20th ECMWF workshop on high performance computing in meteorology

中国气象局地球系统数值预报中心
CMA EARTH SYSTEM MODELING AND PREDICTION CENTRE

清華大學
Tsinghua University

Parallel Software Framework of MCV Model

Qingu Jiang[1,2], Li Liu[2], Xingliang Li[1], Xueshun Shen[1], Xinzhu Yu[2]

1 CMA Earth System Modeling and Prediction Centre (CEMC)

2 Department of Earth System Science, Tsinghua University

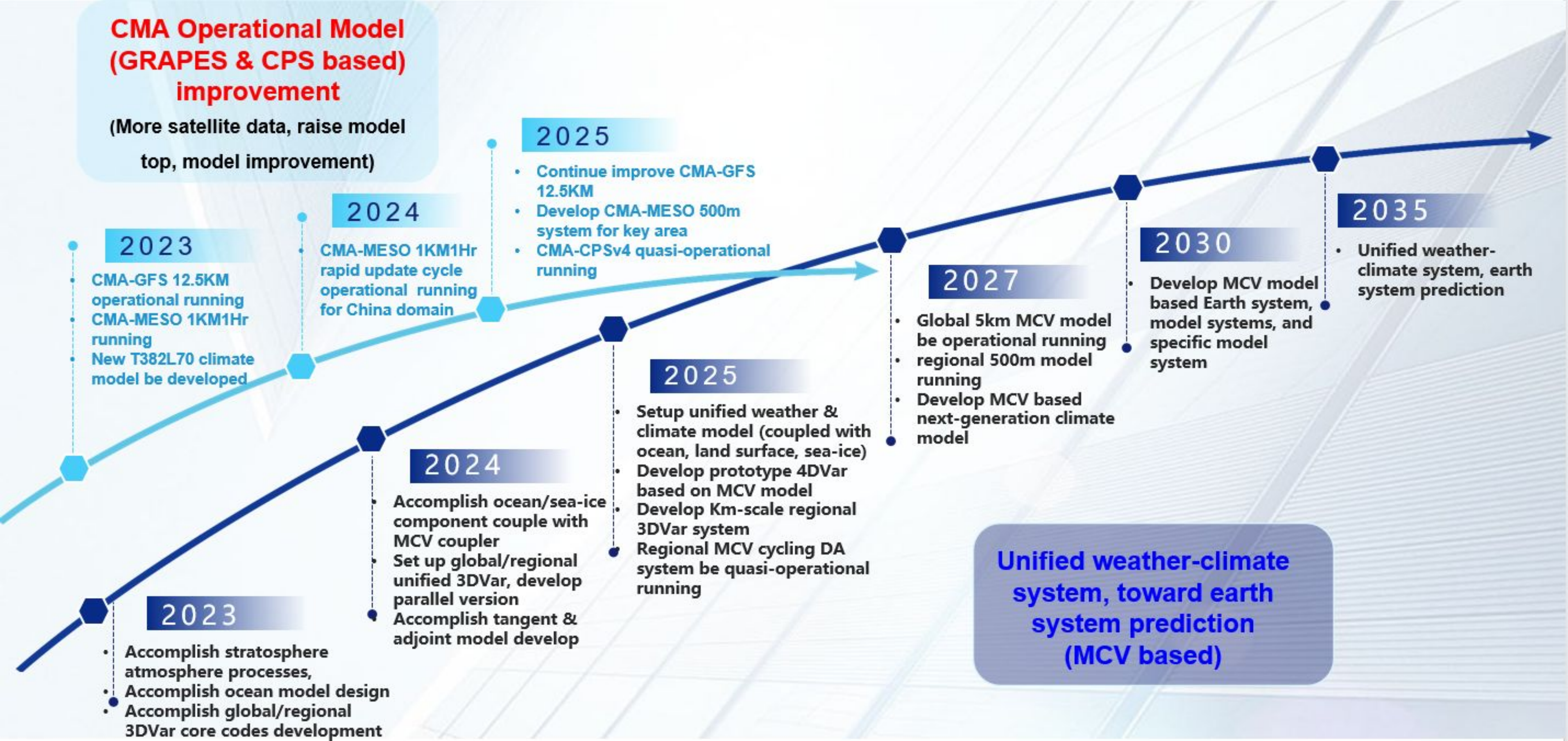
2023/10/13

Contents

01

Background

CEMC’s Roadmap 2023-2035



CMA Operational Model (GRAPES & CPS based) improvement

(More satellite data, raise model top, model improvement)

2023

- CMA-GFS 12.5KM operational running
- CMA-MESO 1KM1Hr running
- New T382L70 climate model be developed

2024

- CMA-MESO 1KM1Hr rapid update cycle operational running for China domain

2025

- Continue improve CMA-GFS 12.5KM
- Develop CMA-MESO 500m system for key area
- CMA-CPSv4 quasi-operational running

2023

- Accomplish stratosphere atmosphere processes,
- Accomplish ocean model design
- Accomplish global/regional 3DVar core codes development

2024

- Accomplish ocean/sea-ice component couple with MCV coupler
- Set up global/regional unified 3DVar, develop parallel version
- Accomplish tangent & adjoint model develop

2025

- Setup unified weather & climate model (coupled with ocean, land surface, sea-ice)
- Develop prototype 4DVar based on MCV model
- Develop Km-scale regional 3DVar system
- Regional MCV cycling DA system be quasi-operational running

2027

- Global 5km MCV model be operational running
- regional 500m model running
- Develop MCV based next-generation climate model

2030

- Develop MCV model based Earth system, model systems, and specific model system

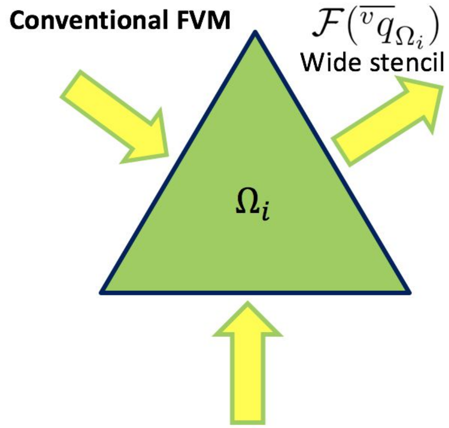
2035

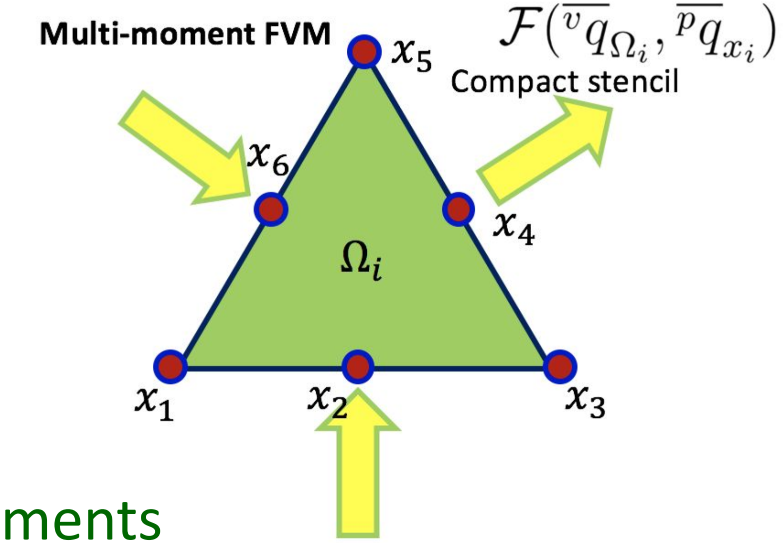
- Unified weather-climate system, earth system prediction

Unified weather-climate system, toward earth system prediction (MCV based)

Multi-moment constrained finite volume (MCV)

“Derivation: Form I” (Ii & Xiao, JCP, 2009)





Moment

Volume integrated average(VIA)

$$\overline{^{v}q}_{V_i}(t) = \frac{1}{|V_i|}\int_{V_i} q(x,t)dV$$

Moments

Volume integrated average **(VIA)**	Point value **(PV)**
$\overline{^{v}q}_{V_i}(t) = \frac{1}{\lvert V_i\rvert}\int_{V_i} q(x,t)dV$	$\overline{^{p}q}_{x_j}(t) = q(x_j,t)$
	$\overline{^{D_k}q}_{x_j}(t) = \frac{\partial^{[k]}}{\partial x^{[k]}}q(x_j,t)$

Multi-moment method uses two or more kinds of moments

A multi-moment FV method distinguishes, memorizes, and updates all of the moments.

A regional/global unified MCV model

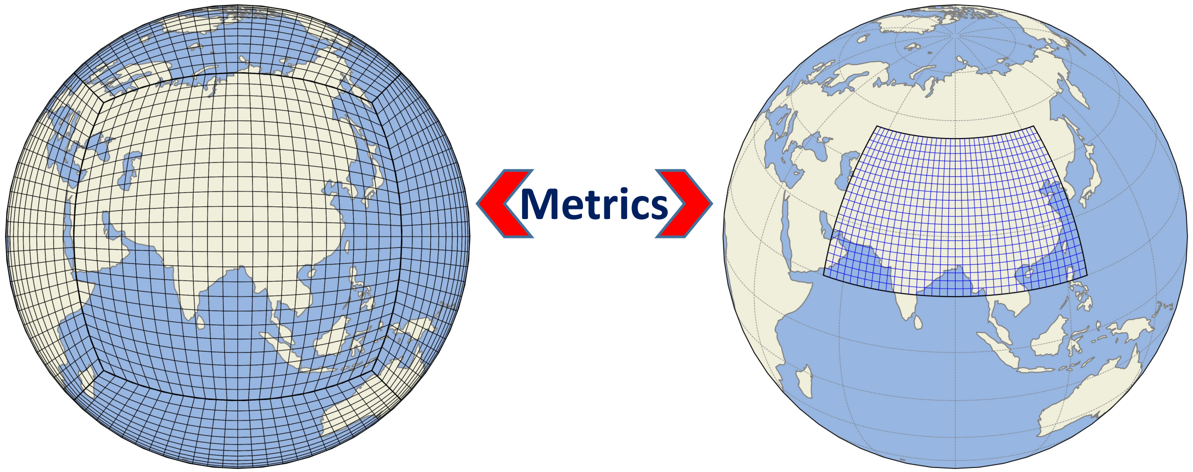
The governing equation on the curvilinear system

$$\frac{\partial \rho_d}{\partial t}+\frac{1}{\sqrt{G}}\left[\frac{\partial(\sqrt{G}\rho_d u^j)}{\partial x^j}\right]=0,$$

$$\frac{\partial \rho_d u^i}{\partial t}+\frac{1}{\sqrt{G}}\frac{\partial}{\partial x^j}\left[\sqrt{G}(\rho_d u^i u^j+G^{ij}P)\right]=F^i_H+F^i_M+F^i_C+F^i_g,$$

$$\frac{\partial \rho_d \theta'}{\partial t}+\frac{1}{\sqrt{G}}\left[\frac{\partial(\sqrt{G}\rho_d \theta' u^j)}{\partial x^j}\right]=-\rho_d\mathbf{u}\cdot\nabla\bar{\theta},$$

$$\frac{\partial \rho_d q_k}{\partial t}+\frac{1}{\sqrt{G}}\left[\frac{\partial(\sqrt{G}\rho_d q_k u^j)}{\partial x^j}\right]=0,$$

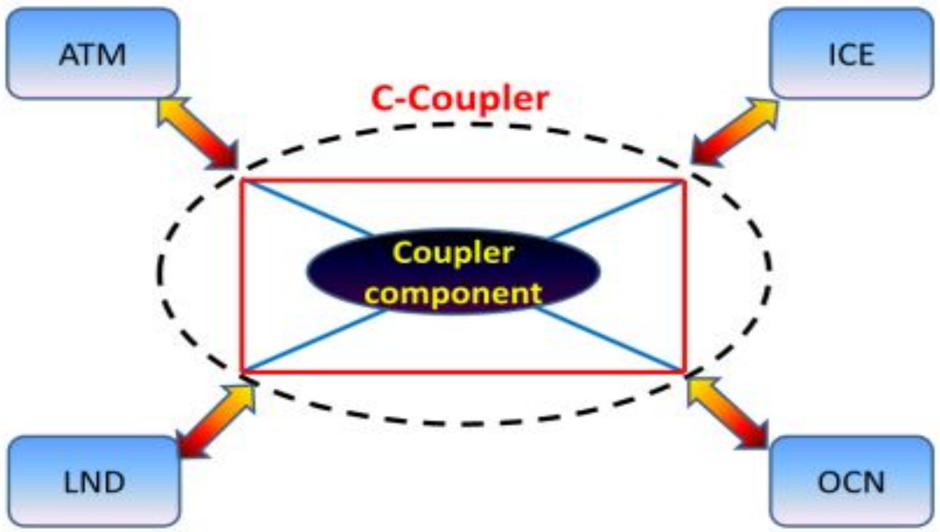


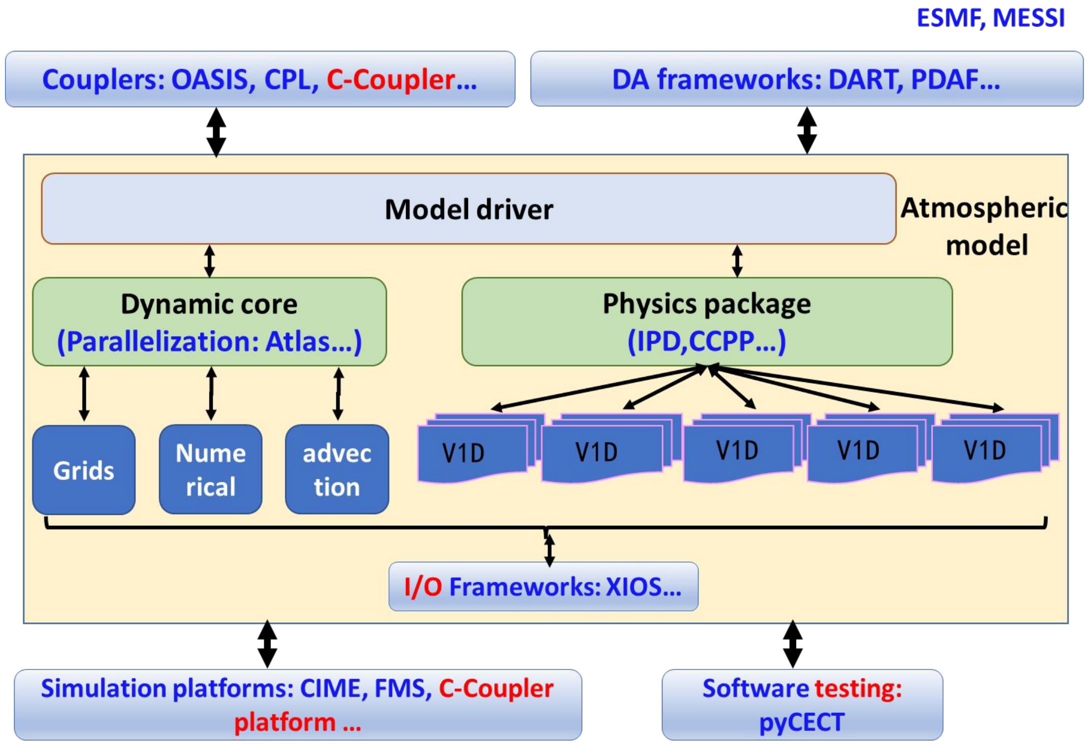
Summary of the main features of MCV model

Model aspects	MCV model
Governing equation	A fully compressible flux-form
Prognostic variables	$(\rho_d', \rho_d u^{\xi}, \rho_d u^{\eta}, \rho_d w, \rho_d\theta', \rho_d r_X)$
Horizontal discretization	4th order MCV scheme
Vertical discretization	2nd or 3rd conservative finite difference
Horizontal coordinate	$[\alpha,\beta]\in\left[-\frac{\pi}{4},\frac{\pi}{4}\right]$ or $[\lambda,\varphi]$
Vertical coordinate	Hybrid terrain-following coordinate
Horizontal staggering	Co-located
Vertical staggering	Co-located
Time marching	3rd order IMEX Runge-Kutta
Advection	A horizontally MCV-BGS/WENO, Vertically PRM

Earth System Approach

- Increased software complexity due to the introduction of more Earth system components
- Increased difficulty in software maintenance and integration development
- Is it possible to use a common software framework to support the Earth system model ?




ESMF, MESSI

Couplers: OASIS, CPL, C-Coupler…

DA frameworks: DART, PDAF…

Model driver

Atmospheric model

Dynamic core
(Parallelization: Atlas…)

Physics package
(IPD,CCPP…)

Grids

Numerical

advection

V1D

V1D

V1D

V1D

V1D

I/O Frameworks: XIOS…

Simulation platforms: CIME, FMS, C-Coupler platform …

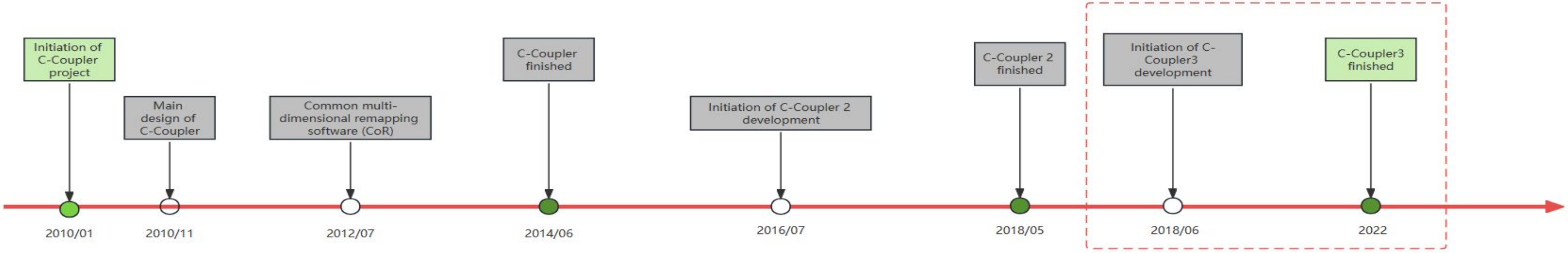
Software testing: pyCECT

02

Coupler-based parallel software framework

The choice of C-Coupler

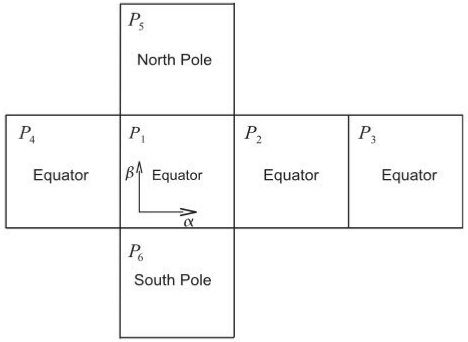
- The C-Coupler, developed at Tsinghua University since 2010, was initially focused on coupling the components of Earth system model.
- Widely used in China for developing coupled models.
- Much of the functionality of the underlying C-Coupler implementation can be used in a parallel software framework for the component model of ESM.
  - General grid management, supporting structured and unstructured grids
  - Parallel decomposition management
  - Non-blocking data transfer
  - Adaptive restart capability
  - …
- Emerging needs for parallel frameworks for atmospheric model development
  - The spatial and temporal resolution of short-wave radiation is different from that of the other components
  - Different resolutions between dynamical core and physical parameterization schemes

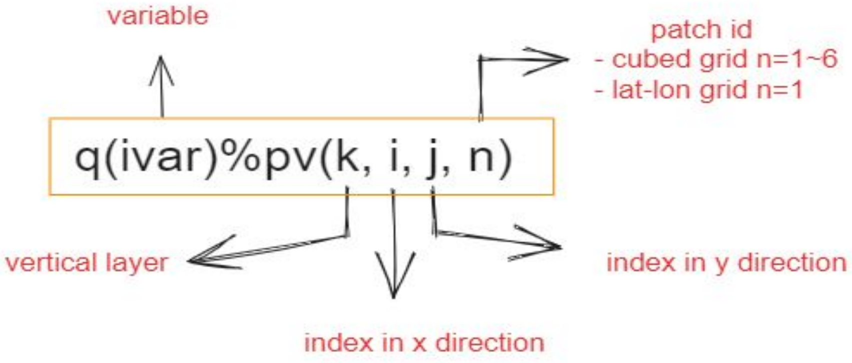


Requirements-driven development from the MCV model
- A new parallel triangulation algorithm for reducing initialization overheads
- Common halo exchange library
- Common I/O framework
- Common module-integration framework

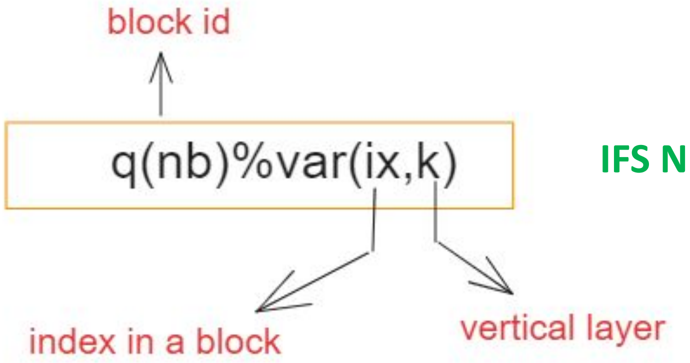
Parallel MCV

- Hybrid MPI + OpenMP
  - Horizontal 2D parallel decomposition of MPI process parallelism
  - Loop-level OpenMP thread parallelism
- Automatic parallel decomposition according to the given number of processes
  - No. of processes should be a multiple of 2, 3, or 5
- Different data structures between the dynamical core and physical package
  - The data structures contain patch information in dynamics allow sequential run.
  - Transformation of data structures required in the physics-dynamics coupling interface





data structure in dynamical core

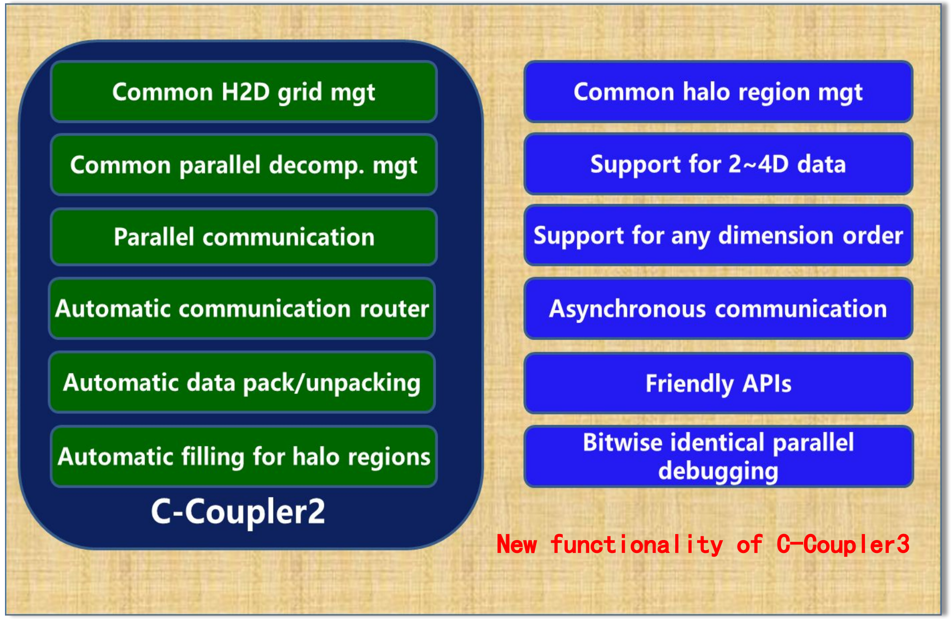


IFS NPROMA-like

data structure in physical package

Halo-exchange of MCV

- Halo exchange is a core feature of the parallel software framework.
- Add halo exchange functions in C-Coupler 3 to support MCV.
- As a global/regional unified model, MCV's halo exchange needs to support a global cubed-sphere grid and a regional latitude/longitude grid.
  - The communication pattern within each patch of the cubed sphere grid is the same as the latitude/longitude grid
  - In addition, global cubed-sphere grid need to deal with boundary conditions between patches, involving scalar and vector one-dimensional interpolated reconstructions.

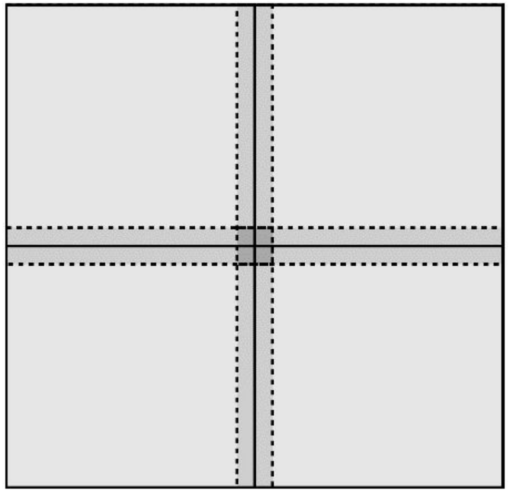


C-Coupler2

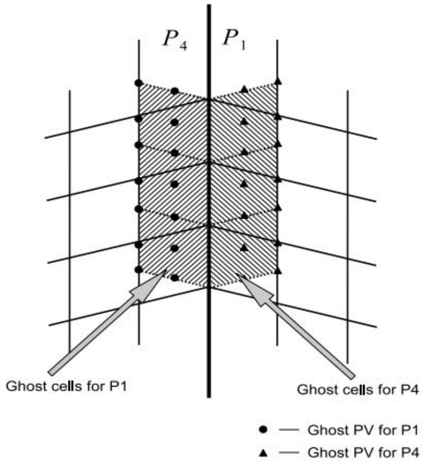
- Common H2D grid mgt
- Common parallel decomp. mgt
- Parallel communication
- Automatic communication router
- Automatic data pack/unpacking
- Automatic filling for halo regions

New functionality of C-Coupler3

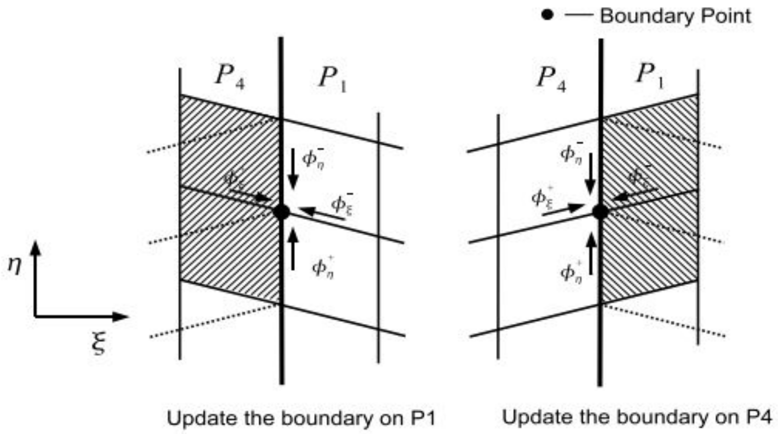
- Common halo region mgt
- Support for 2~4D data
- Support for any dimension order
- Asynchronous communication
- Friendly APIs
- Bitwise identical parallel debugging



Inside one patch of cubed-sphere or regional lat-lon grid



Ghost cells for cubed grid's patch boundary

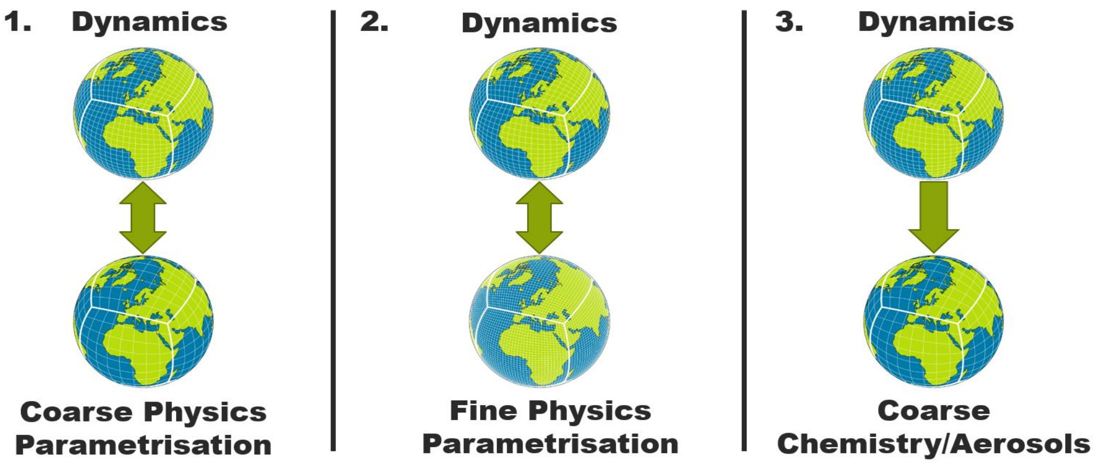


Common boundary between adjacent patches

Coupler-based Physics-Dynamics Coupling

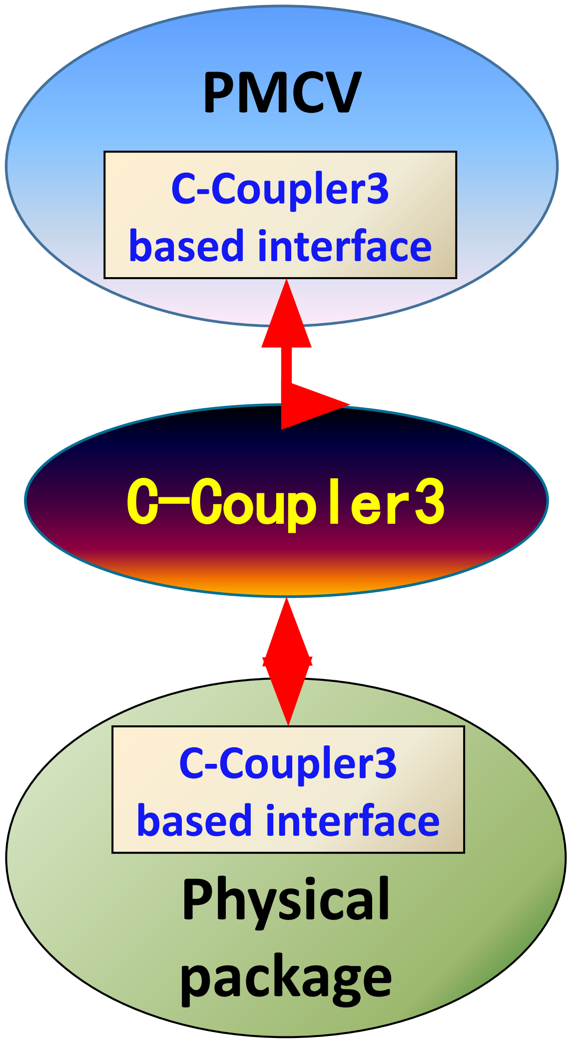
Unlike traditional implementations of physics-dynamics coupling interface, MCV use C-Coupler to connect the dynamical core to physical parameterization:

- Supports conversion of different data structures between dynamics and physics (k,i,j,n) □□ (blocksize,k,nblocks)
- Allows for different parallel decompositions and different resolutions between dynamics and physics (e.g. Radiation)
- But it also introduces some additional performance overheads



Key aspect is how fields are communicated between meshes

Bendall and Brown, 2022


PMCV

C-Coupler3 based interface

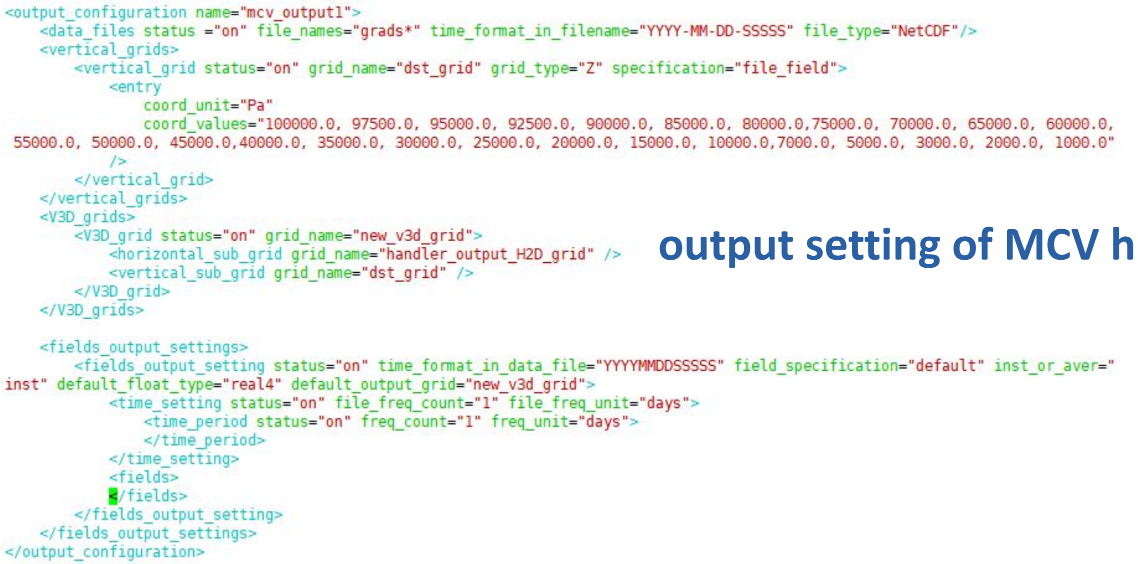
C-Coupler3

C-Coupler3 based interface

Physical package

Parallel I/O

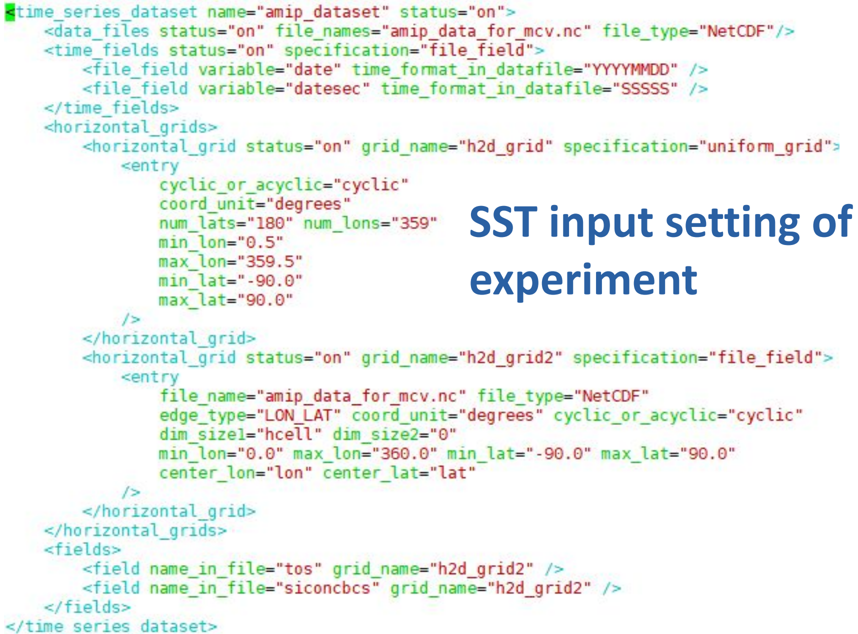
- Developed a parallel I/O scheme for MCV model based on CIOFC/C-Coupler 3
- Support irregular output time series
- Online parallel post-processing with dynamic 3-D interpolation
- Automatic input of time-series data: automatic input field data at different time points and automatic time interpolation
- Friendly and flexible XML configuration files



```xml
<output_configuration name="mcv_output1">
    <data_files status ="on" file_names="grads*" time_format_in_filename="YYYY-MM-DD-SSSSS" file_type="NetCDF"/>
    <vertical_grids>
        <vertical_grid status="on" grid_name="dst_grid" grid_type="Z" specification="file_field">
            <entry
                coord_unit="Pa"
                coord_values="100000.0, 97500.0, 95000.0, 92500.0, 90000.0, 85000.0, 80000.0,75000.0, 70000.0, 65000.0, 60000.0,
 55000.0, 50000.0, 45000.0,40000.0, 35000.0, 30000.0, 25000.0, 20000.0, 15000.0, 10000.0,7000.0, 5000.0, 3000.0, 2000.0, 1000.0"
            />
        </vertical_grid>
    </vertical_grids>
    <V3D_grids>
        <V3D_grid status="on" grid_name="new_v3d_grid">
            <horizontal_sub_grid grid_name="handler_output_H2D_grid" />
            <vertical_sub_grid grid_name="dst_grid" />
        </V3D_grid>
    </V3D_grids>

    <fields_output_settings>
        <fields_output_setting status="on" time_format_in_data_file="YYYYMMDDSSSSS" field_specification="default" inst_or_aver="
inst" default_float_type="real4" default_output_grid="new_v3d_grid">
            <time_setting status="on" file_freq_count="1" file_freq_unit="days">
                <time_period status="on" freq_count="1" freq_unit="days">
                </time_period>
            </time_setting>
            <fields>
            </fields>
        </fields_output_setting>
    </fields_output_settings>
</output_configuration>
```

output setting of MCV history files

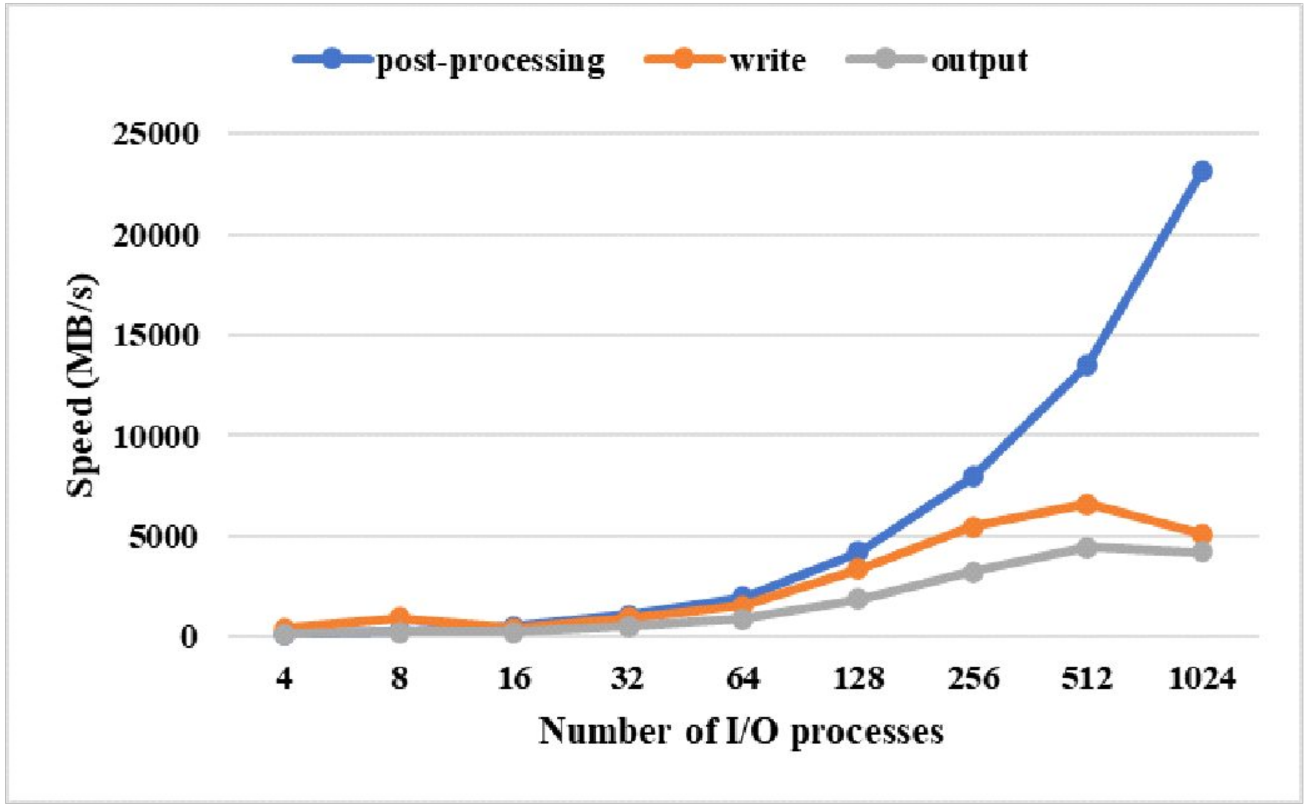


```xml
<time_series_dataset name="amip_dataset" status="on">
    <data_files status="on" file_names="amip_data_for_mcv.nc" file_type="NetCDF"/>
    <time_fields status="on" specification="file_field">
        <file_field variable="date" time_format_in_datafile="YYYYMMDD" />
        <file_field variable="datesec" time_format_in_datafile="SSSSS" />
    </time_fields>
    <horizontal_grids>
        <horizontal_grid status="on" grid_name="h2d_grid" specification="uniform_grid">
            <entry
                cyclic_or_acyclic="cyclic"
                coord_unit="degrees"
                num_lats="180" num_lons="359"
                min_lon="0.5"
                max_lon="359.5"
                min_lat="-90.0"
                max_lat="90.0"
            />
        </horizontal_grid>
        <horizontal_grid status="on" grid_name="h2d_grid2" specification="file_field">
            <entry
                file_name="amip_data_for_mcv.nc" file_type="NetCDF"
                edge_type="LON_LAT" coord_unit="degrees" cyclic_or_acyclic="cyclic"
                dim_size1="hcell" dim_size2="0"
                min_lon="0.0" max_lon="360.0" min_lat="-90.0" max_lat="90.0"
                center_lon="lon" center_lat="lat"
            />
        </horizontal_grid>
    </horizontal_grids>
    <fields>
        <field name_in_file="tos" grid_name="h2d_grid2" />
        <field name_in_file="siconcbcs" grid_name="h2d_grid2" />
    </fields>
</time_series_dataset>
```

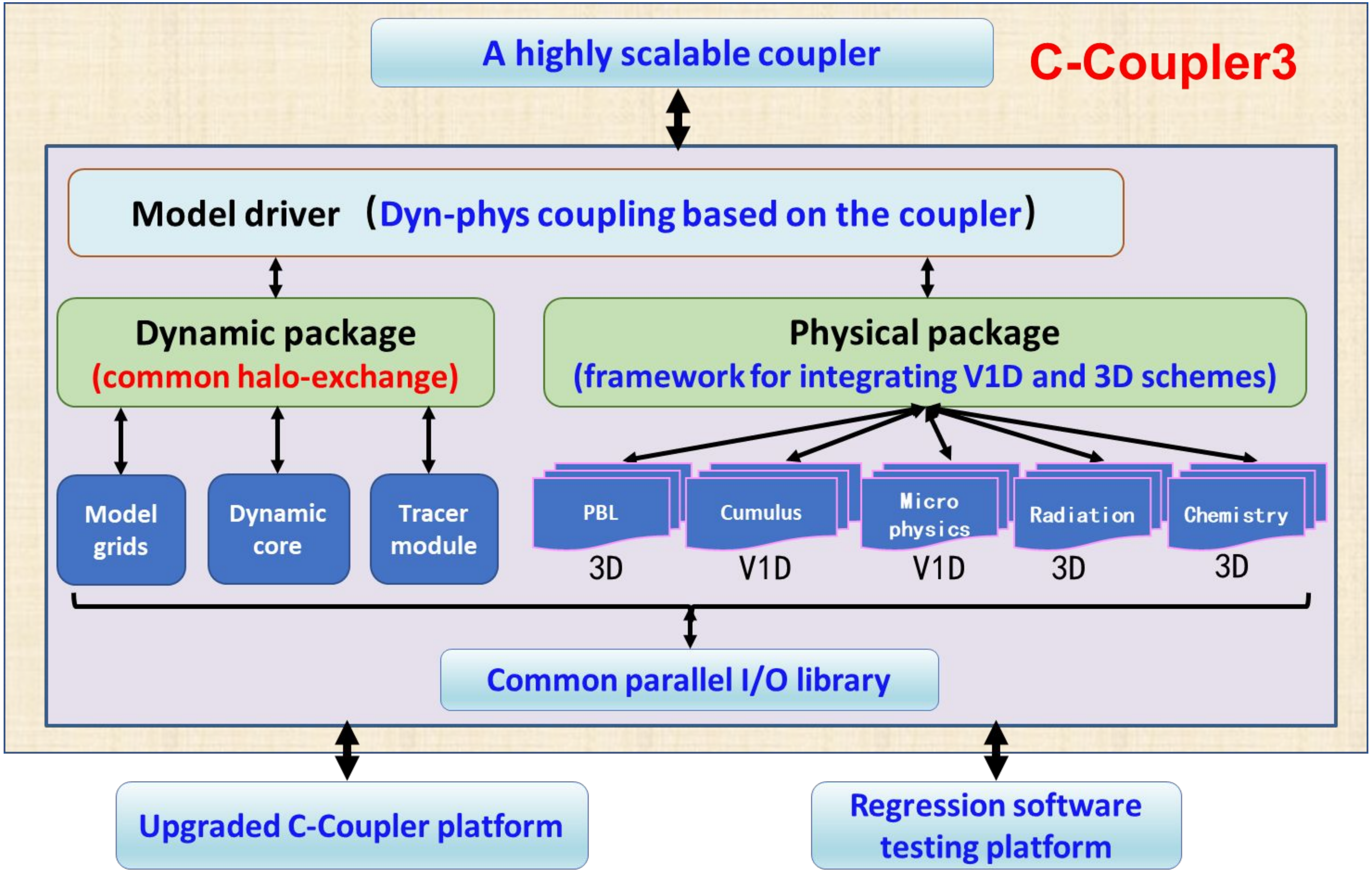
SST input setting of AMIP experiment

Parallel I/O cont.

The MCV model at a global 12.5km resolution
Size of a 3-D field in data file: 3.3GB



Software structure of MCV



A highly scalable coupler

C-Coupler3

Model driver （Dyn-phys coupling based on the coupler）

Dynamic package

(common halo-exchange)

Physical package

(framework for integrating V1D and 3D schemes)

Model grids

Dynamic core

Tracer module

PBL

3D

Cumulus

V1D

Micro physics

V1D

Radiation

3D

Chemistry

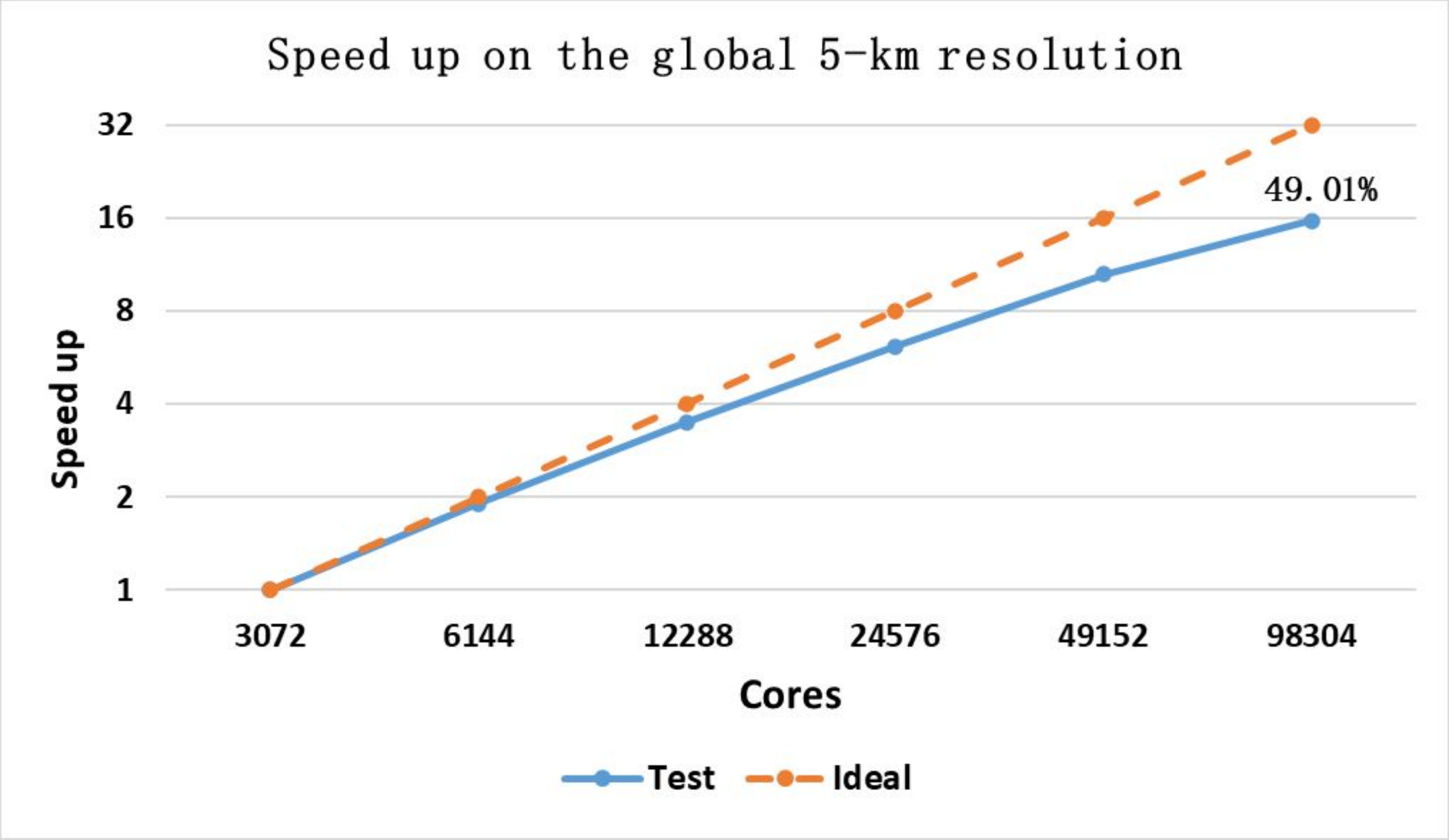
3D

Common parallel I/O library

Upgraded C-Coupler platform

Regression software testing platform

Scalability of the MCV model



HPC in National Supercomputing Center in Jinan	
CPU	2*Intel Xeon Gold 6258R 28C 2.7GHz
Memory/node	192GB
Network	Infiniband EDR 100Gb/s
PFS	Lustre
Software	Intel C/Fortran, Intel MPI

03

Progresses in optimizations

Resource requirement for developing MCV model

Test MCV model on the new CMA HPC system

	Horizontal resolution	Vertical level	Time step(sec)	Forecast length(hour)	MPI task	OpenMP thread	Wall time(sec)
Global	5	60	8	1	24576	2	774
Global	9	60	15	6	12288	2	1480
China	9	60	12	6	3072	1	883

	Cores	Wall time		CPU-hours
glob5km10d	49152	185760 sec	2-03:36:00	2,536,244
glob9km10d	24576	142080 sec	1-15:28:00	969,933
chn9km24h	3072	3532 sec	00:48:52	3,014

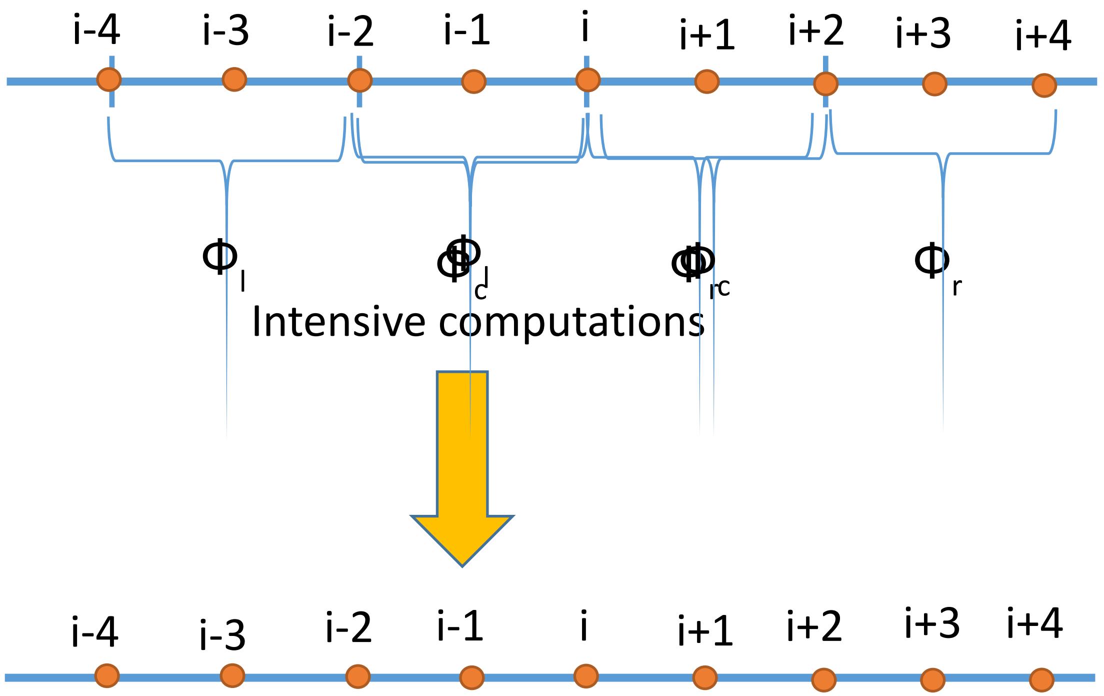
Total:3,509,191 CPU-hours,
30% of the whole new HPC system

- Huge resource demand for research and development
- Great efforts need to improve the computing performance

From Dr. Huadong Xiao's slide

Optimization 1: Derivative Operator

Derivative Operator: $\frac{\partial \phi}{\partial x} / \frac{\partial f}{\partial x}$

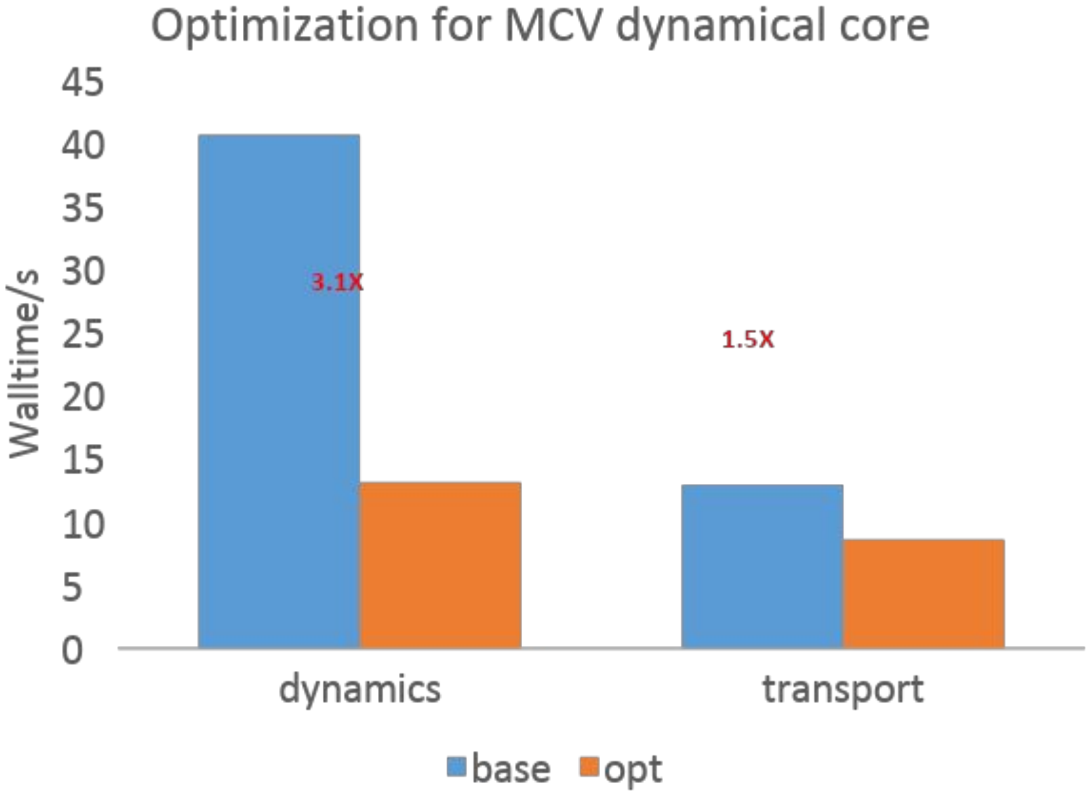


Intensive computations

Φ(ids:ide) / flux(ids:ide)

Computation once

Space for Time Optimization Strategy



Optimization for MCV dynamical core

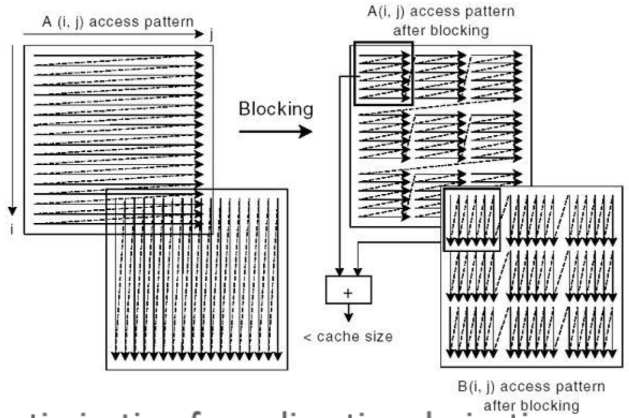
Configuration : dx=25km, 768 tasks

Optimization 2: loop tiling for tracer transport

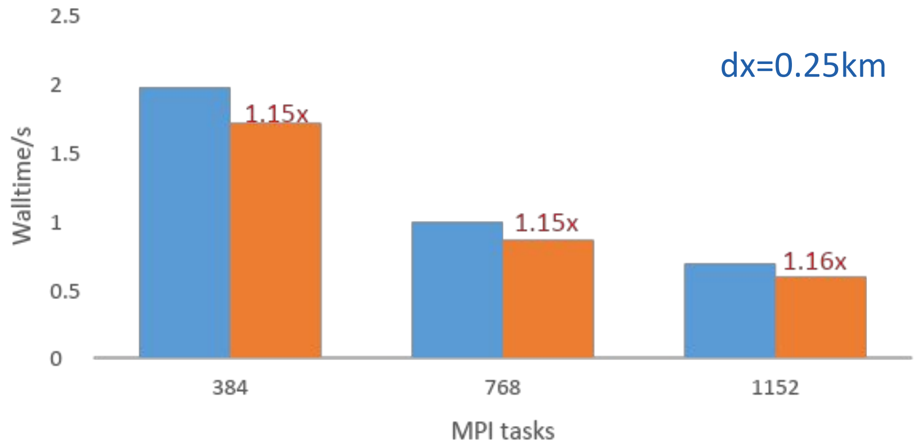
- Performance analysis reveals that the scalability of tracer transport is not good
- One of the reasons for this is the inefficiency of cache access when calculating horizontal derivatives
- Improve data locality and cache access efficiency using loop tiling technique

```
do n=fms,fme   ! Loop in patches
   do j=jds,jde    ! Loop in y direction
      do k=kvs,kve          ! loop in vertical direction
         do ivar=1,nadvect    ! Loop in advection tracers
            do i=ims,ime          ! Loop in x direction
              qmtx(i,ivar)=qmoist(ivar)%pv(k,i,j,n)
              …
```

```
do n=fms,fme
   do j=jds,jde
      do t=0, tile_size-1
         ks=tile_cap*t + kvs
         ke=max(tile_cap*(t+1)-1+kvs, kve)
         tile_count=ke-ks+1
         do ivar=1,nadvect
            do i=ims,ime
               do k=1,tile_count
                 qmtx(k,i,ivar)=qmoist(ivar)%pv(ks+k-1,i,j,n)
                 …
```



loop tiling optimization for x-direction derivatives in transport



04

Summary and future work

Summary

- As the multi-moment method has a large computation-to-communication ratio, the model based on it has good computational scalability.
- A coupler-based parallel software framework is a meaningful practice used for unified software infrastructures in developing ESM.
- There are significant computational optimization challenges in meeting the time-to-solution goal of the MCV model.
- Doing computational optimization on a rapid development model is not an easy work.

Future work

- Advancing the computational efficiency of the MCV model: MPI + “X”

(1) Continue to focus on single-core computing performance of MCV

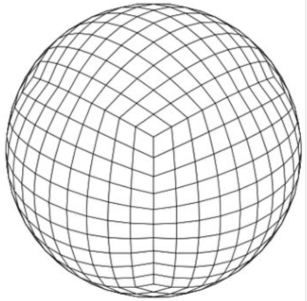
  - Vectorization (expect 2~4x speedup)
  - Refactoring code, loop-tiling optimization for cache-friendly (expect 1.5x speedup)
  - Change double precision to single precision (expect 2x speedup)

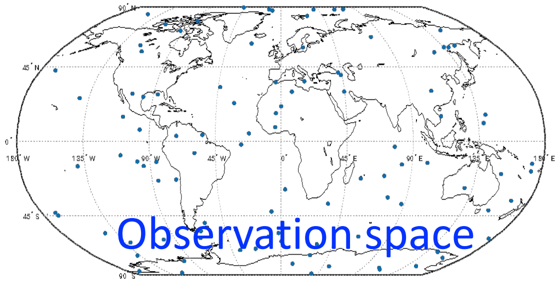
(2) Inter-node communication optimization

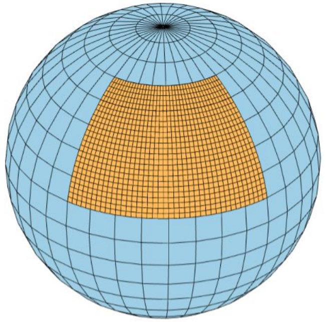
  - Optimize halo-exchange implementation
  - Try asynchronous communication of halo-exchange

(3) Another team is porting MCV to GPU platform (expect 3~5x speedup)

Regional & global unified Data Assimilation System


Global cubed-sphere grid


Observation space


Regional LAT-LON grid

A unified DA framework suitable for both Cubed sphere grid and latitude-longitude grid

- Global spectral filtering and regional recursive filtering techniques
- Construction of multiscale background error covariance matrix.
- Gravity wave noise control technology

Efficient tangent-linear and adjoint models of non-hydrostatic atm. model in cubed-sphere space

- TL/AD of Dynamic core and main physics modules
- Linkage between linear models and variational assimilation framework
- Common framework for highly scalable parallelization

Regional non-hydrostatic 3DVAR and Rapid Update Cycle

- Non-hydrostatic control variables
- Modelling and estimation of Km-scale background error covariances
- Use of spatially dense and frequent-in-time Obs (radar, etc)
- Noise control during rapid update cycle

THANKS !

CEMC

Email: jiangqg@cma.cn

References

- MCV method

  - Chen, C., Li, X., Xiao, F., Shen, X., 2023. A nonhydrostatic atmospheric dynamical core on cubed sphere using multi-moment finite-volume method. Journal of Computational Physics 473, 111717.
  - Tang, J., Chen, C., Shen, X., Xiao, F., Li, X., 2022. A three-dimensional positivity-preserving and conservative multimoment finite-volume transport model on a cubed-sphere grid. Quarterly Journal of the Royal Meteorological Society 148, 3622–3638.
  - Chen, C., Li, X., Shen, X., Xiao, F., 2014. Global shallow water models based on multi-moment constrained finite volume method and three quasi-uniform spherical grids. Journal of Computational Physics 271, 191–223.
  - Li, X., Chen, C., Shen, X., Xiao, F., 2013. A Multimoment Constrained Finite-Volume Model for Nonhydrostatic Atmospheric Dynamics. Monthly Weather Review 141, 1216–1240.
  - Ii, S., Xiao, F., 2009. High order multi-moment constrained finite volume method. Part I: Basic formulation. Journal of Computational Physics 228, 3669–3707.

- C-Coupler

  - Liu, L., Sun, C., Yu, X., Yu, H., Jiang, Q., Li, X., Li, R., Wang, B., Shen, X., Yang, G., 2023. C-Coupler3.0: an integrated coupler infrastructure for Earth system modelling. Geoscientific Model Development 16, 2833–2850.
  - Yu, X., Liu, L., Sun, C., Jiang, Q., Zhao, B., Zhang, Z., Yu, H., Wang, B., 2022. CIOFC1.0: a common parallel Input/Output framework based on C-Coupler2.0. Geoscientific Model Development Discussions 2022, 1–58.
  - Liu, L., Zhang, C., Li, R., Wang, B., Yang, G., 2018. C-Coupler2: a flexible and user-friendly community coupler for model coupling and nesting. Geosci. Model Dev. 11, 3557–3586.

- Others

  - Bendall, T., Brown, A., 2022. Physics-Dynamics-Chemistry coupling with components of different resolutions in LFRic. PDC22 workshop, Princeton University, USA, Available at: https://intranet.gfdl.noaa.gov/Members/Lucas.Harris/physics-dynamics-coupling-workshop-2022/bendall_multires_pdc_workshop_2022.pptx(Last Access: 20 Oct 2022)