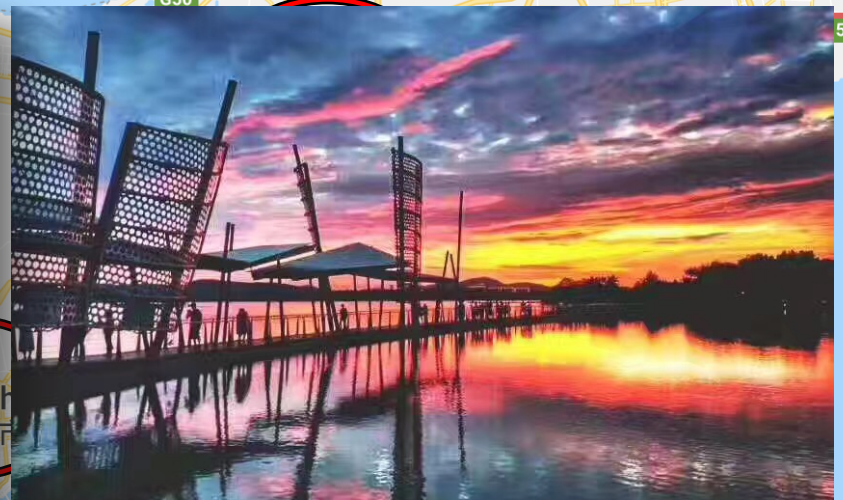
219 systems of the TOP 500

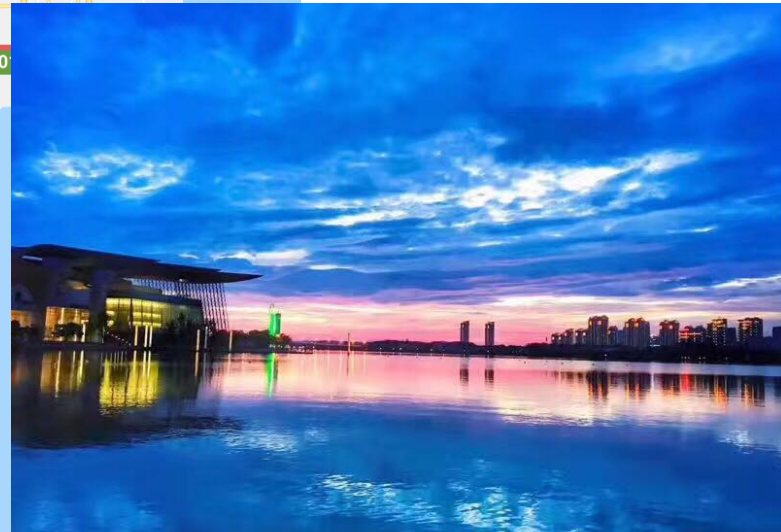
7 national supercomputing centers

Sunway BlueLight

Location of Wuxi

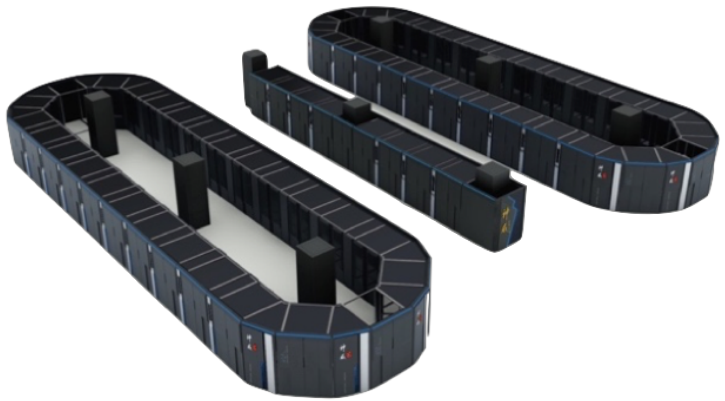
Taihu Lake





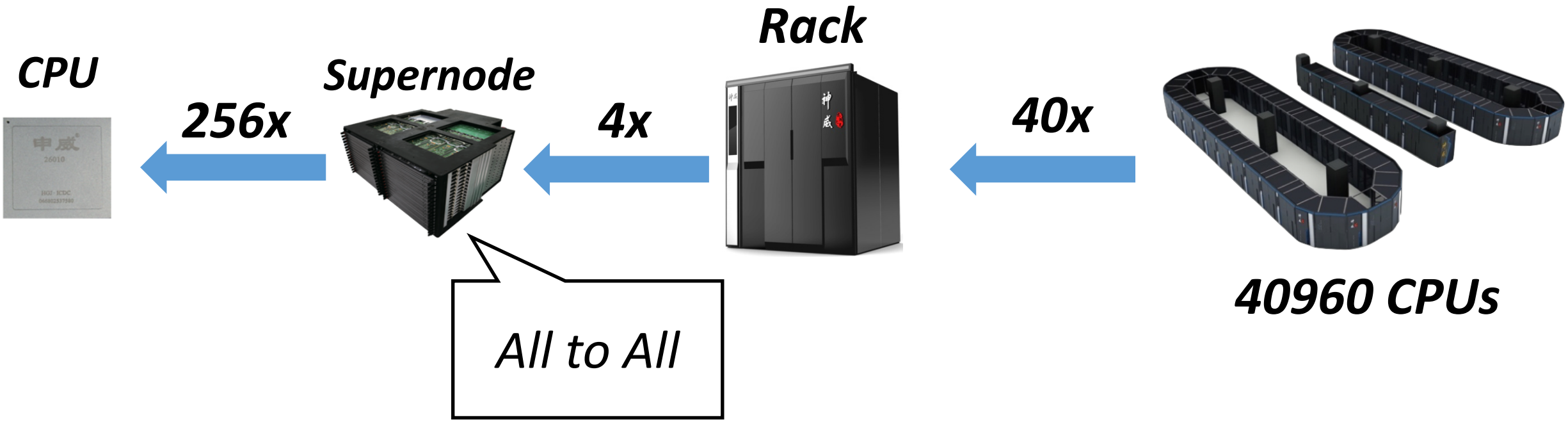
国家超级计算无锡中心
National Supercomputing Center in Wuxi

- ***National Supercomputing Center in Wuxi***
  - *Sponsored by MOST and state Gov.*
  - *Operated by Tsinghua University*
  - *Goal: world-leading SC center for important academic and industrial applications*

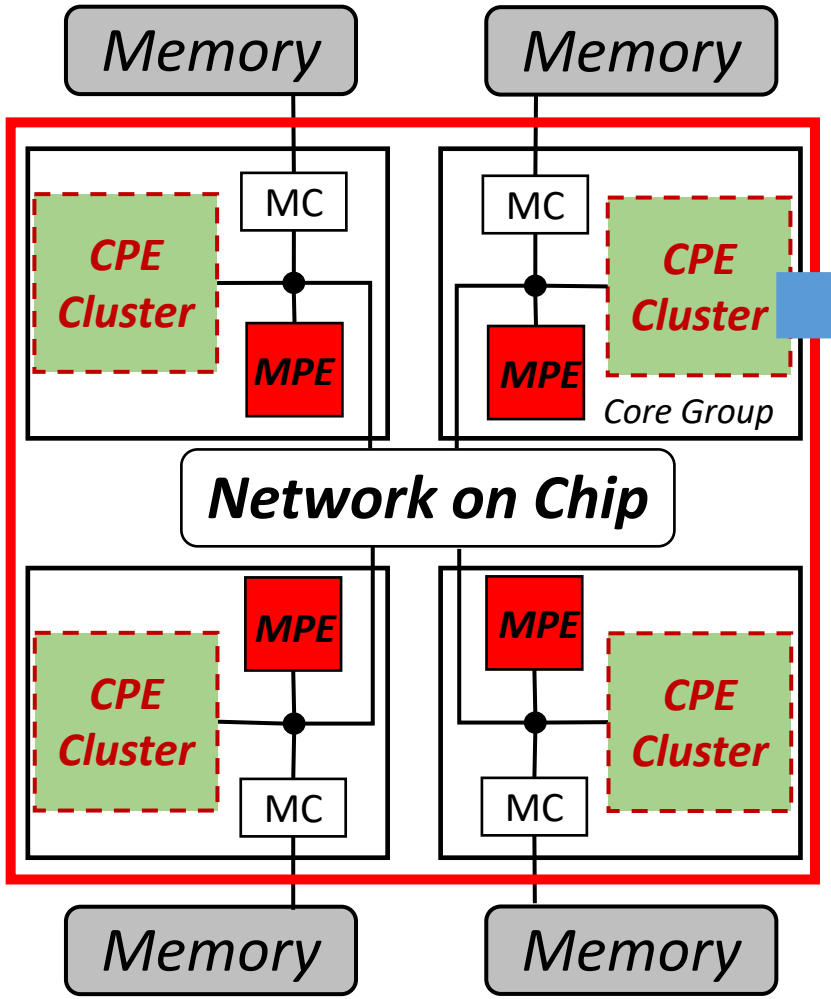


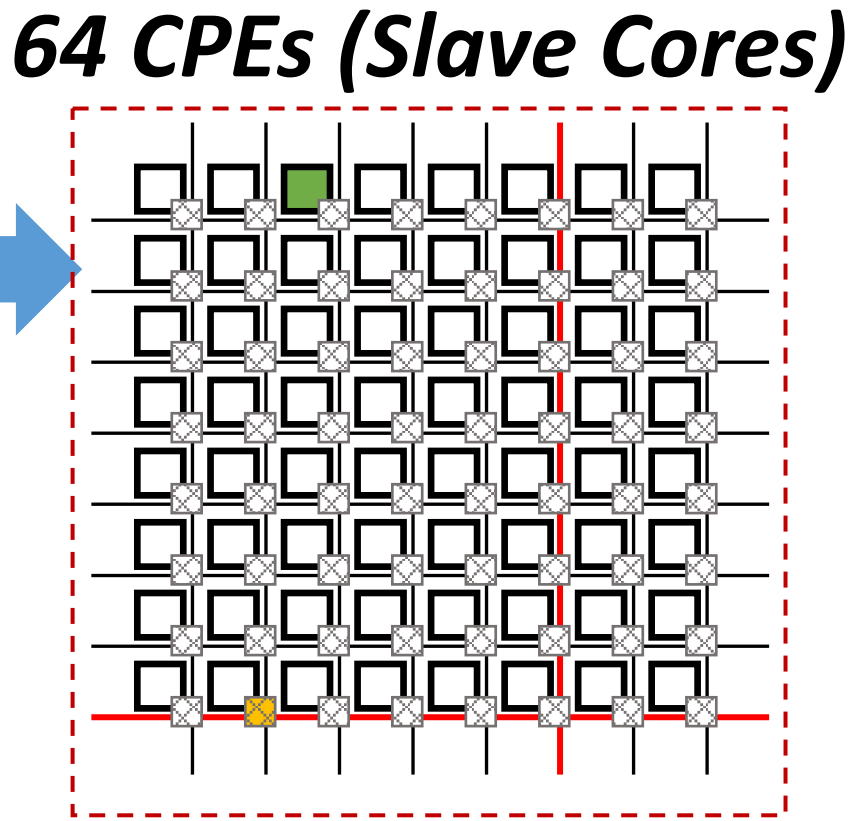
- ***Sunway TaihuLight Supercomputer***
  - *125 PFlops peak*
  - *93 PFlops LINPACK (No.3 in TOP 500)*
  - *6.05 GFlops/Watt*
  - *40, 960 homegrown 260-core CPUs*

System Hierarchy



SW26010 CPU


Memory
Memory
MC
MC
CPE Cluster
CPE Cluster
MPE
MPE
Core Group
Network on Chip
MPE
MPE
CPE Cluster
CPE Cluster
MC
MC
Memory
Memory

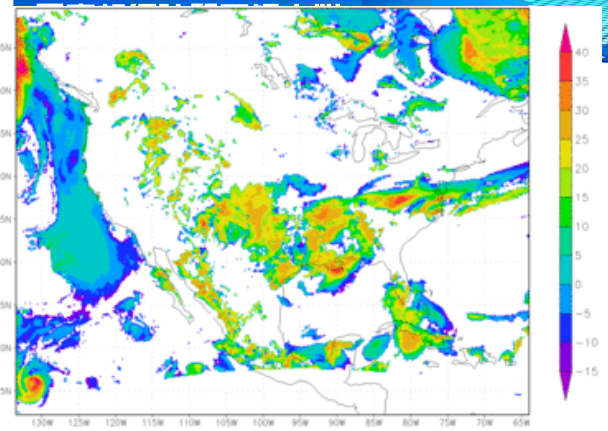

64 CPEs (Slave Cores)

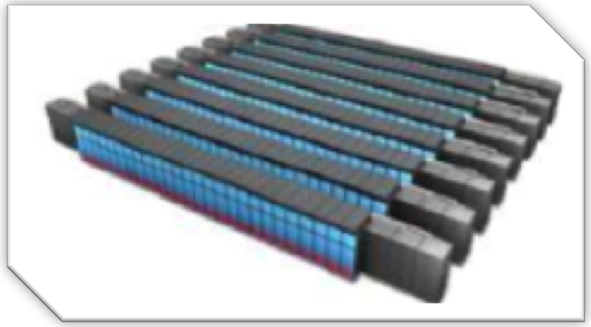
Application Overview

Since 2016, over 200 large-scale applications from over 300 research institutes covering 19 application domains, 22 full-scale applications, 40 half-scale, 100 million-core-scale, 6 Gordon Bell Finalists, and 2 Gordon Bell Prizes.


ACM
Gordon Bell Prize
2017
presented to
Haohuan Fu, Zekun Yin,
Tingjian Zhang, Wanwang Yin,
Conghui He, Zhenguo Zhang,
Wei Xue, Guangwen Yang,
Bingwei Chen, Wenqiang Zhang,
Weiguo Liu, Xiaofei Chen
For 18.9-Pflops Nonlinear Earthquake Simulation on Sunway TaihuLight: Enabling Depiction of 18-Hz and 8-Meter Scenarios


ACM
Gordon Bell Prize
2016
presented to
Chao Yang, Wei Xue, Haohuan Fu,
Hongtao You, Xinliang Wang,
Yulong Ao, Fangfang Liu, Lin Gan,
Ping Xu, Lanning Wang,
Guangwen Yang, Weimin Zheng
FOR 10M-Core Scalable Fully-Implicit Solver for Nonhydrostatic Atmospheric Dynamics





Climate Modeling

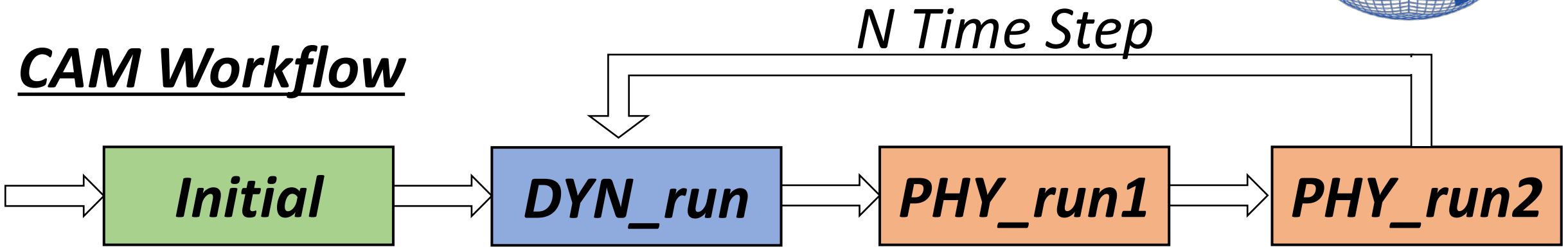
surface

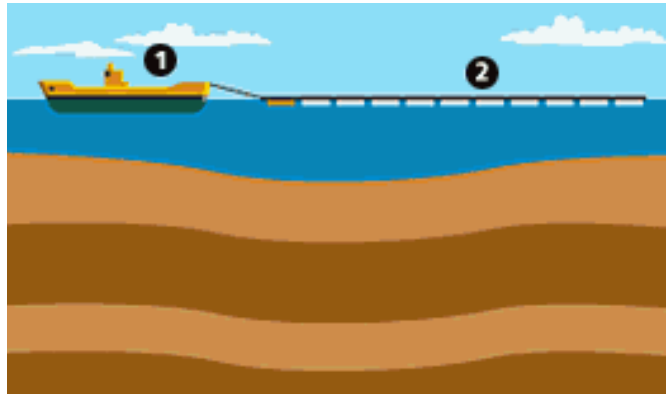
Exploration Geophysics

Community Atmosphere Model (CAM)

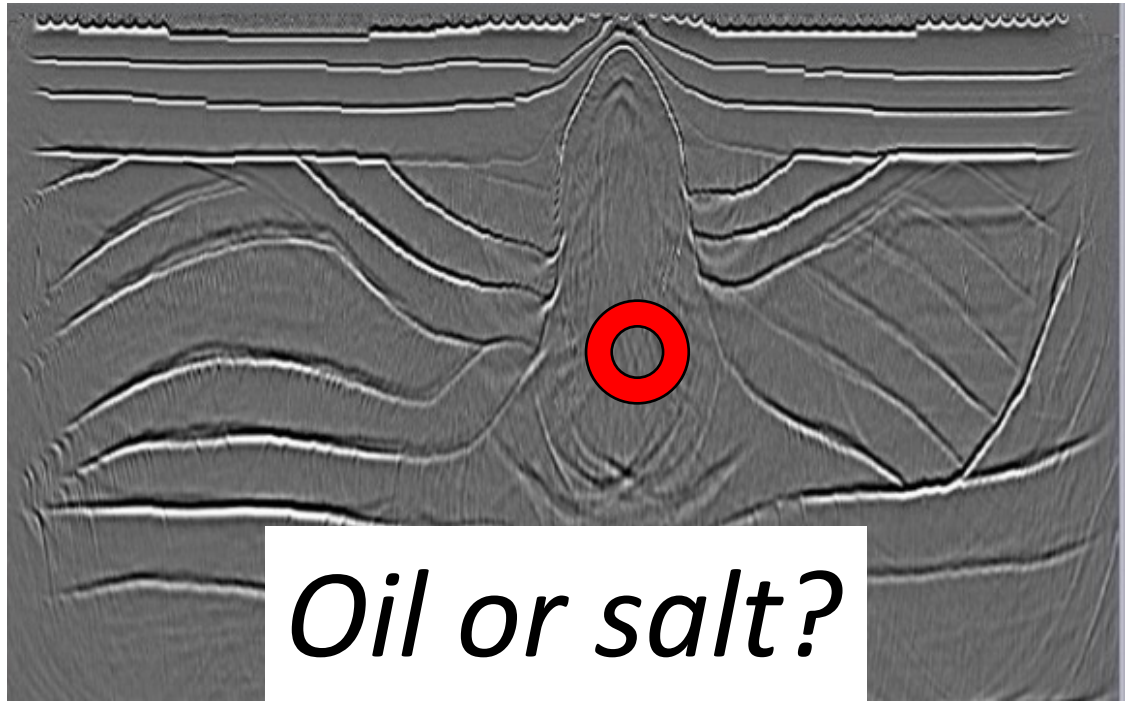
- *Atmospheric model for CESM(climate model) developed by* ***the National Center for Atmosphere Research (NCAR)***
- *One of the most computationally consuming part of CESM*
- *A popular & killer HPC application*
- ***Dynamics core (DNY)*** *and* ***Physics scheme (PHY)***

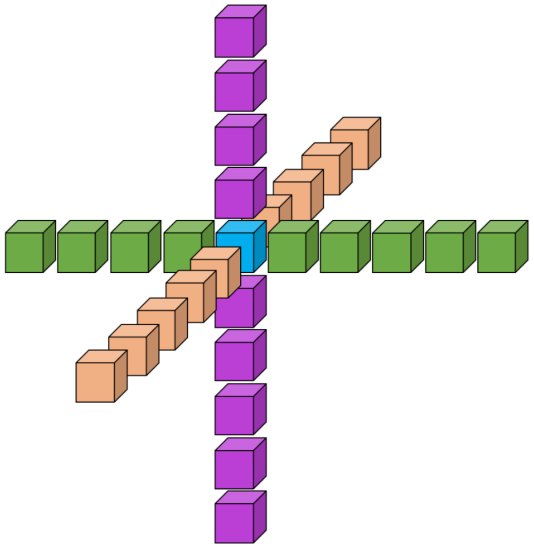
CAM Workflow





- ***Elastic* Reverse Time Migration (RTM)**
  - *Major hot spot for migration algorithm*
  - *Evolving to be more complex*





RTM stencil

Geoscience Applications to Sunway TaihuLight

- OpenACC Refactoring
- Athread Redesigning

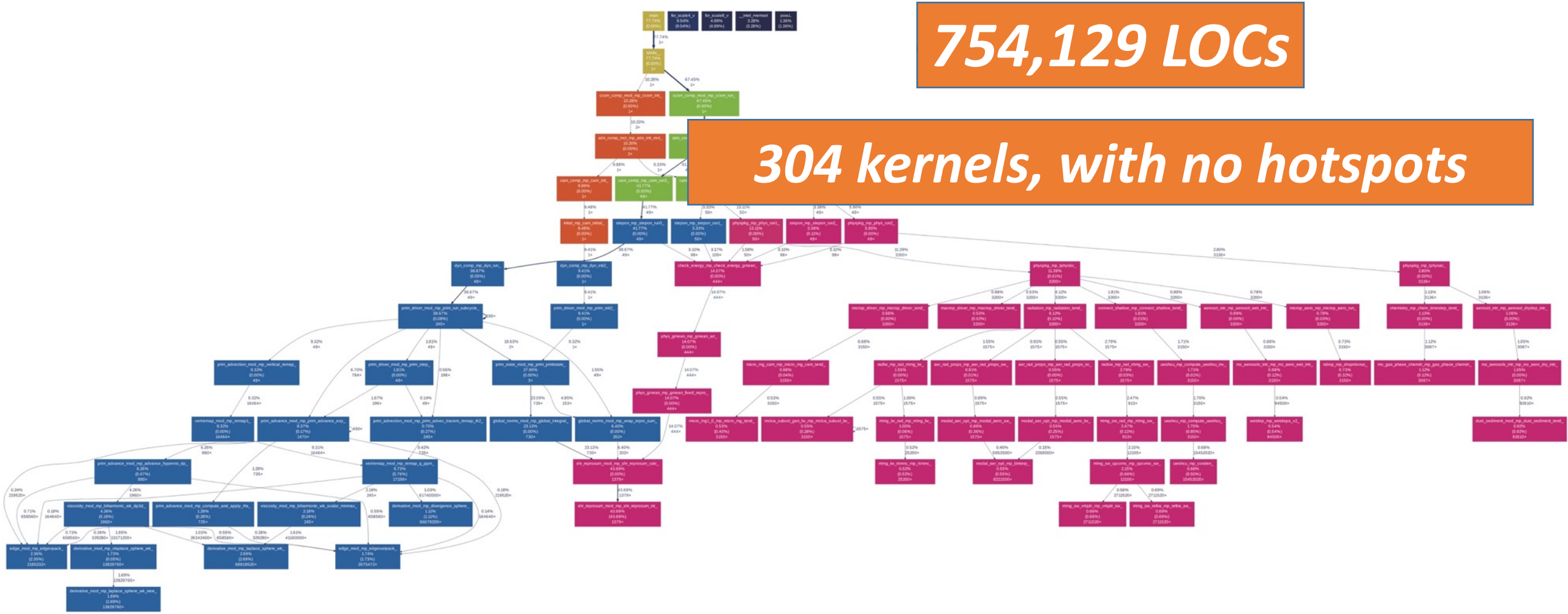
OpenACC

- *Sunway OpenACC Compiler*
  - *A customized version based on OpenACC 2.0*
  - *Directive-based and source-to-source compiler*

Workflow of CAM

754,129 LOCs

304 kernels, with no hotspots



OpenACC Refactoring

- Sunway OpenACC Compiler
  - A customized version based on OpenACC 2.0
  - Directive-based and source-to-source compiler

- To efficiently apply OpenACC on numerous Loops
  - Loop transformation tool to expose the right level of parallelism and data size

OpenACC for Refactoring (exp. 1)

Loop Abstraction

LOOP A

Body A

LOOP B

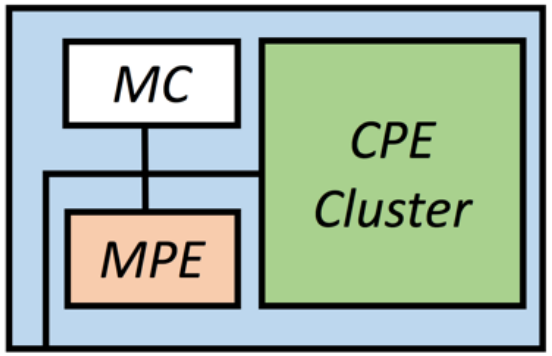
Body B

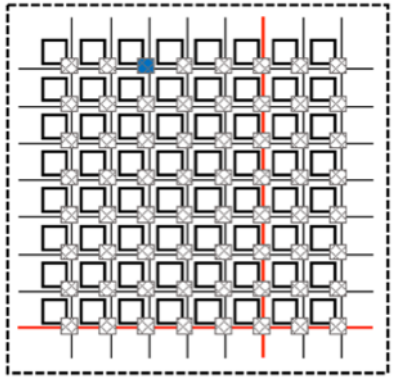
LOOP C

Originally designed for multi-core system with cache

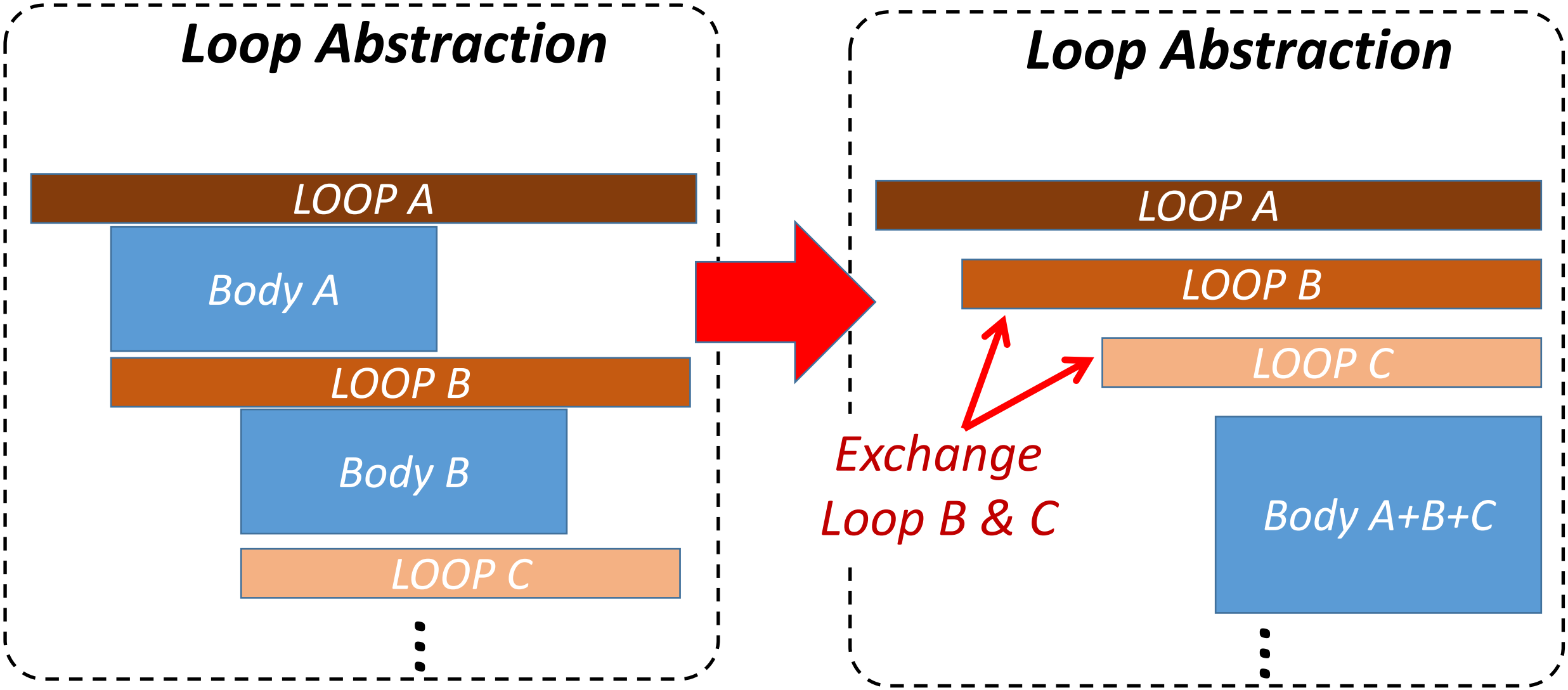
Small loop size

Fastest direction

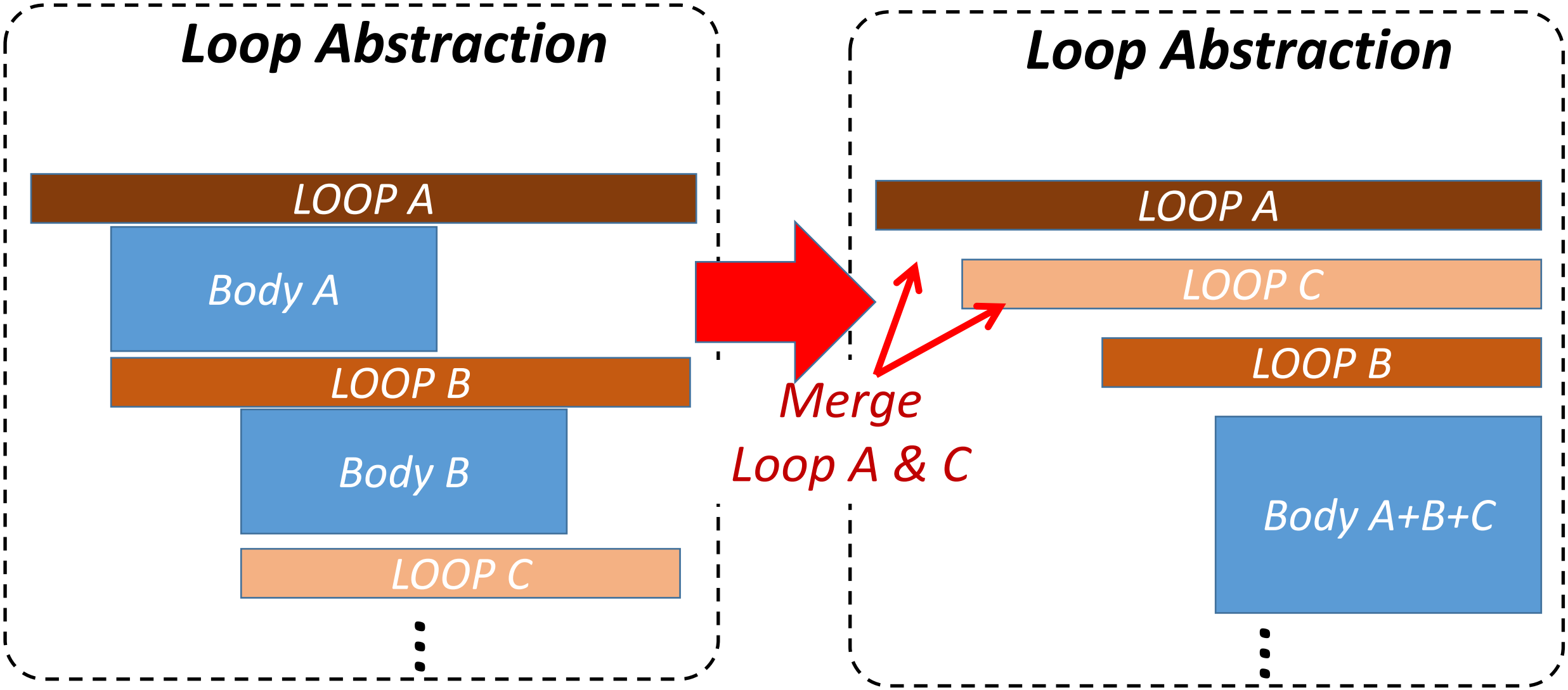




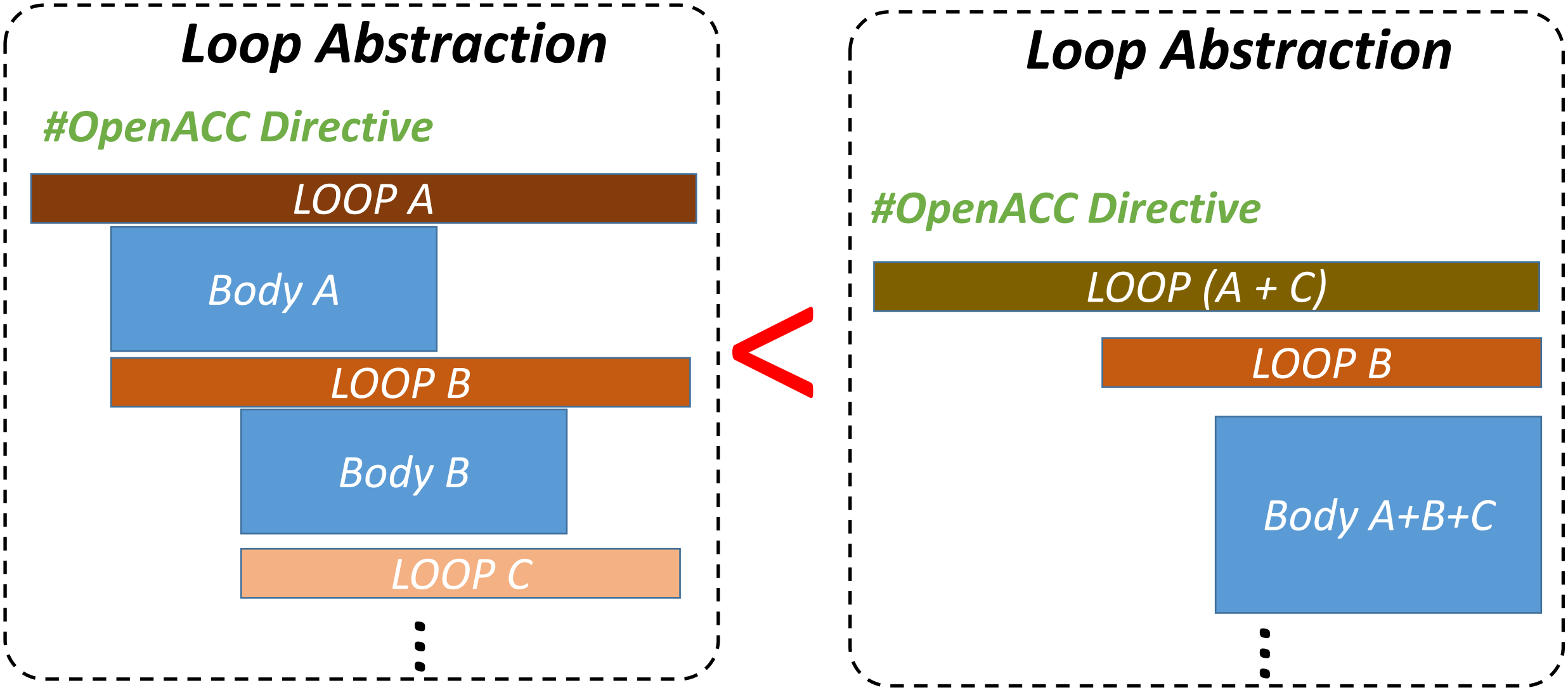
OpenACC for Refactoring (exp. 1)



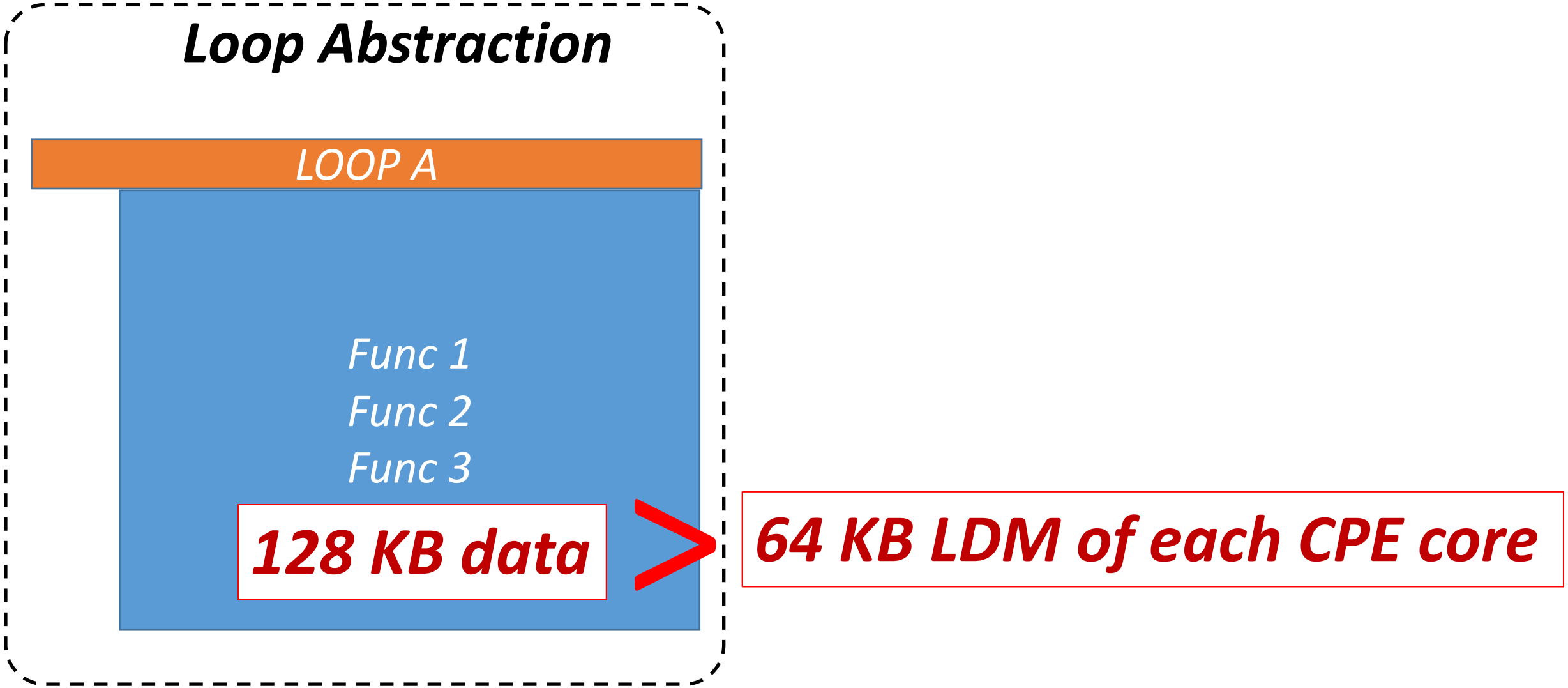
OpenACC for Refactoring (exp. 1)



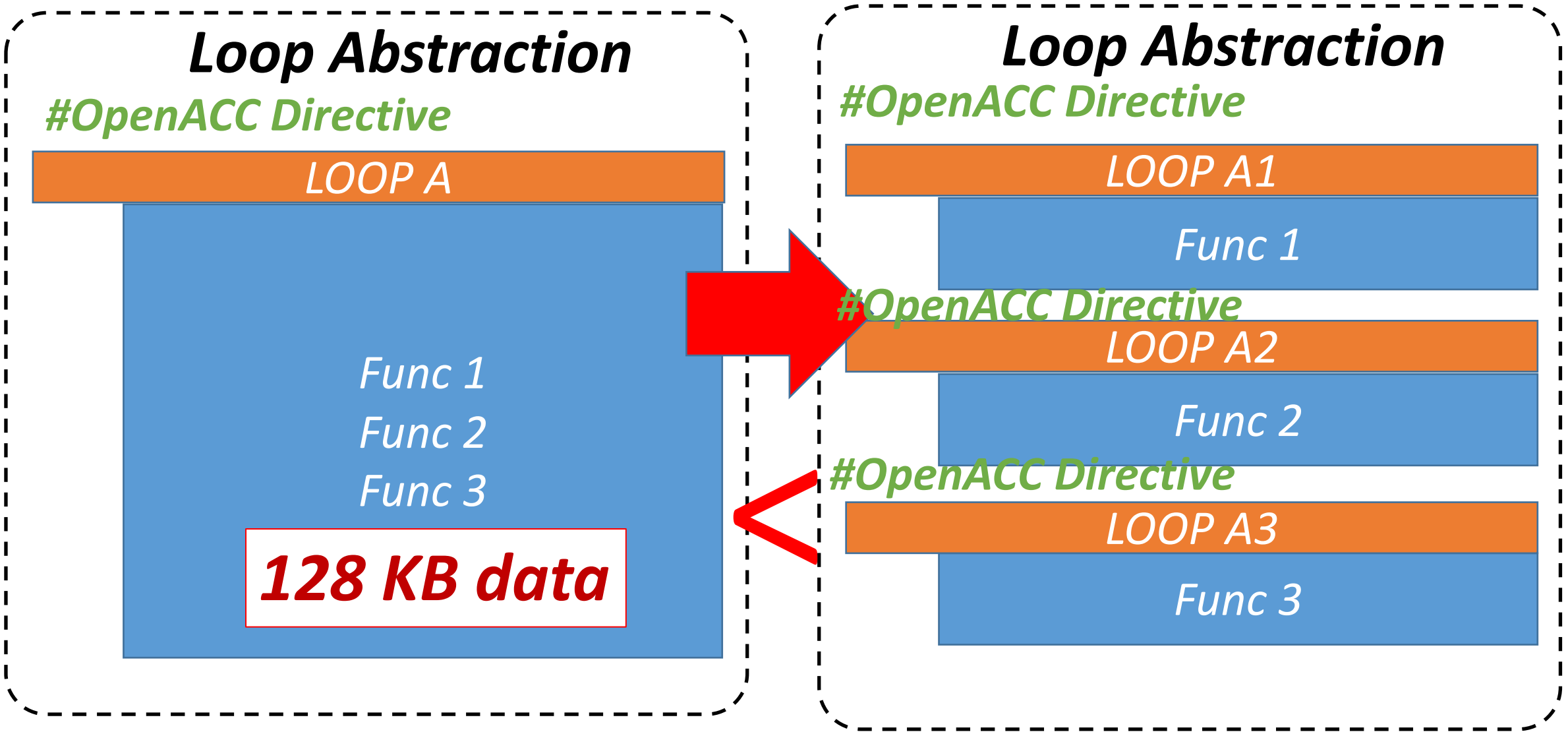
OpenACC for Refactoring (exp. 1)



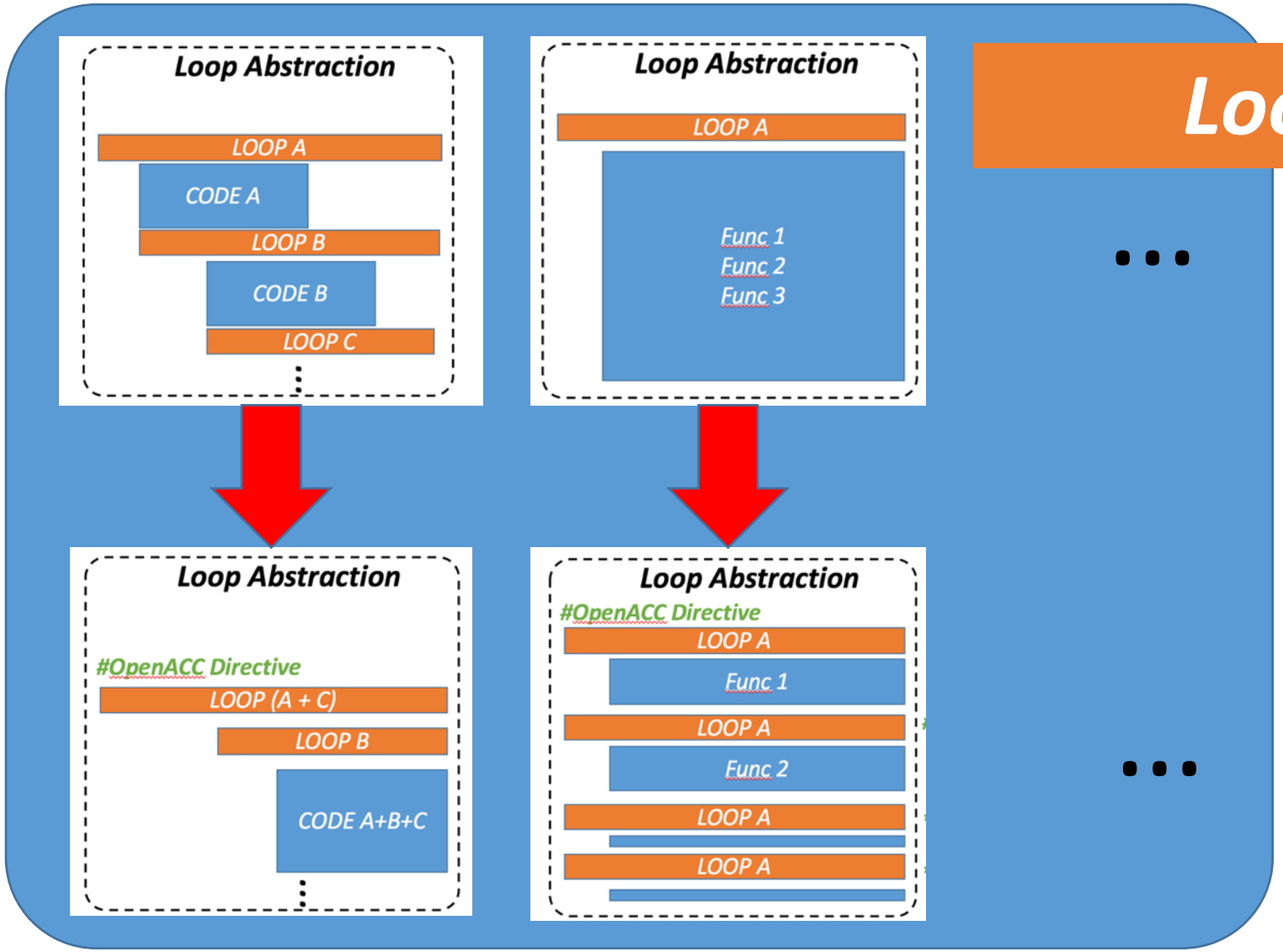
OpenACC for Refactoring (exp. 2)



OpenACC for Refactoring (exp. 2)



Tools for OpenACC Implementation



Loop Transform.

Tools for auto-refactoring

DYN and PHY

Restrictions

CAM Model

- *Difficulty in data locality to fit with the 64 KB LDM*
- *Computation patterns that are unfriendly to vectorization*
- *Absence of overlapping schemes to hide communications*

OpenACC

- *OpenACC removes options to achieve finer control of computing or memory operations or register communication*
- *Threading overhead becomes a huge issue for no-hotspot module*

Redesign is necessary

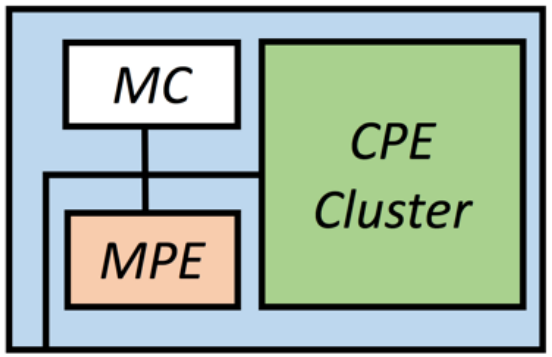
Geoscience Applications to Sunway TaihuLight

- OpenACC Refactoring
- Athread Redesigning

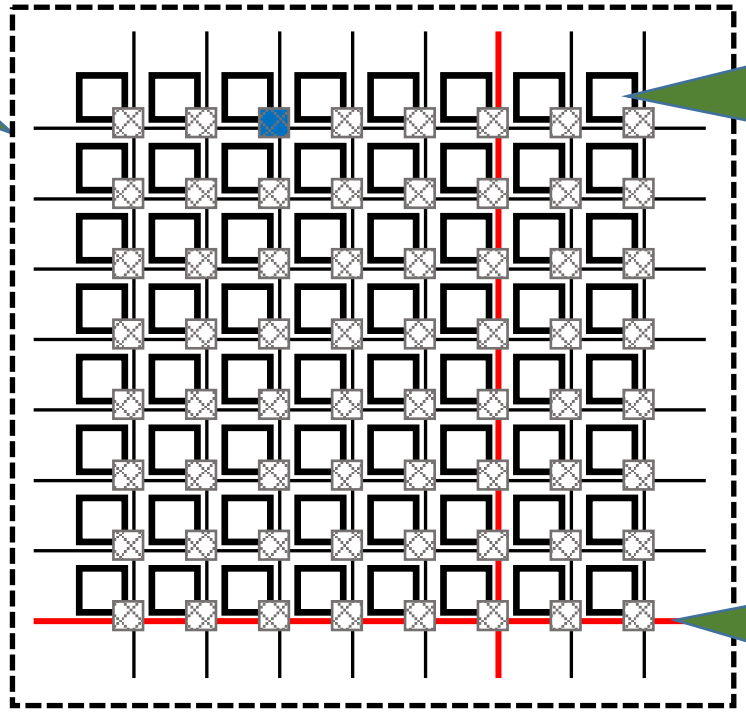
Ahtread: Fine-Grained Parallel Approach

- **Take more aggressive operations to redesign**

Parallel scheme using 64 CPES



Heterogeneous arch.



One CPE
64 KB Local data memory (LDM)

Between CPE
Fast register communication

Parallel Scheme

Element

Direct

One CPE for one Element

OpenACC

Decomp.

Sub-Element 0
Sub-Element 1
Sub-Element i-1

CPE 0
CPE 1
CPE i-1

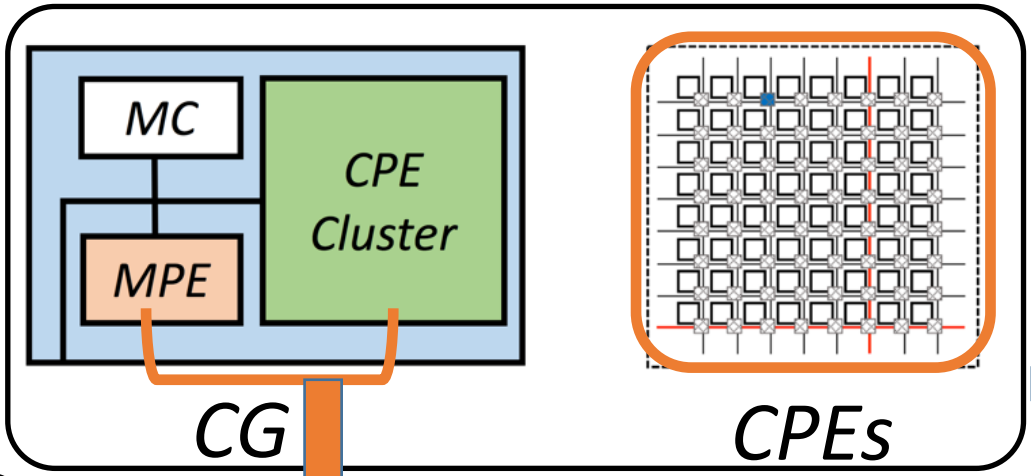
#i CPEs for one Element

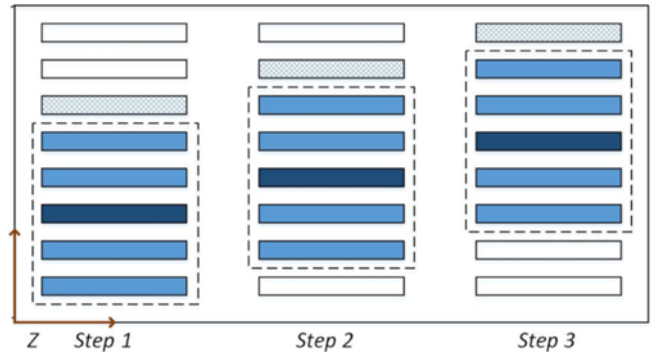
More parallelism or lower LDM requirement

Cubed-sphere mesh

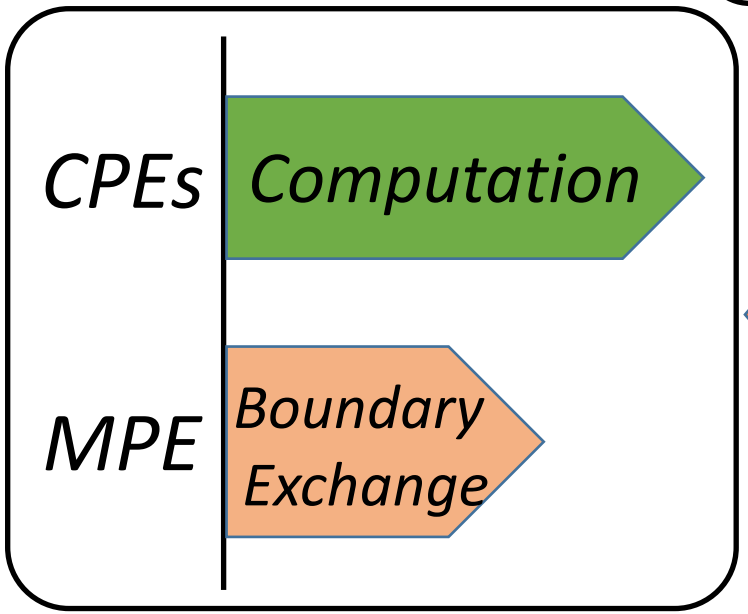
Overlapping Strategies

Heterogeneous
MPE for inter-Comm.
CPEs for Comp.

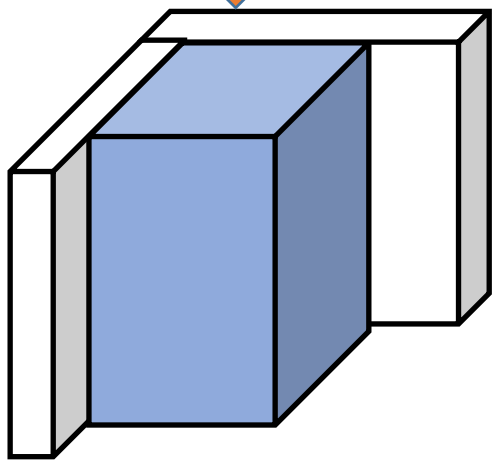




Pipeline Scheme

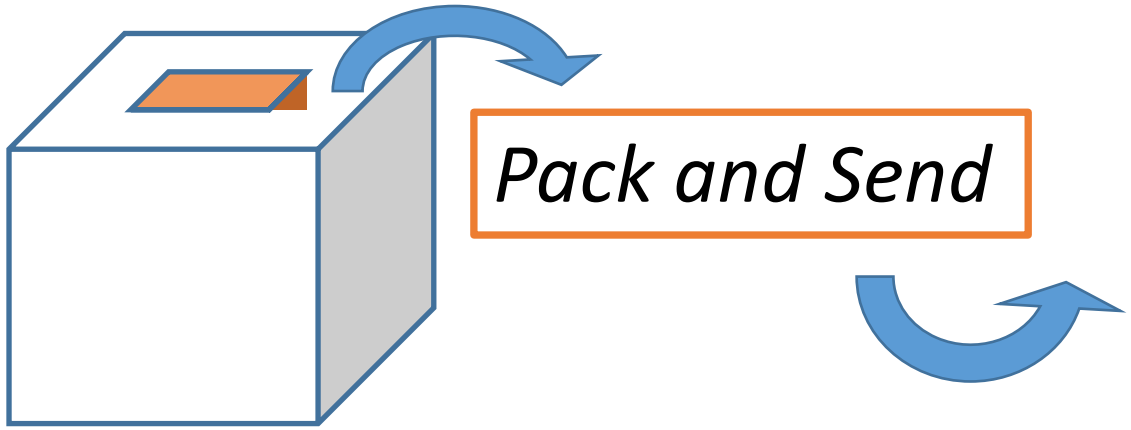


Stencil-like
Hybrid decomp. into Inner and Outer parts for halo overlapping

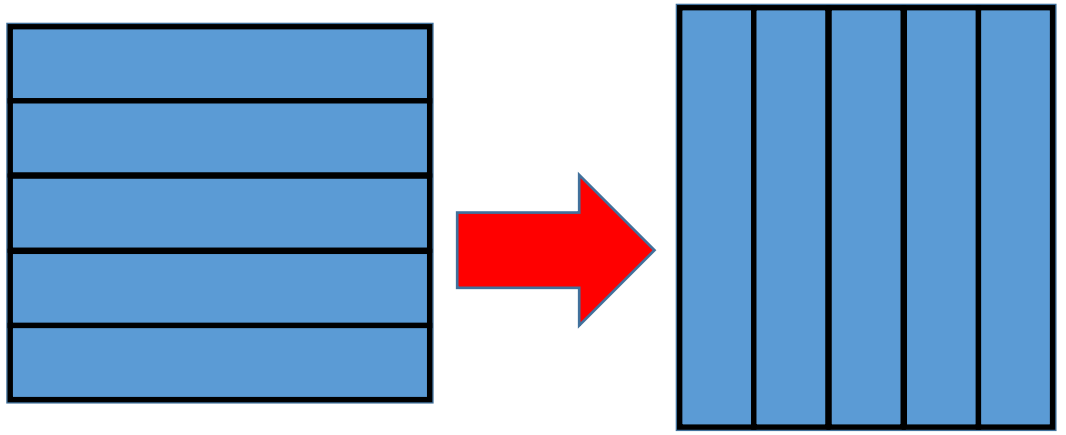


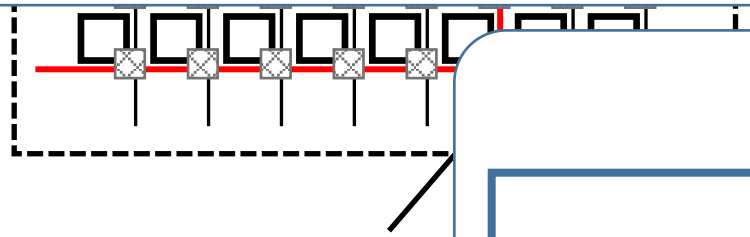
Array Transposition by Register Communication

Boundary Exchange



On-the-fly Matrix Transposition

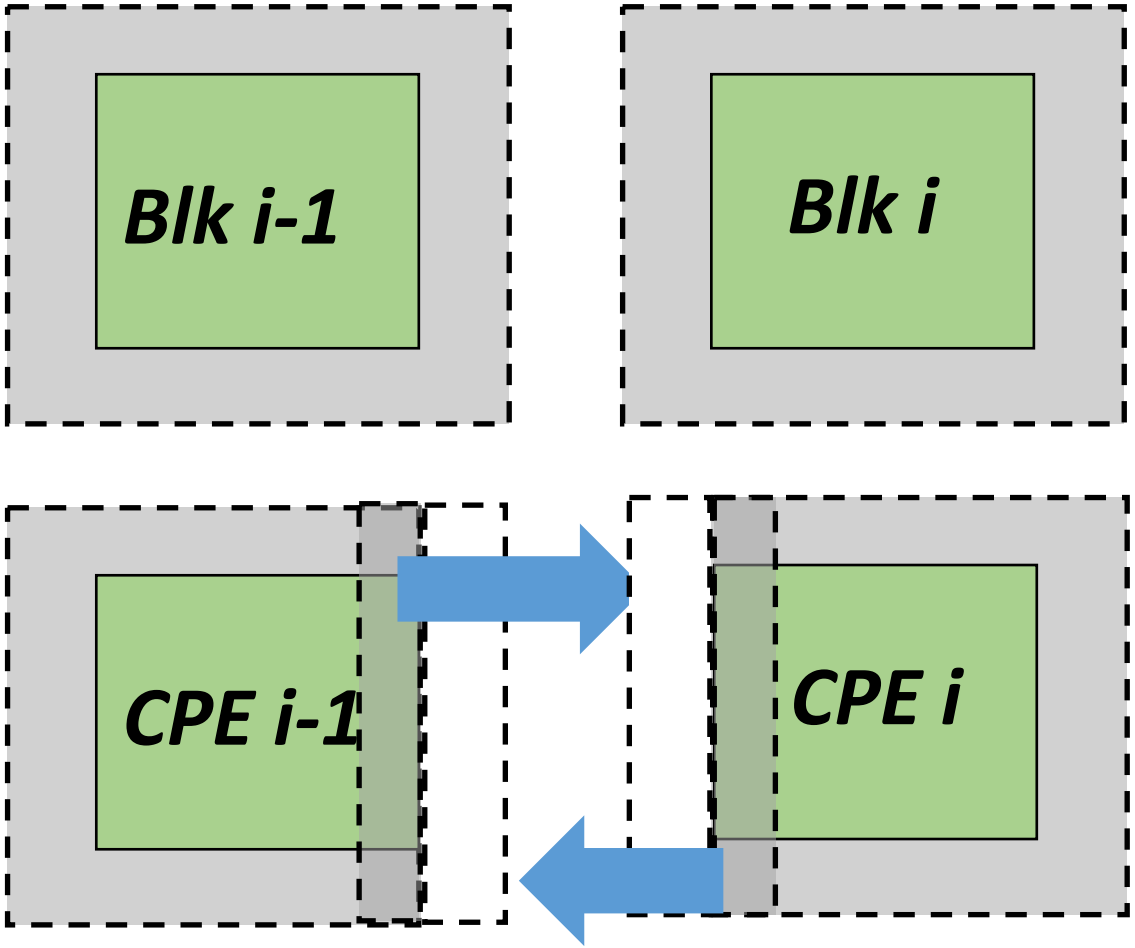


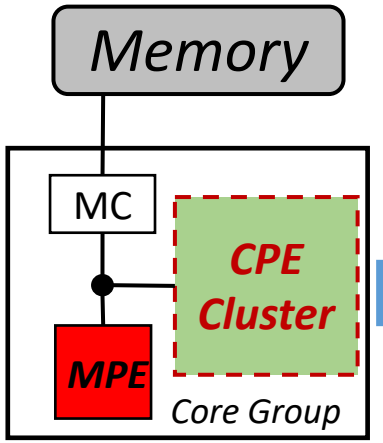


Registe
communic
(~10cycl

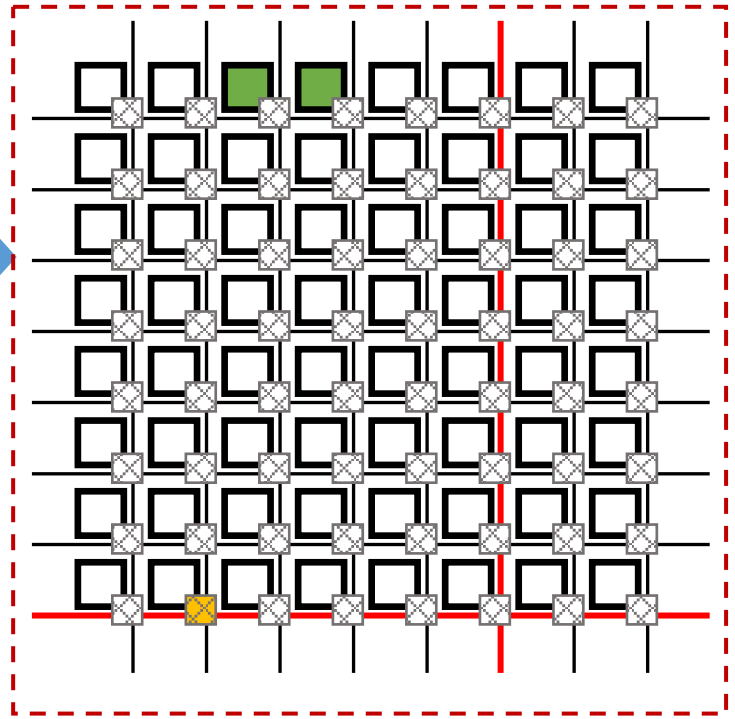
Matrix-Matrix Multiplication

Data Sharing via Register Communication

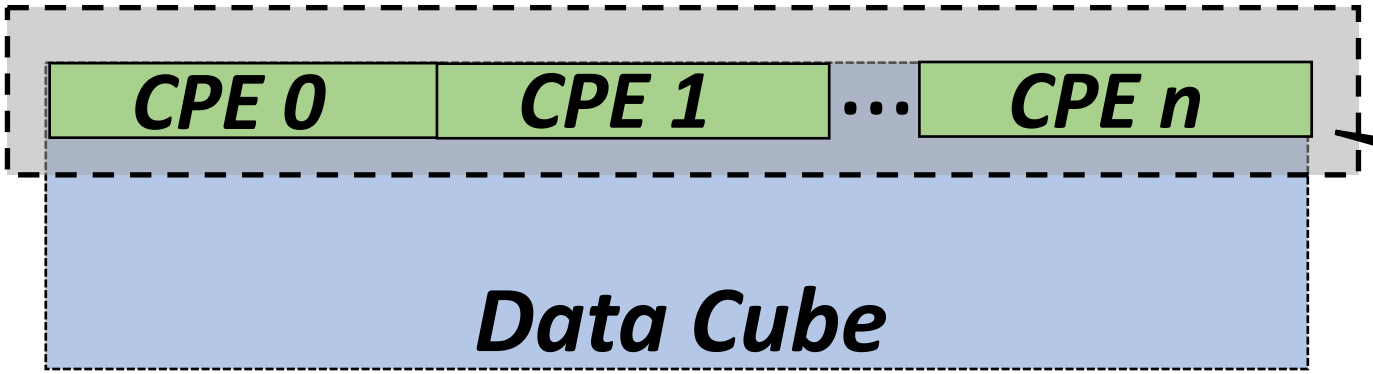




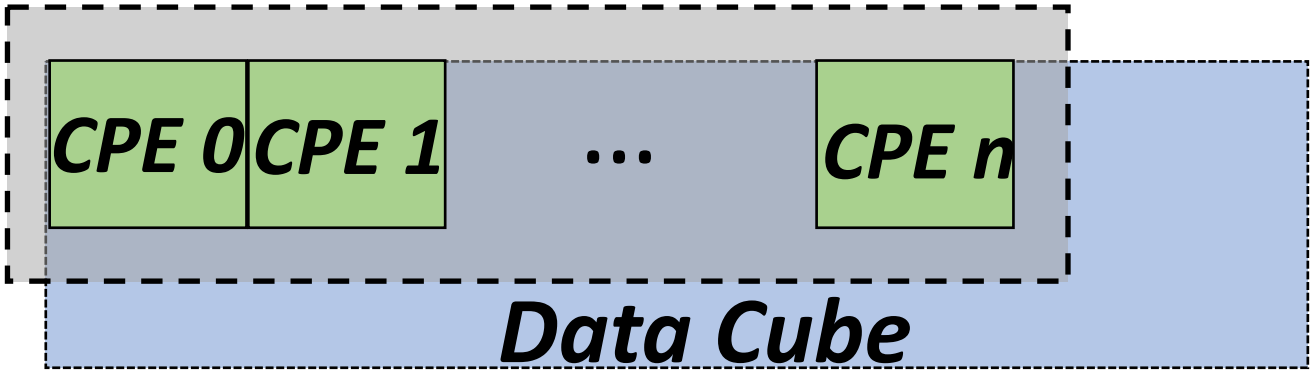
64 CPEs (Slave Cores)

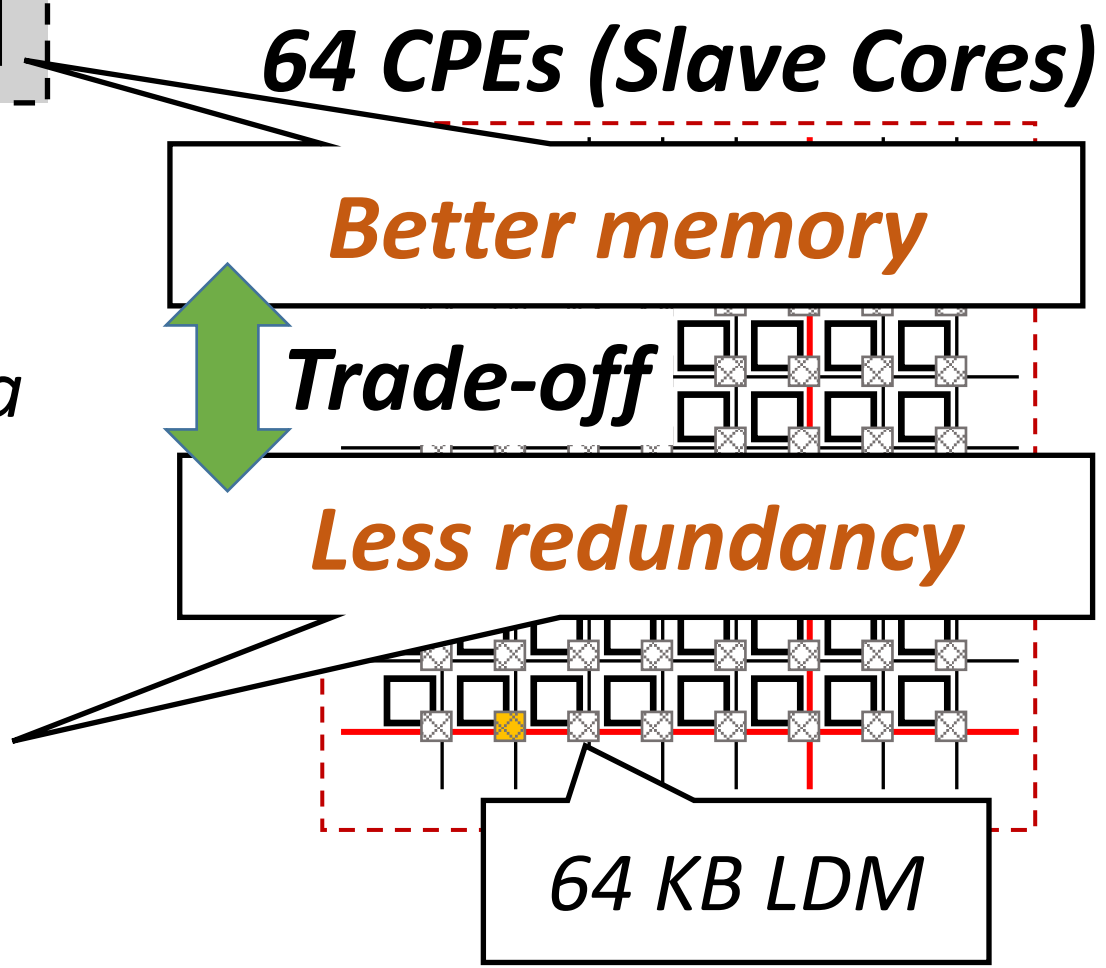


Locality-aware Memory Design

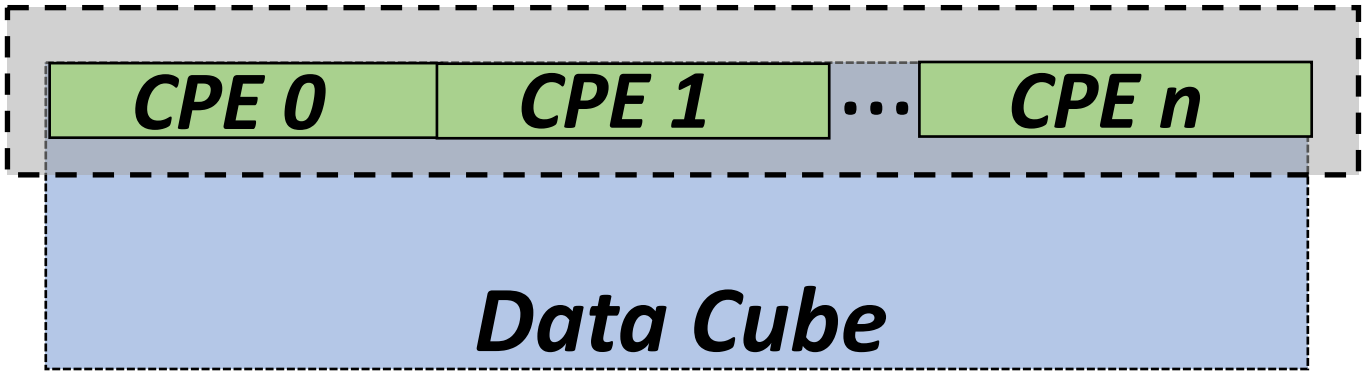


*One CPE to control a **stride/block** data*

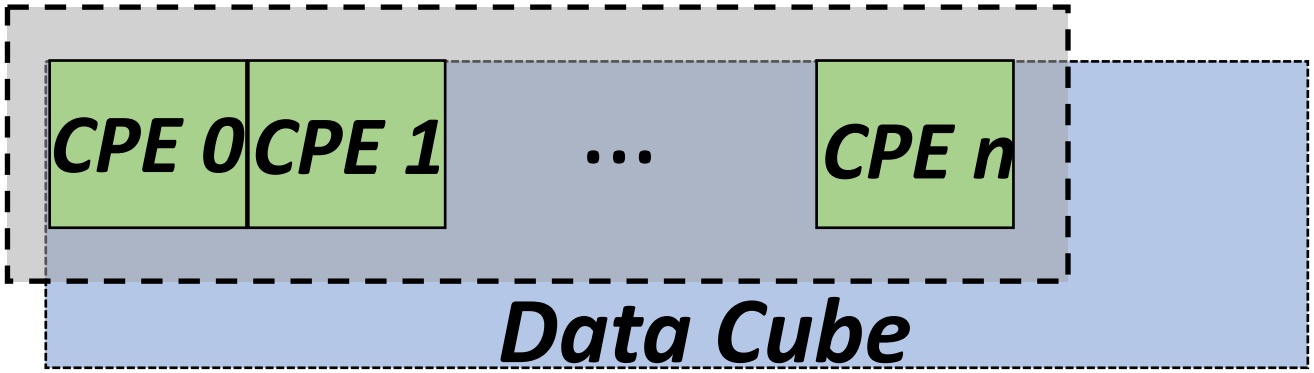




Locality-aware Memory Design



*One CPE to control a **stride/block** data*

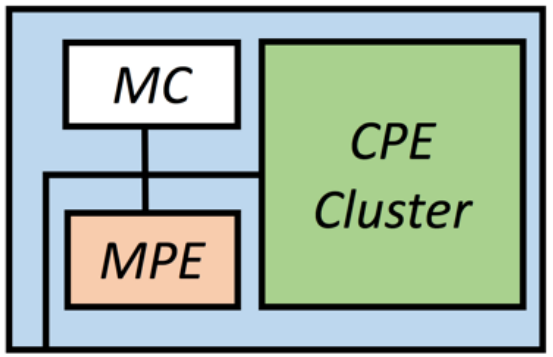


Cache-like Mechanism for better locality

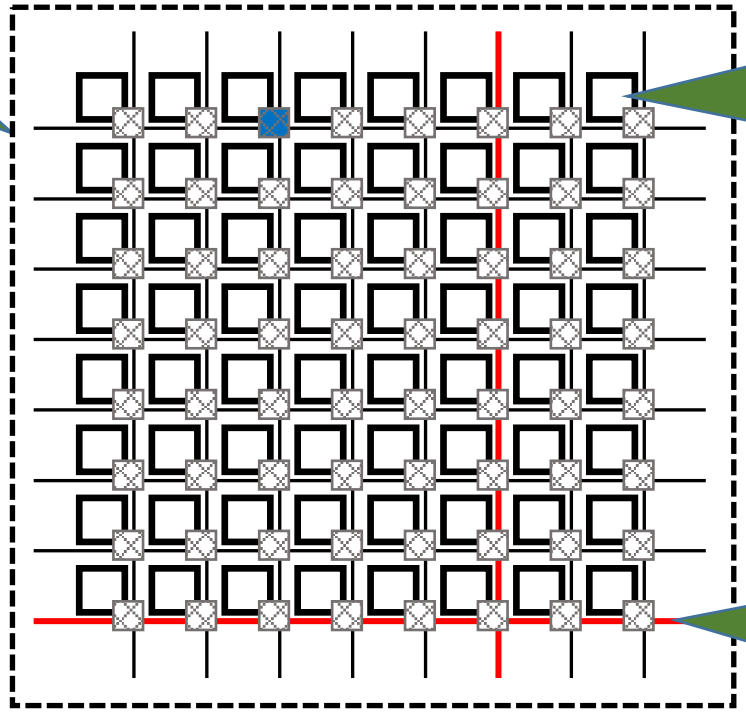
Systematic Redesign Solution by Athread

- **Take more aggressive operations to redesign**

Parallel scheme using 64 CPES



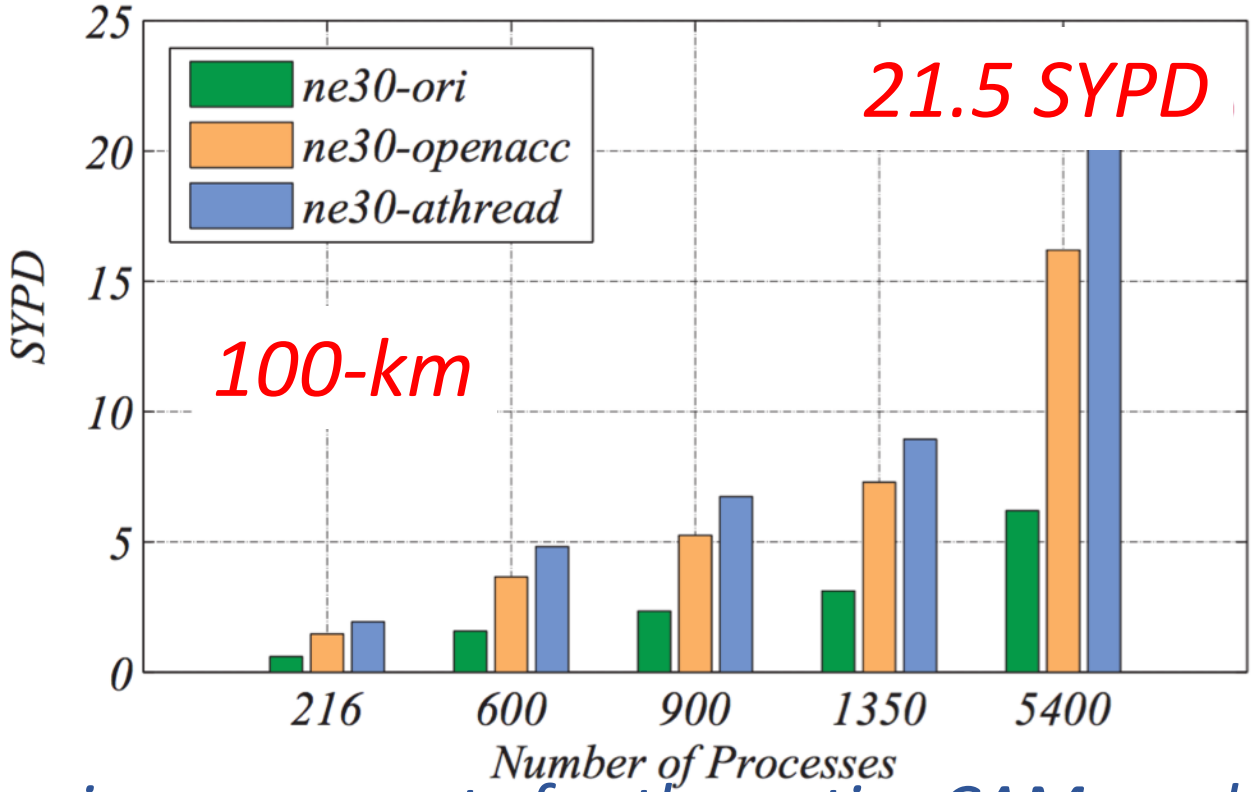
Heterogeneous arch.



One CPE
64 KB Local data memory (LDM)

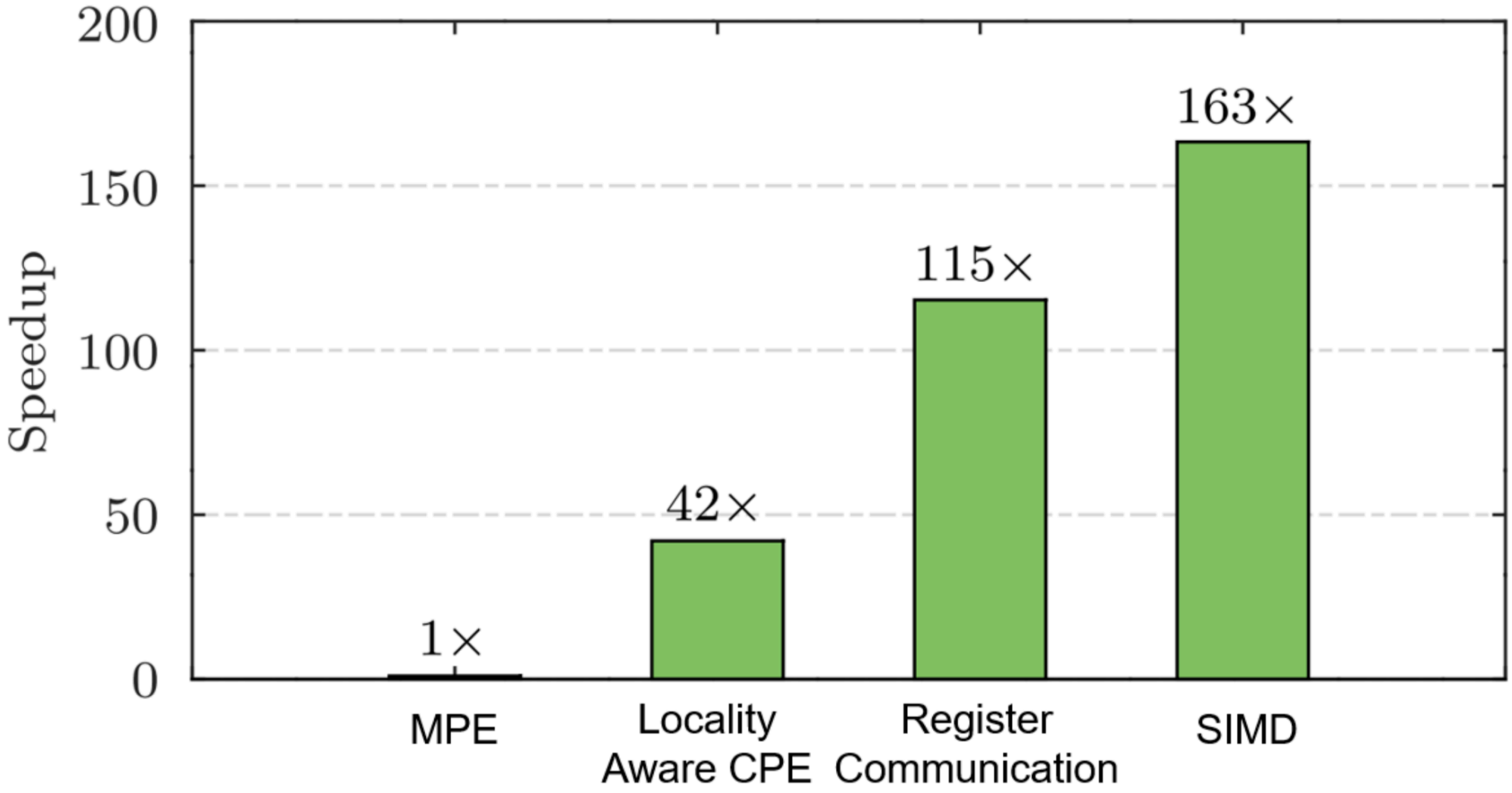
Between CPE
Fast register communication

Performance Speedups for CAM on TaihuLIght

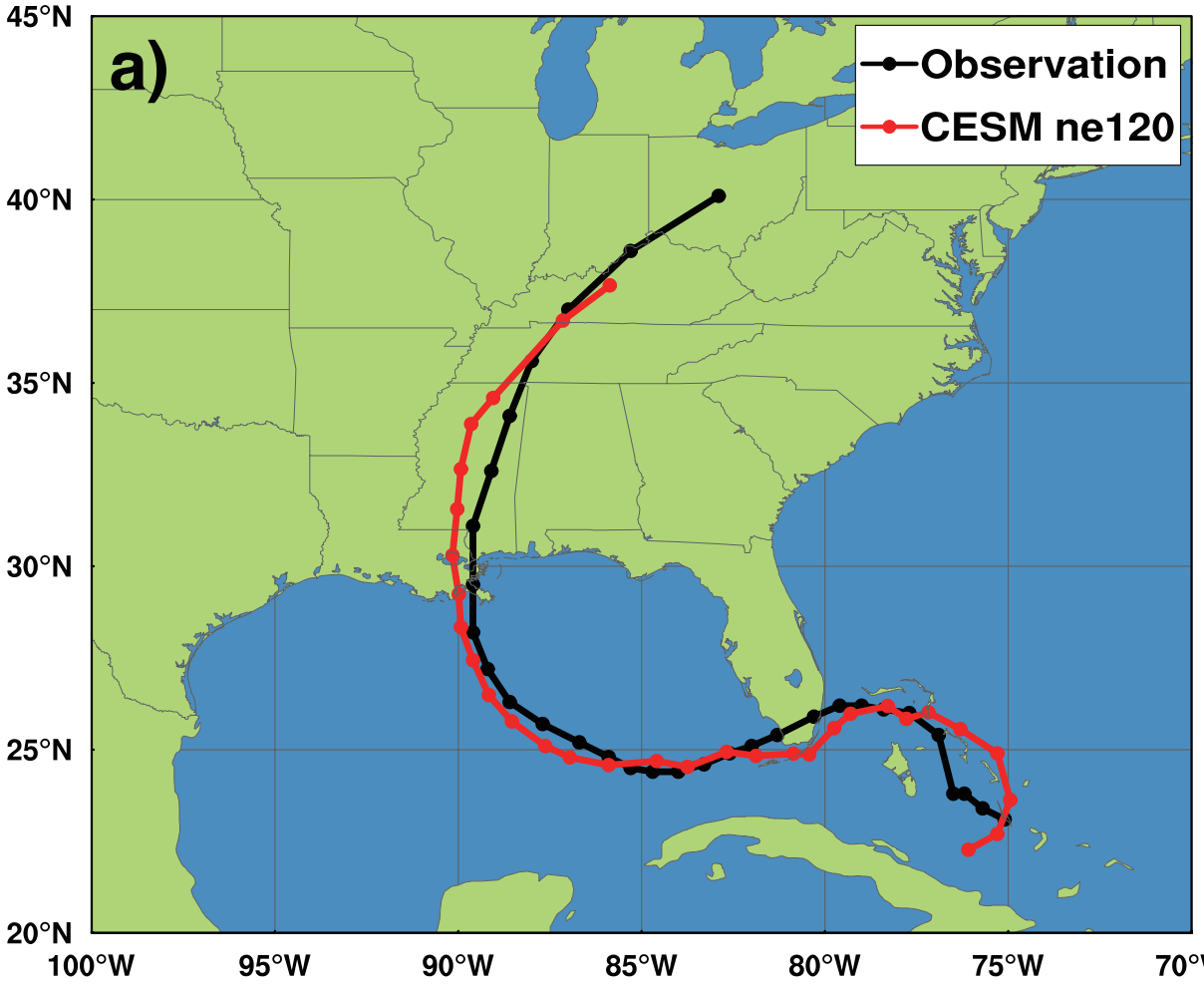


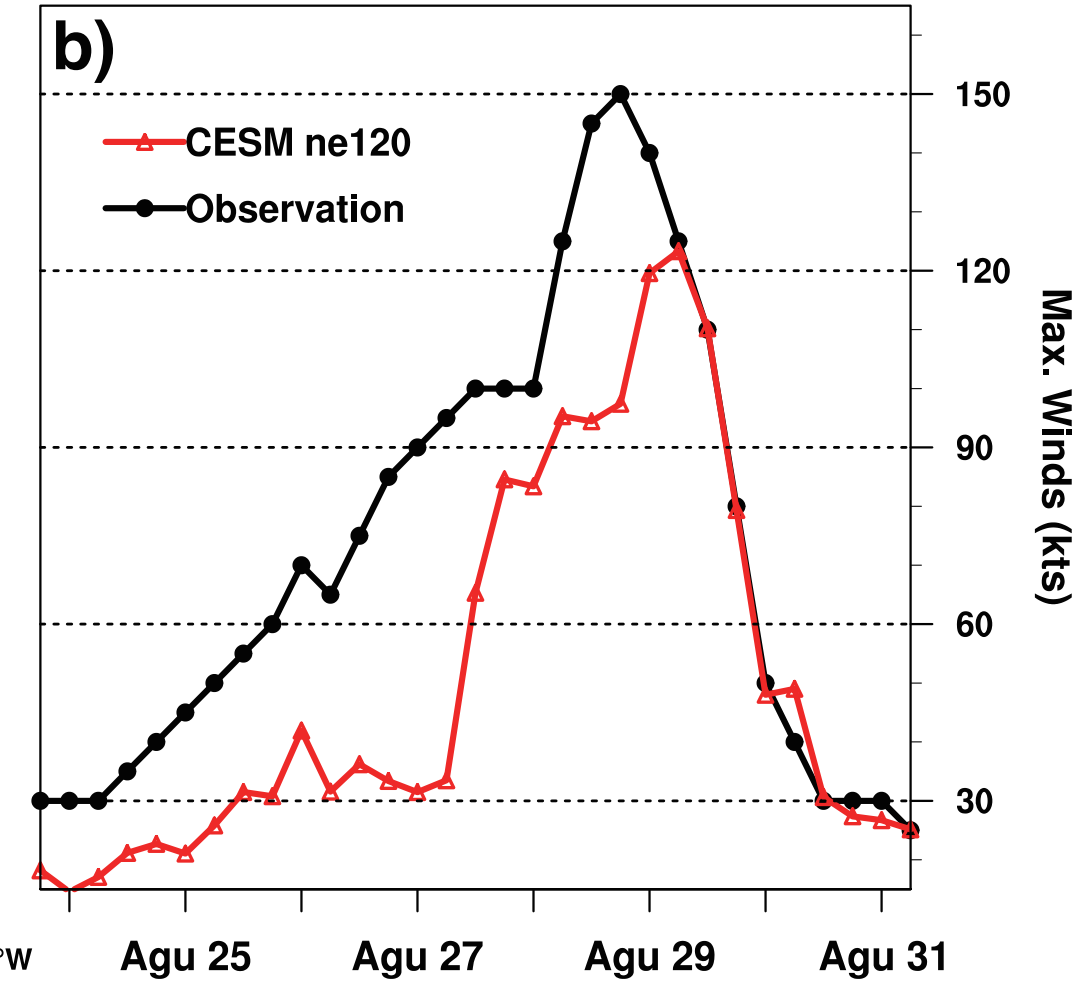
The performance improvements for the entire CAM model in ne30 and ne120. ori refers to the original version based on MPE, openacc refers to the usage of OpenACC directive, and athread refers to the further usage of Athread.

Performance Speedup of Elastic RTM



Hurricane Simulation – Katrina (2005)

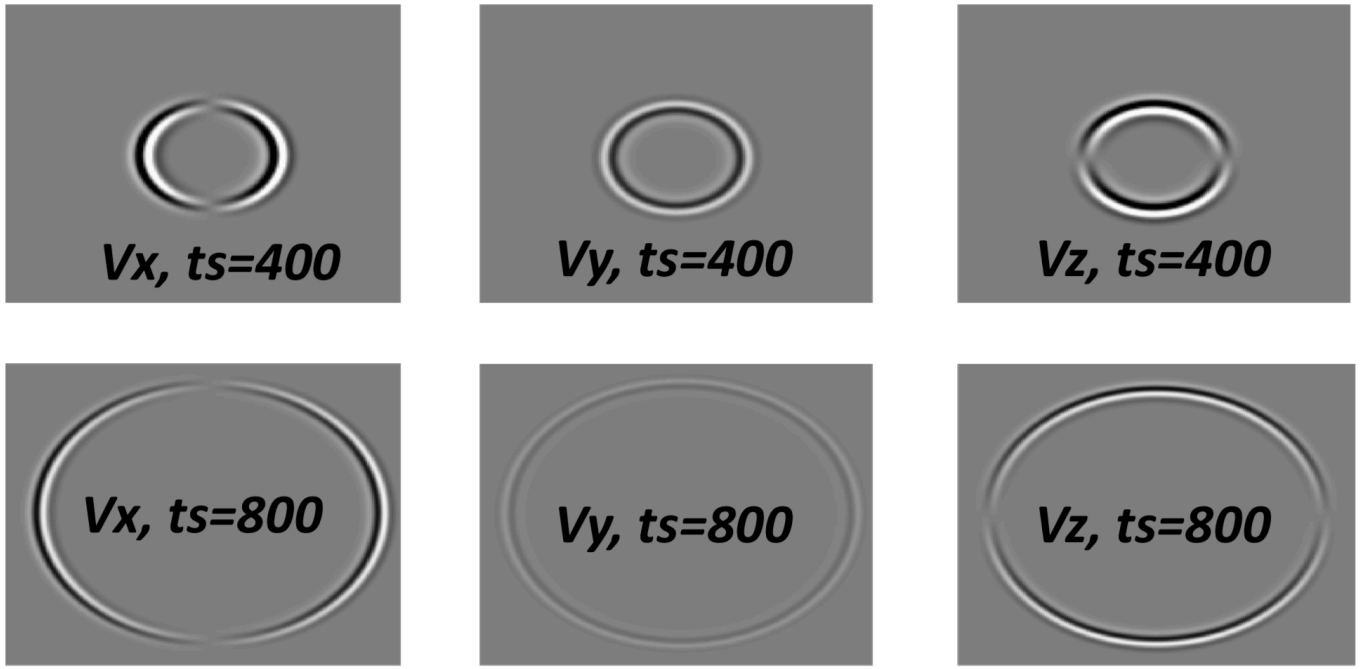




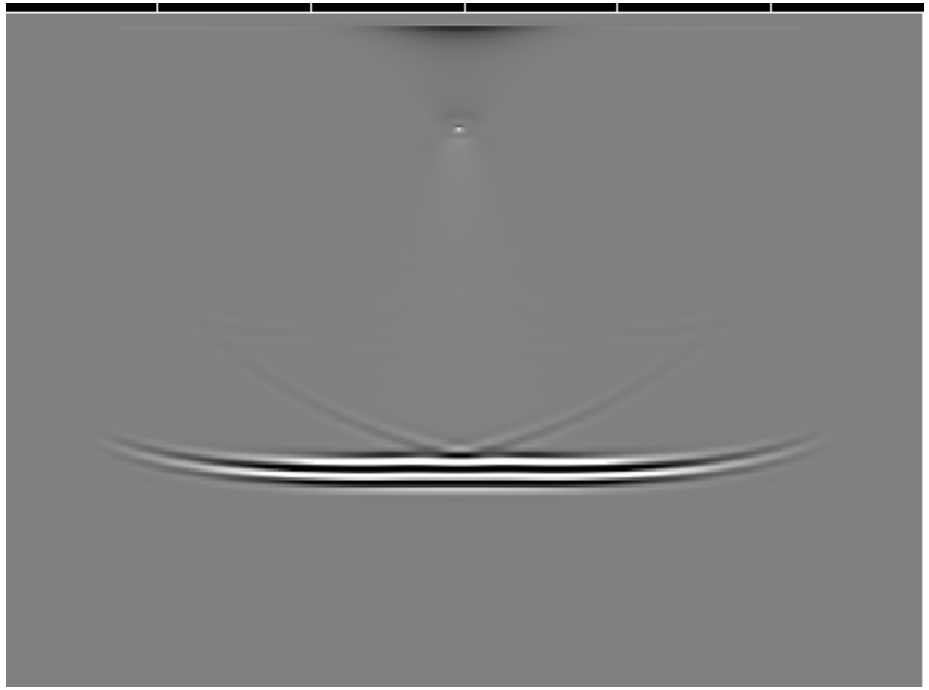
Track

Intensity

Validation of Elastic RTM



Wave propagation of three particle velocities at step 400 and 800, views from xoy, yoz, and xoz

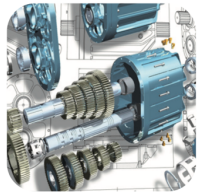
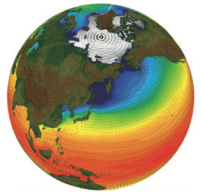
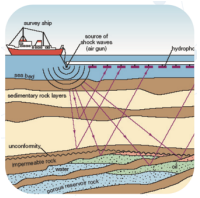


Drho-Layer Benchmark

Summary

- ***Solutions for dealing with real-world applications***
  - *OpenACC Refactoring, and the automatic tools for loop transformation*
  - *Athread Redesign, with different solutions towards architectural features*
- ***Proved thread-level comm. a good option***
- ***Virtual cache: make up the con of the Sunway architecture***
- ***Provide better support for applications & current and upcoming (Exascale) many-core supercomputers***

Big Computing & Big Data



Software Ecosystem

xMath	Thunder	VASP	swLBM
Petsc	OpenFoam	LAMMPS	swAllies
bowtie	Star	Gromacs	swRTM

Deep Learning

SWCAFFE

Auto swDNN

SWDNN

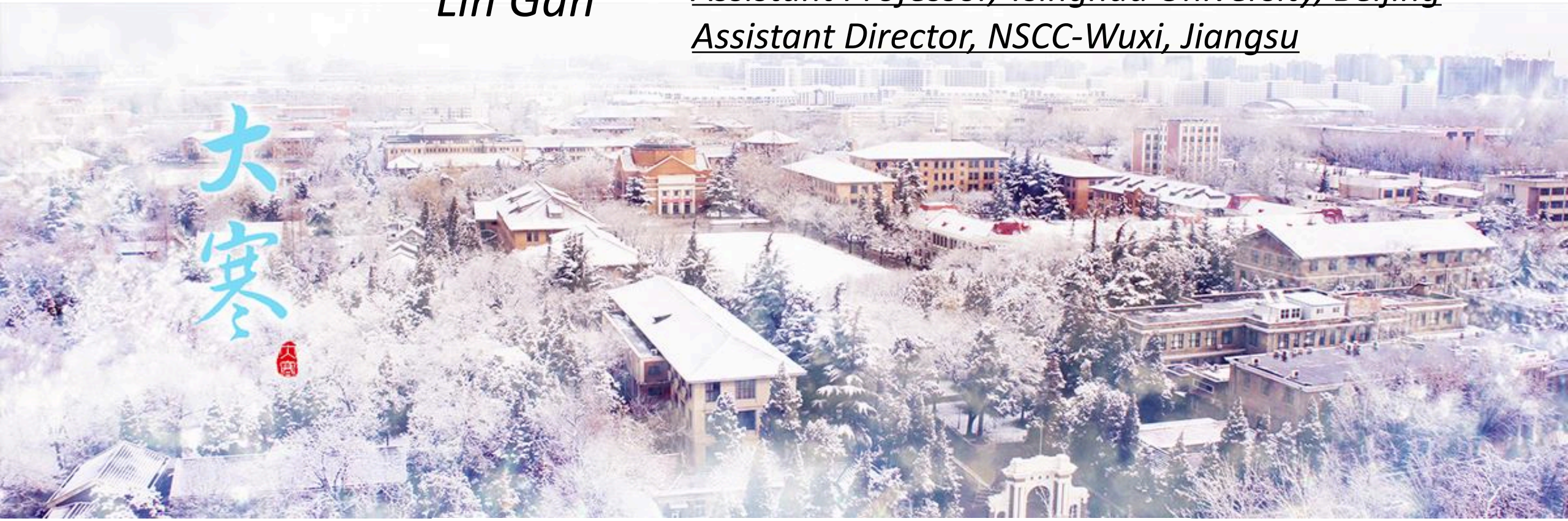
SWGEMM

C | C++ | Fortran | OpenACC | Athread | gprof | Jperf | swLU

Scaling Geoscience Applications on Sunway Supercomputer

Lin Gan

Assistant Professor, Tsinghua University, Beijing
Assistant Director, NSCC-Wuxi, Jiangsu



Sep. 10, 2019, PPAM 2019 @ Bialystok, Poland