https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

Geoscientific Model Development
Discussions
Open Access

GPU-HADVPPM4HIP V1.0: higher model accuracy on China's domestically GPU-like accelerator using heterogeneous compute interface for portability (HIP) technology to accelerate the piecewise parabolic method (PPM) in an air quality model (CAMx V6.10)

Kai Cao[1], Qizhong Wu[1,5], Lingling Wang[2], Hengliang Guo[3], Nan Wang[2], Huaqiong Cheng[1,5], Xiao Tang[4], Lina Liu[3], Dongqing Li[1], Hao Wu[3], and Lanning Wang[1,5]

[1]College of Global Change and Earth System Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[2]Henan Ecological Environmental Monitoring Centre and Safety Center, Henan Key Laboratory of Environmental Monitoring Technology, Zhengzhou 450008, China

[3]National Supercomputing Center in Zhengzhou, Zhengzhou, 450001, China

[4]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Science, Beijing 100029, China

[5]Joint Center for Earth System Modeling and High Performance Computing, Beijing Normal University, Beijing, 100875, China

Correspondence to: Qizhong Wu (wqizhong@bnu.edu.cn); Lingling Wang(928216422@qq.com); Lanning Wang (wangln@bnu.edu.cn)

Abstract. The graphics processing units (GPUs) are becoming a compelling acceleration strategy for geoscience numerical model due to their powerful computing performance. In this study, AMD's heterogeneous compute interface for portability (HIP) was implemented to port the GPU acceleration version of the Piecewise Parabolic Method (PPM) solver (GPU-HADVPPM) from the NVIDIA GPUs to China' s domestically GPU-like accelerators as GPU-HADVPPM4HIP, and further introduced the multi-level hybrid parallelism scheme to improve the total computational performance of the HIP version of CAMx (CAMx-HIP) model on the China' s domestically heterogeneous cluster. The experimental results show that the acceleration effect of GPU-HADVPPM on the different GPU accelerator is more obvious when the computing scale is larger, and the maximum speedup of GPU-HADVPPM on the domestic GPU-like accelerator is 28.9 times. The hybrid parallelism with a message passing interface (MPI) and HIP enables achieve up to 17.2 times speedup when configure 32 CPU cores and GPU-like accelerators on the domestic heterogeneous cluster. And the OpenMP technology is introduced to further reduce the computation time of CAMx-HIP model by 1.9 times. More importantly, by comparing the simulation results of GPU-HADVPPM on NVIDIA GPUs and domestic GPU-like accelerators, it is found that the simulation results of GPU-HADVPPM on domestic GPU-like accelerators have less difference than the NVIDIA GPUs, and the reason for this difference may be related to the fact that the NVIDIA GPU sacrifices part of the accuracy for improved computing performance. All in all, the domestic GPU-like accelerators are more accuracy for scientific computing in the field of geoscience numerical models. Furthermore, we also exhibit that the data transfer efficiency between CPU and GPU has an important impact on heterogeneous computing, and point out that optimizing the data transfer efficiency between CPU and GPU is one of the important directions to improve the computing efficiency of geoscience numerical models in heterogeneous clusters in the future.

1. Introduction

Over the recent years, GPUs have become an essential part of providing processing power for high performance computing (HPC) application, and heterogeneous supercomputing based on CPU processors and GPU accelerators has become the trend of global advanced supercomputing development. The 61st edition of the top 10 list, released in June 2023, reveals that 80% of advanced supercomputers adopt the heterogeneous architectures (https://www.top500.org/lists/top500/2023/06/, last access: 20 October 2023), and the Frontier system equipped with AMD Instinct MI250X GPU at the Oak Ridge National Laboratory remains the only true exascale machine with the High-Performance Linpack benchmark (HPL) score of 1.194 Exaflop/s (https://www.top500.org/news/frontier-remains-sole-exaflop-machine-and-retains-top-spot-improving-upon-its-previous-hpl-score/, last access: 20 October 2023). It is worth noting that in addition to the second-place Fugaku supercomputer using a general-purpose CPU architecture, the third-ranked LUMI system also uses AMD Instinct MI250X GPUs as accelerators and its HPL score reaches 309.1 PFlop/s. The much-watched AMD Instinct MI250X GPU achieves

https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

95.7 TFlop/s for peak double precision matrix performance (https://www.amd.com/en/products/server-accelerators/Instinct-mi250x, last access: 20 October 2023), and its performance is 2.7 times that of EARTH-SIMULATOR which is the top 1 supercomputer in 2003. How to realize the large-scale parallel computing and improve the computational performance of geoscience numerical models on the GPU has become one of the significant directions for the future development of numerical models.

In terms of the heterogeneous porting for ~~the~~ atmospheric chemical models, many scholars have carried out research on chemical modules. For example, Sun et al. (2018) used CUDA technology to port the second-order Rosenbrock solver of chemistry module of CAM4-Chem to NVIDIA Tesla K20X GPU and achieved up 11.7x speedup for computation alone. Alvanos and Christoudias (2017) developed a software that automatically generates CUDA kernels to solve chemical kinetics equation in the chemistry module for the global climate model ECHAM/MESSy Atmospheric Chemistry (EMAC) and performance evaluation shows a 20.4x speedup for the kernel execution. Linford et al. (2011) presented the Kinesthetic PreProcessor: Accelerated (KPPA) to generate the chemical mechanism code in CUDA language which can be implemented on NVIDIA Tesla C1060 GPU. The KPPA-generated SAPRC'99 mechanism from CMAQ model achieved a maximum speedup of 13.7x and KPPA-generated RADM2 mechanism from WRF-chem model achieved an 8.5x speedup over the serial implementation. Horizontal advection module for the atmospheric chemical models, Cao et al. (2023) used the Fortran-C-CUDA C scheme and implemented a series of optimizations, including reduce the CPU–GPU communication frequency, optimize the GPU memory access, and thread and block co-indexing, to increase the computational efficiency of the HADVPPM advection solver in the CAMx model by 18.8 times on the NVIDIA Tesla V100 GPU.

The CUDA technology was implemented to carry out heterogeneous porting for the atmospheric chemical models from the CPU processors to different NIVIDA GPU accelerators. In this study, the Heterogeneous-computing Interface for Portability (HIP) interface was introduced to implement the porting of GPU-HADVPPM from the NVIDIA GPU to the China' s domestically GPU-like accelerators based on the research of Cao et al. (2023). First, we compared the simulation result of Fortran version CAMx model with CUDA version of CAMx (CAMx-CUDA) and CAMx-

https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

HIP model which were coupled with CUDA and HIP version of GPU-HADVPPM program, respectively. And then, the computing performance of GPU-HADVPPM programs on different GPUs are compared. Finally, we tested total coupling performance of CAMx-HIP model with multi-level hybrid parallelization on the China’ s domestically heterogeneous cluster.

2. Model and experimental platform

2.1. The CAMx model description and configuration

The Comprehensive Air Quality Model with Extensions version 6.10 (CAMx v6.10; ENVIRON, 2014) is a state-of-the-art air quality model which simulates the emission, dispersion, chemical reaction, and removal of the air pollutants on a system of nested three-dimensional grid boxes (https://www.camx.com/, last access: last access: 20 October 2023). The Eulerian continuity equation is expressed as shown Cao et al. (2023), the first term on the right-hand side represents horizontal advection, the second term represents net resolved vertical transport across an arbitrary space and time varying height grid, and the third term represents turbulent diffusion on the sub-grid scale. Pollutant emission represents both point source emissions and grided source emissions. Chemistry is treated by solving a set of reaction equations defined by specific chemical mechanisms. Pollutant removal includes both dry deposition and wet scavenging by precipitation.

In terms of the horizontal advection term on the right-hand side, this equation is solved using either the Bott (1989) scheme or the Piecewise Parabolic Method (PPM) (Colella and Woodward, 1984; Odman and Ingram, 1996) scheme. The PPM horizontal advection scheme (HADVPPM) was selected in this study because it provides higher accuracy with minimal numerical diffusion. The other numerical scheme selected during the CAMx model running are listed in Table S1. As described by Cao et al. (2023), the -fp-model precise compile flag which can force the compiler to use the vectorization of some computation under value safety is 41.4% faster than -mieee-fp compile flag which comes from the Makefile of the official CAMx version with the absolute errors of the simulation results are less than ±0.05 ppbV. Therefore, the -fp-model precise compile flag was selected when compiling the CAMx model in this research.

2.2. CUDA and ROCm introduction

Compute Unified Device Architecture (CUDA) (NVIDIA, 2020) is a parallel programming paradigm which was released in 2007 by NVIDIA. CUDA is a proprietary application programming interface (API) and as such is only supported on NVIDIA’s GPUs ~~that are based on Tesla Architecture~~. For the CUDA programming, it uses a programming language similar to standard C, which achieves efficient parallel computing of programs on NVIDIA GPUs by adding some keywords. In the previous study, CUDA technology was implemented to port the HADVPPM program from CPU to NVIDIA GPU (Cao et al., 2023).

Radeon Open Compute platform (ROCm) (AMD, 2023) is an open-source software platform developed by AMD in 2015 for HPC and hyperscale GPU computing. In general, ROCm for the AMD GPU is equivalent to CUDA for NVIDIA GPU. On the ROCm software platform, it uses the AMD’s HIP interface which is a C++ runtime API to allows developers to run programs on AMD GPUs. Table 1 shows the difference between the CUDA programming and HIP programming on the NVIDIA GPU and AMD GPU. In general, it is very similar between the CUDA and HIP programming and their code can be converted directly by replacing the character “cuda” with “hip” in the most cases. More information about HIP API can be available on https://rocm.docs.amd.com/projects/HIP/en/latest/index.html (last access: 20 October 2023). Similar to AMD GPU, developers can also use ROCM-HIP programming interface to implement programs running on the China’ s domestically GPU-like accelerator.

Table 1. The difference between the CUDA programming and HIP programming on the NVIDIA GPU and AMD GPU.

	CUDA programming	**HIP programming**
Header file	cuda_runtime.h	hip_runtime.h
Gets the number of compute-capable GPUs.	cudaGetDeviceCount	hipGetDeviceCount
Set device to be used for GPU executions.	cudaSetDevice	hipSetDevice
Allocates memory on the GPU.	cudaMalloc	hipMalloc
Copies data between CPU and GPU.	cudaMemcpy	hipMemcpy

Kernel function	mykernel<<< >>>	hipLaunchKernelGGL(mykernel)
Frees memory on the GPU.	cudaFree	hipFree

2.3. Hardware components and software environment of the testing system

Table 2 listed four GPU clusters which are conducted the experiments, two NVIDIA heterogeneous clusters which have the same hardware configuration as Cao et al. (2023) and two China’ s domestically heterogeneous clusters newly used in this research. The NVIDIA K40m cluster is equipped with two 2.5 GHz 16 cores Intel Xeon E5-2682 v4 CPU and one NVIDIA Tesla K40m GPU. Each NVIDIA Tesla K40m GPU accelerator has 2880 CUDA cores with 12 GB of video memory. The NVIDIA V100 cluster contains two 2.7 GHz 24 cores Intel Xeon Platinum 8168 processors and eight NVIDIA Tesla V100 GPU accelerators. Each NVIDIA Tesla V100 GPU accelerator is configured with 5120 CUDA cores and 16 GB video memory.

For the China’ s domestically heterogeneous cluster A (domestic cluster A), each compute node contains a 2.0 GHz China’ s domestically CPU processor A of 32 cores (domestic CPU processor A) and four China’ s domestically GPU-like accelerator A (domestic GPU-like accelerator A). Each CPU processor A has 32 cores with 4 Non-Uniform Memory Access nodes, each NUMA node has 8 X86 based processors. The GPU-like accelerator A has 64 compute unit, for totaling 60 threads on each compute unit. The China’ s domestically heterogeneous cluster B (domestic cluster B) is the next generation of cluster A, and its CPU and GPU hardware have been upgraded, especially the data transfer bandwidth between CPU and GPU. The CPU and GPU configuration scheme on the cluster B is the same as the cluster B, with one 2.5 GHz China’ s domestically CPU processor B (domestic CPU processor B) on a single node equipped with four China’ s domestically GPU-like accelerator B (domestic GPU-like accelerator B). The domestic GPU-like accelerator B also contains 64 compute units with 128 threads each.

In term of the software environment, the Intel Toolkit (including compiler and MPI library) version 2021.4.0, 2019.1.144, and 2021.3.0 are employed for compiling on Intel CPU and China’ s domestically series CPU, respectively. The drivers and libraries of NVIDIA Tesla K40m and V100 GPU accelerator, domestic GPU-like accelerator A and B were CUDA version 10.2, CUDA version

https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

10.0, ROCm version 4.0.1/ DTK toolkit version 23.04, and DTK toolkit version 23.04. DTK toolkit, like ROCm, supports developers to develop GPU-like applications using HIP programming interface in C++ language.

Table 2. Configurations of NVIDIA K40m cluster, NVIDIA V100 cluster, China' s domestically cluster A, and China' s domestically cluster B.

	Hardware components	
	CPU	GPU
NVIDIA K40m cluster	Intel Xeon E5-2682 v4 CPU @2.5 GHz, 16 cores	NVIDIA Tesla K40m GPU, 2880 CUDA cores, 12 GB video memory
NVIDIA V100 cluster	Intel Xeon Platinum 8168 CPU @2.7 GHz, 24 cores	NVIDIA Tesla V100 GPU, 5120 CUDA cores, 16 GB video memory
China' s domestically cluster A	China' s domestically CPU processor A, 2.0GHz, 32 cores	China' s domestically GPU-like accelerator A, 3840 stream processors, 16 GB memory
China' s domestically cluster B	China' s domestically CPU processor B, 2.5GHz, 32 cores	China' s domestically GPU-like accelerator B, 8192 stream processors, 16 GB memory

	Software environment	
	Compiler and MPI	Programming model
NVIDIA K40m cluster	Intel Toolkit 2021.4.0	CUDA-10.2
NVIDIA V100 cluster	Intel Toolkit 2019.1.144	CUDA-10.0
China' s domestically cluster A	Intel Toolkit 2021.3.0	ROCm-4.0.1/ DTK-23.04
China' s domestically cluster B	Intel Toolkit 2021.3.0	DTK-23.04

3. Implementation details

This section mainly introduced the strategy of porting HADVPPM program from CPU to NVIDIA GPU and domestic GPU-like accelerator, as well as the proposed multi-level hybrid parallelism technology to make full use of computing resources.

https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

3.1. Porting the HADVPPM program from CPU to NVIDIA GPU and domestic GPU-like accelerator

Fig. 1 shows the heterogeneous porting process of HADVPPM from CPU to NVIDIA GPU and domestic GPU-like accelerator. First, the original Fortran code was refactored using standard C language. ~~And~~ then the CUDA and HIP technology were used to convert the standard C code into CUDA C and HIP C code to make it computable on the NIVIDA GPU and domestic GPU-like accelerator. To facilitate the portability of applications across different GPU platforms, ROCm provides hipify toolkits to help transcode. In this study~~ing~~, the ROCm HIP technology was used to implement the operation of GPU-HADVPPM on domestic GPU-like accelerator based on the CUDA version of GPU-HADVPPM which was developed by Cao et al. (2023). ~~During the compiling,~~ the HIP code was compiled using the “hipcc” compiler driver with the library flag “-lamdhip64”.

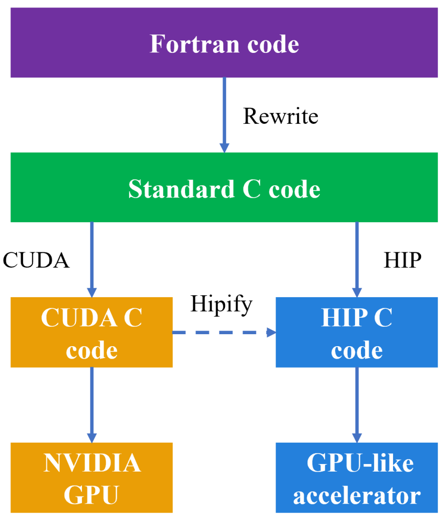

Figure 1. The heterogeneous porting process of HADVPPM Fortran code from CPU to NVIDIA GPU and domestic GPU-like accelerator.

3.2. Multi-level hybrid parallelization of CAMx model on heterogeneous platform

The original CAMx model running on the CPUs supports two types of parallelization (ENVIRON, 2014): (1) OpenMP (OMP), which supports multi-platform (e.g., multi-core) shared-memory programming in C/C++ and Fortran; (2) Message Passing Interface (MPI), which is a message passing interface standard for developing and running parallel applications on the distributed-memory computer cluster. In the original CAMx model, MPI+OMP hybrid parallel can be used to maximize computational efficiency.

In the previous study~~ing~~, Cao et al. (2023) adopt a parallel architecture with an MPI and CUDA (MPI+CUDA) hybrid paradigm to expand the parallel scale of CAMx-CUDA model in NVIDIA heterogeneous cluster. Adopting this strategy, GPU-HADVPPM can run on multiple NVIDIA GPUs. When the CUDA C code of GPU-HADVPPM is converted to HIP C code, GPU-HADVPPM with an MPI and HIP (MPI+HIP) heterogeneous hybrid programming technology can also run on multiple domestic GPU-like accelerators. The MPI and HIP hybrid parallel scheme can configure one GPU-like accelerator for each CPU process participating in the computation. However, the number of GPU-like accelerators in a single compute node is usually much smaller than the number of CPU cores in the super-large heterogeneous cluster. Therefore, in order to make full use of the remaining CPU computing resources, OMP technology is further introduced into the CAMx-HIP model which was coupled the HIP version of GPU-HADVPPM. In the framework of the multi-level hybrid parallelism, the horizontal advection module is accelerated by MPI and HIP technology, and the other modules are accelerated by MPI and OMP.

4. Results and evaluation

The coupling performance experiments of CUDA and HIP version GPU-HADVPPM were conducted in this section. First, we compared the simulation result of Fortran version CAMx model with CAMx-CUDA and CAMx-HIP model which were coupled with CUDA and HIP version of GPU-HADVPPM program, respectively. Then, the computing performance of GPU-HADVPPM programs on the NVIDIA GPU and domestic GPU-like accelerator are compared. Finally, we tested total coupling performance of CAMx-HIP model with multi-level hybrid parallelization on the domestic cluster A. For ease of description, the CAMx versions of the HADVPPM program written in Fortran, CUDA C and HIP C code are named F, CUDA and HIP, respectively.

4.1. Experimental setup

There are three test cases were used to evaluate the coupling performance of CUDA and HIP version GPU-HADVPPM. The experimental setup for the three test cases is shown in Table 3. The Beijing case (BJ) covers Beijing, Tianjin, and part of the Hebei Province with 145 × 157 grid boxes, and simulation of BJ case starts on 1 November, 2020. The Henan case (HN) mainly covers the Henan Province with 209 × 209 grid boxes. The starting date of simulation in HN case is 1 October, 2022. The Zhongyuan case (ZY) has the widest coverage of the three cases, with Henan Province as the center, covering the Beijing-Tianjin-Hebei region, Shanxi Province, Shaanxi Province, Hubei Province, Anhui Province, Jiangsu Province, and Shandong Province, with 531 × 513 grid boxes. ZY case started simulation on 4 January, 2023. All of the three performance test cases are 3km horizontal resolution, 48 hours of simulation, and 14 vertical model layers. The number of three-dimensional grid boxes in BJ, HN, and ZY cases are totally 318,710, 611,534 and 3,813,642, respectively. The meteorological fields inputting the different versions of the CAMx model in the three cases were provided by the Weather Research and Forecasting Model (WRF). In terms of emission inventories, the emission for BJ case is consistent with the Cao et al. (2023), HN case uses the Multi-resolution Emission Inventory for China (MEIC) and ZY case uses the emission constructed by Sparse Matrix Operator Kernel Emission (SMOKE) model in this study.

Table 3. The experimental setup for the BJ, HN, and ZY case.

	BJ	HN	ZY
Start date	November 1, 2020	October 1, 2022	1 January, 2023
Horizontal resolution	3km	3km	3km
Grid boxes	145 × 157 × 14	209 × 209 × 14	531 × 513 × 14
Meteorological fields	WRF	WRF	WRF
Emission	Cao et al. (2023)	MEIC	SMOKE

4.2. Error analysis

The hourly concentrations of four major species, i.e. O_3, PSO_4, CO, and NO_2, outputted by Fortran, CUDA, and HIP version of CAMx for the BJ case are compared to verify the results reasonableness before testing the computation performance. Fig. 2 present the four major species simulation results of three CAMx version, including Fortran version on the Intel E5-2682 v4 CPU, CUDA version on the NVIDIA K40m cluster and HIP version on the domestic cluster A, after 48 hours integration, as well as the absolute errors (AEs) of their concentrations. The species' spatial pattern of three CAMx versions on different platform are visually very consistent, and the AEs between the HIP and Fortran version is much smaller than the CUDA and Fortran version. For example, the AEs between the CUDA and Fortran version for O_3, PSO_4, and NO_2 are in the range of ± 0.04 ppbV, ± 0.02 $\mu g \cdot m^{-3}$, and ± 0.04 ppbV. And the AEs between the HIP and Fortran version for above the three species are fall into the range of ± 0.01 ppbV, ± 0.005 $\mu g \cdot m^{-3}$, and ± 0.01 ppbV. For CO, AEs is relatively large due to its high background concentration. However, the AEs between the HIP and Fortran versions is also less than that between the CUDA and Fortran versions where were in the range of ± 0.4 ppbV and ± 0.1 ppbV, respectively.

https://doi.org/10.5194/gmd-2023-222
Preprint. Discussion started: 17 January 2024

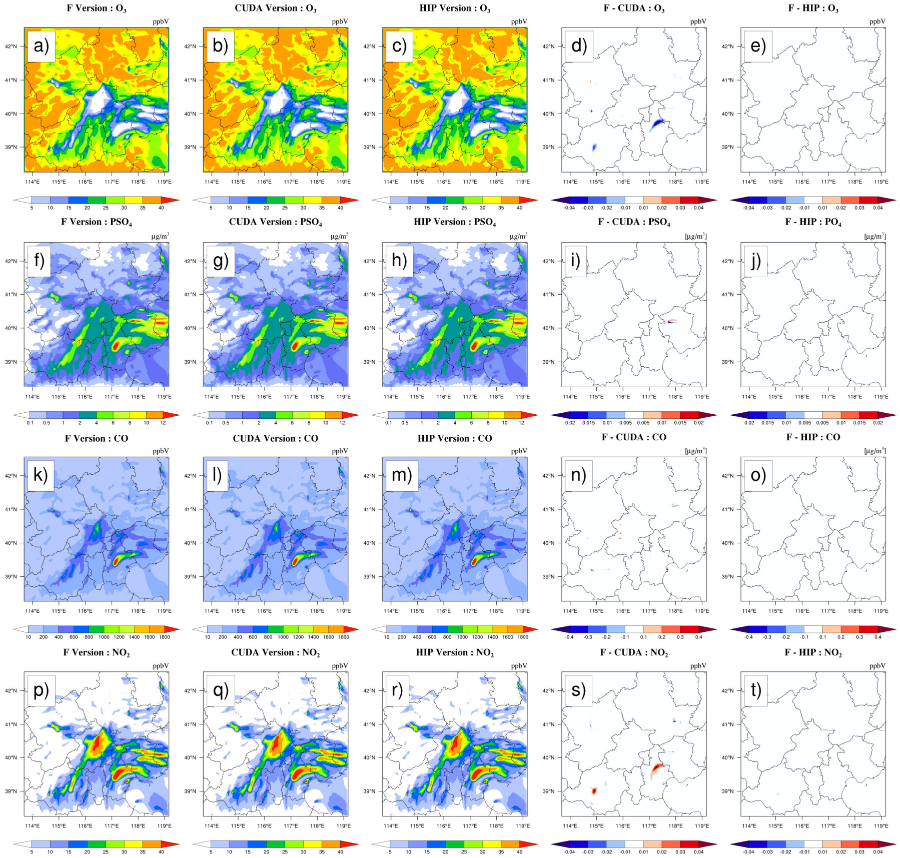

Figure 2. O$_3$, PSO$_4$, CO, and NO$_2$ concentrations outputted by the CAMx Fortran version on the Intel E5-2682 v4 CPU, CUDA version on the NVIDIA K40m cluster and HIP version on the domestic cluster A under the BJ case. Panels (a), (f), (k), and (p) are from the Fortran version of simulation results for four species. Panels (b), (g), (l), and (q) are from the CUDA version of simulation results for four species. Panels (c), (h), (m), and (r) are from the HIP version of simulation results for four species. Panels (d), (i), (n), and (s) are the AEs between the Fortran and CUDA versions. Panels (e), (j), (o), and (t) are the AEs between the Fortran and HIP versions.

Fig. 3 presents the boxplot of the relative errors (REs) in all grid boxes for the O$_3$, PSO$_4$, CO, and NO$_2$ during the 48 hours simulation under the BJ case. Statistically, the REs between the CUDA version on the NVIDIA K40m cluster and Fortran version on the Intel E5-2682 v4 CPU for the above four species are in the range of ±0.002%, ±0.006%, ±0.002%, and ±0.002%. In terms of

REs between the HIP version on the domestic cluster A and Fortran version on the Intel E5-2682 v4 CPU, the values are much smaller than REs between CUDA and Fortran versions which are fall into the range of $\pm0.00006\%$, $\pm0.0005\%$, $\pm0.00004\%$, and $\pm0.00008\%$, respectively.

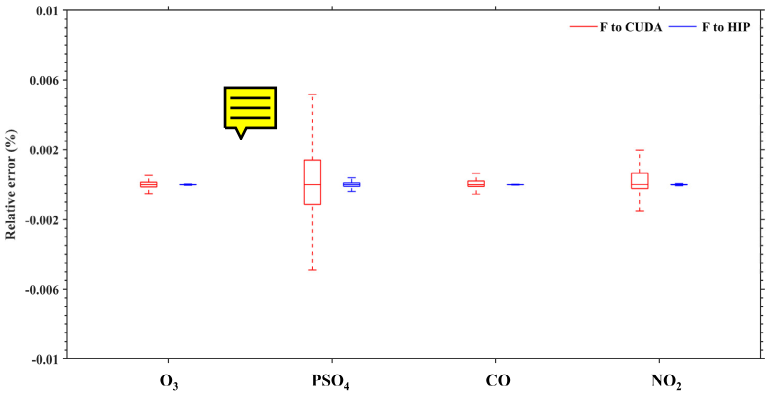

Figure 3. The distribution of REs in all grid boxes for the O_3, PSO_4, CO, and NO_2 under the BJ case. The red boxplot represents the REs between the CUDA version on the NVIDIA K40m cluster and Fortran version on the Intel E5-2682 v4 CPU, and blue boxplot represents the REs between the domestic cluster A and Fortran version on the Intel E5-2682 v4 CPU.

Wang et al. (2021) verified the applicability of the numerical model in scientific research by computing the ratio of root mean square error (RMSE) between two different model versions to system spatial variation (standard deviation, std). If the ratio is smaller, it is indicated that the difference in the simulation results of the model on the GPU is minimal compared with the spatial variation of the system, that is to say, the simulation results of the model on the GPU are accepted for scientific research. Here, we compute the standard deviation of O_3, PSO_4, CO and NO_2 on the Intel Xeon E5-2682 v4 CPU, and their root mean square error (RMSE) between the NVIDIA V100 cluster, NVIDIA K40m cluster and domestic cluster A and the Intel Xeon E5-2682 v4 CPU, which are presented in Table 4. The std for the above four species on the Intel Xeon E5-2682 v4 CPU are 9.6 ppbV, 1.7 $\mu g \cdot m^{-3}$, 141.9 ppbV, and 7.4 ppbV, respectively, and their ratios of RMSE and std on domestic cluster A are $5.8 \times 10^{-5}\%$, $4.8 \times 10^{-6}\%$, $5.7 \times 10^{-8}\%$, and $2.1 \times 10^{-4}\%$, which are smaller than two NVIDIA clusters, especially much smaller than the NVIDIA V100 cluster. For example, the ratio on the NVIDIA K40m cluster for four species are 1.2×10^{-4}%, 6.6×10^{-5}%, 7.0×10^{-5}%, and 4.1×10^{-4}%, and ratio on the NVIDIA V100 cluster are 1.5×10^{-2}%, 2.5×10^{-3}%, 6.4×10^{-3}%, and 1.3×10^{-3}%, respectively.

From AEs, REs, and ratio of RMSE and std between different CAMx versions, it can be identified that the HIP version of the GPU-HADVPPM program runs on domestic cluster A with less difference, and the reason for this difference may be related to the fact that the NVIDIA GPU sacrifices part of the accuracy for improved computing performance. In other words, domestic cluster A are more accuracy for scientific computing in the field of the geoscience numerical models.

Table 4. The standard deviation (std) of O_3, PSO_4, CO and NO_2 on the Intel Xeon E5-2682 v4 CPU, root mean square error (RMSE) and its ratio on the NVIDIA V100 cluster, NVIDIA K40m cluster and domestic cluster A

		NIVIDA V100 cluster		**NIVIDA K40m cluster**		**domestic cluster A**	
	std	RMSE	RMSE/std	RMSE	RMSE/std	RMSE	RMSE/std
O_3 (ppbV)	9.6	1.5×10^{-3}	1.5×10^{-2}	1.1×10^{-5}	1.2×10^{-4}	7.4×10^{-6}	7.7×10^{-5}
PSO_4 ($\mu g \cdot m^{-3}$)	1.7	4.3×10^{-5}	2.5×10^{-3}	1.1×10^{-6}	6.6×10^{-5}	2.5×10^{-7}	1.5×10^{-5}
CO (ppbV)	141.9	9.0×10^{-3}	6.4×10^{-3}	1.0×10^{-4}	7.0×10^{-5}	4.4×10^{-7}	3.1×10^{-7}
NO_2 (ppbV)	7.4	9.3×10^{-5}	1.3×10^{-3}	3.0×10^{-5}	4.1×10^{-4}	2.0×10^{-5}	2.7×10^{-4}

4.3. Application performance

4.3.1. GPU-HADVPPM on a single GPU accelerator

As described in Sect. 4.2, we validate the 48 hours simulation results outputted by the CAMx model which coupling the Fortran version HADVPPM, CUDA and HIP version of GPU-HADVPPM. And then, the coupling computational performance of the Fortran version of HADVPPM on the Intel Xeon E5-2682 v4 CPU and domestic CPU processor A, the CUDA version of GPU-HADVPPM on the NVIDIA Tesla K40m and V100 GPU accelerators, and the HIP version of GPU-HADVPPM on the domestic GPU-like accelerator A were compared under BJ, HN, and ZY case. The simulation time in this section is 1 hour unless otherwise specified.

Table 5 listed the elapsed time and speedup of the different versions of HADVPPM on the CPU processors and GPU accelerators for BJ, HN, and ZY cases. Using CUDA and HIP technology to port HADVPPM from CPU to GPU can significantly improve its computational efficiency. Moreover, the optimization of thread and block co-indexing is used to simultaneously compute the grid point in the horizontal direction (Cao et al., 2023), the larger the computing scale, the more obvious the acceleration. For example, for the BJ case, the elapsed time of HADVPPM on the domestic CPU processor A and Intel Xeon E5-2682 v4 CPU was 57.8 and 37.7 seconds, and it takes the only 29.6, 6.8, and 1.6 seconds when porting to the NVIDIA Tesla K40m GPU, the domestic GPU-like accelerator A, and NVIDIA Tesla V100 GPU, with speedup of 2.0x, 8.5x, and 36.1x. The HN case has a slightly larger grid number, the acceleration of GPU-HADVPPM on NVIDIA Tesla K40m GPU, the domestic GPU-like accelerator A, and NVIDIA Tesla V100 GPU is obvious which were 2.7x, 11.5x, and 41.8x, respectively. The ZY case had the largest number of grids in the three cases and exceeded the memory of a single NVIDIA Tesla K40m GPU accelerator, so it was not possible to test its elapsed time on it. But as far as the domestic GPU-like accelerator A and NVIDIA Tesla V100 GPU are concerned, the ZY case gets 28.9x and 80.2x acceleration on it compared to the domestic CPU processor A.

Table 5. The elapsed time and speedup of the Fortran version of HADVPPM on the Intel Xeon E5-2682 v4 CPU and the domestic CPU processor A, the CUDA version of GPU-HADVPPM on the NVIDIA Tesla K40m GPU, NVIDIA Tesla V100 GPU, and the HIP version of GPU-HADVPPM on the domestic GPU-like accelerator A for BJ, HN, and ZY case. The unit of elapsed time is in seconds (s).

		BJ case		HN case		ZY case	
		Elapsed time (s)	Speedup	Elapsed time (s)	Speedup	Elapsed time (s)	Speedup
CPU processor	domestic CPU processor A	57.8	1.0x	71.1	1.0x	609.2	1.0x
	Intel Xeon E5-2682 v4 CPU	37.7	1.5x	48.1	1.5x	395.7	1.5x
GPU accelerator	NVIDIA Tesla K40m GPU	29.6	2.0x	26.3	2.7x	-	-
	domestic GPU-like	6.8	8.5x	6.2	11.5x	21.1	28.9x

accelerator A						
NVIDIA Tesla V100 GPU	1.6	36.1x	1.7	41.8x	7.6	80.2x

In the above experiments to test the coupling performance of GPU-HADVPPM on NVIDIA GPU and domestic GPU-like accelerator, the data transfer time between CPU and GPU was not considered. However, the communication bandwidth of data transfer between the CPU and GPU is one of the most significant factors that restrict the performance of numerical model on the heterogeneous cluster Mielikainen et al., 2012; Mielikainen et al., 2013; Huang et al., 2013). To exhibit the significant impact of CPU-GPU data transfer efficiency, the coupled computing performance of GPU-HADVPPM with and without data transfer time for the BJ case is tested on the domestic cluster A and B with the same DTK version 23.04 software environment. The elapsed time of GPU-HADVPPM on domestic GPU-like accelerator A with and without taking into account the data transfer time between CPU and GPU are 6.8 and 93.1 seconds, respectively, which means that only 7.3% of the time is spent on GPU computing, and the rest of the time is spent on data transfer. Although, the domestic cluster B upgrade the hardware component and network bandwidth, and the elapsed time of GPU-HADVPPM on it with and without taking into account the data transfer time are 5.7 and 23.9 seconds respectively, the GPU computing time is still only 23.8%. Optimizing the data transfer efficiency between CPU and GPU is one of the most important directions for the porting and adaptation of numerical models to heterogeneous clusters.

4.3.2. CAMx-HIP model on the heterogeneous cluster

Generally, the super-large heterogeneous clusters have thousands of compute nodes which are equipped with one or more GPUs on each node. To make full use of multiple GPUs, a parallel architecture with an MPI and CUDA hybrid paradigm was implemented to improve the overall computational performance of CAMx-CUDA model (Cao et al., 2023). In this studying, the hybrid parallelism with an MPI and HIP paradigm was used to implement the HIP version of GPU-HADVPPM run on multiple domestic GPU-like accelerators.

Fig.4 shows the total elapsed time and speedup of CAMx-HIP model which coupled with the HIP version GPU-HADVPPM on the domestic cluster A under the BJ, HN, and ZY cases. The simulation of above three cases for one hour took 488 seconds, 1135 seconds and 5691 seconds respectively when launching two domestic CPU processors and two GPU-like accelerators. When the number of CPUs and GPUs reaches 24, the speedup of BJ and HN cases reaches the maximum, 8.1x and 11.6x, respectively. In terms of the ZY case, it can achieve up to the 17.2 times speedup when equipped with 32 domestic CPU processors and GPU-like accelerators.

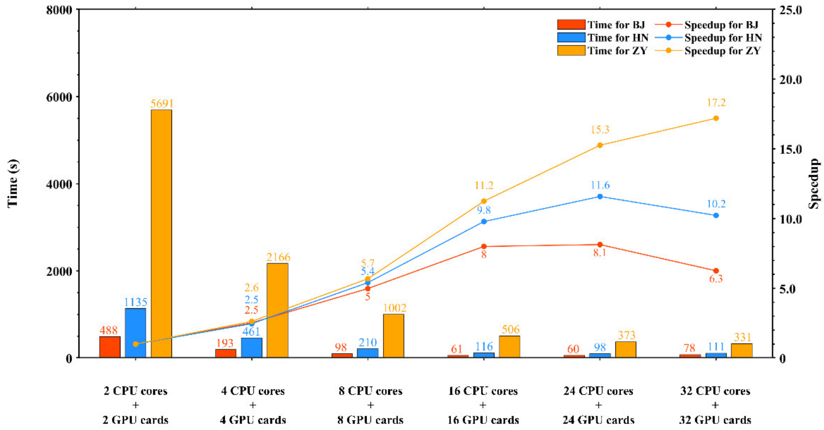

Figure 4. The total elapsed time and speedup of CAMx-HIP model on the domestic cluster A under the BJ, HN, and ZY cases. The unit is in seconds (s).

The number of GPU accelerators in a single compute node is usually much smaller than the number of CPU cores in the super-large heterogeneous cluster. Using the hybrid parallel paradigm with MPI and HIP to configure one GPU accelerator for each CPU process results in idle computing resources for the remaining CPU cores. Therefore, the multi-level hybrid parallelism scheme was introduced to further improve the total computational performance of the CAMx-HIP model. As described in the Sect. 3.2, the horizontal advection module is accelerated by MPI and HIP technology and the other modules which runs on the CPU are accelerated by MPI and OMP under the framework of the multi-level hybrid parallelism.

The ZY case achieved the maximum speed-up when launching the 32 domestic CPU processors and GPU-like accelerators. In the same configuration, Fig. 5 shows the total elapsed time and speedup of CAMx-HIP model when further implementing the multi-level hybrid parallelism on the domestic cluster A. The AEs of the simulation results between the CAMx-HIP model and CAMx-HIP model with the OMP technology is within ±0.04 ppbV, and the specified results are shown in Figure S1. As the number of threads increases, the elapsed time of CAMx-HIP model is further reduced. When a CPU core launching 8 threads, the one-hour integration time in CAMx-HIP model has been reduced from 338 seconds to 178 seconds, with a maximum acceleration of 1.9x.

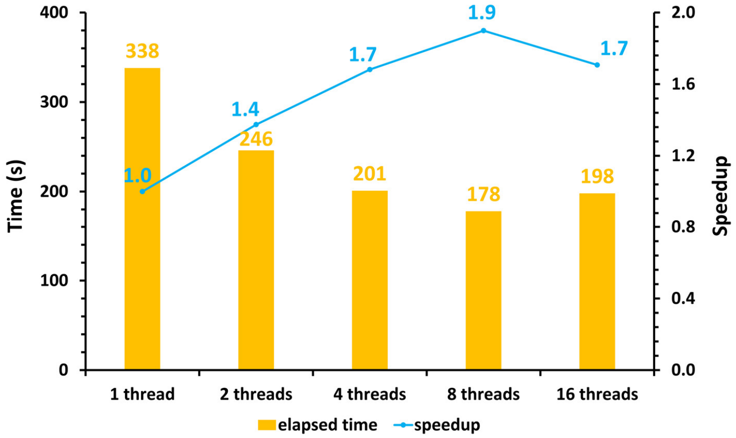

Figure 5. The total elapsed time and speedup of CAMx-HIP model when implementing the multi-level hybrid parallelism in the ZY case. The unit is in seconds (s).

5. Conclusions and discussion

GPUs have become an essential part of providing processing power for high performance computing application, especially in the field of geoscience numerical models, implementing super-large scale parallel computing of numerical models on GPUs has become one of the significant directions of its future development. In this study~~ing~~, the ROCm HIP technology was implemented to port the GPU-HADVPPM from the NVIDIA GPUs to China’ s domestically GPU-like accelerators, and further introduced the multi-level hybrid parallelism scheme to improve the total computational performance of the CAMx-HIP model on the China’ s domestically heterogeneous cluster.

The consistency of model simulation results is a significant prerequisite for heterogeneous porting, although the experimental results show that the simulation difference of CAMx-CUDA and CAMx-HIP models are within an acceptable range, the simulation difference of CAMx-HIP model is smaller, which indicates that the domestic GPU-like accelerator is more accuracy for scientific computing in the field of geoscience numerical models. Moreover, the BJ, HN, and ZY test cases can achieve 8.5x, 11.5x, and 28.9x speedup, respectively, when the GPU-HADVPPM program is ported to the domestic GPU-like accelerator A. ~~And~~ the larger the computing scale, the more obvious the acceleration effect of the GPU-HADVPPM program, which means that the GPU is more suitable for super-large scale parallel computing. The data transfer bandwidth between CPU and GPU is one of the most important factors affecting the computational efficiency of numerical model in heterogeneous clusters, the elapsed time of GPU-HADVPPM program on GPU only accounts for 7.3% and 23.8% when considering the data transfer time between CPU and GPU on the domestic cluster A and B. Therefore, optimizing the data transfer efficiency between CPU and GPU is one of the important directions for the porting and adaptation of geoscience numerical models on heterogeneous clusters in the future.

There is still potential to further improve the computational efficiency of the CAMx-HIP model in the further. First, improve the data transfer efficiency of GPU-HADVPPM between the CPU and the GPU and reduce the data transfer time. Secondly, increase the proportion of HIP C code in CAMx-HIP model on the domestic GPU-like accelerator, and port other modules of CAMx-HIP model to the domestic GPU-like accelerator for computing. Finally, the data type of some variables can be changed from double precision to single precision, and the mixing-precision method is used to further improve the CAMx-HIP computing performance.

Code and data availability. The source codes of CAMx version 6.10 are available at https://camx-wp.azurewebsites.net/download/source/ (ENVIRON, 2023). The datasets related to this paper and the CAMx-HIP codes are available online via ZENODO (https://doi.org/10.5281/zenodo.10158214), and the CAMx-CUDA code is available online via ZENODO (https://doi.org/10.5281/zenodo.7765218, Cao et al., 2023).

Author contributions. KC and QW conducted the simulation and prepared the materials. QW, LiW and LaW planned and organized the project. KC, QW, HG, HW, XT and LL refactored and optimized the codes. LiW, NW, HC, and DL collected and prepared the data for the simulation. KC, HW, QW, and HG validated and discussed the model results. KC, QW, LiW, NW, XT, HG, and LaW took part in the discussion.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The National Key R&D Program of China (grant no. 2020YFA0607804), the National Supercomputing Center in Zhengzhou Innovation Ecosystem Construction Technology Special Program (grant no. 201400210700), GHfund A (grant no. 202302017828), and the Beijing Advanced Innovation Program for Land Surface funded this work. The authors would like to thank the High Performance Scientific Computing Center (HSCC) of Beijing Normal University for providing some high-performance computing environment and technical support.

Reference

Alvanos, M. and Christoudias, T.: GPU-accelerated atmospheric chemical kinetics in the ECHAM/MESSy (EMAC) Earth system model (version 2.52), Geoscientific Model Development, 10, 3679-3693, 10.5194/gmd-10-3679-2017, 2017.

AMD: ROCm Documentation Release 5.7.1, https://rocm.docs.amd.com/_/downloads/en/latest/pdf/ (last access: 20 October 2023), 2023.

Bott, A.: A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advective Fluxes, Monthly Weather Review - MON WEATHER REV, 117, 10.1175/1520-0493(1989)117<1006:APDASO>2.0.CO;2, 1989.

Cao, K., Wu, Q., Wang, L., Wang, N., Cheng, H., Tang, X., Li, D., and Wang, L.: GPU-HADVPPM V1.0: a high-efficiency parallel GPU design of the piecewise parabolic method (PPM) for horizontal advection in an air quality model (CAMx V6.10), Geosci. Model Dev., 16, 4367-

4383, 10.5194/gmd-16-4367-2023, 2023.

Cao, K., Wu, Q., Wang, L., Wang, N., Cheng, H., Tang, X.,Li, D., and Wang, L.: The dataset of the manuscript “GPUHADVPPM V1.0: high-efficient parallel GPU design of the Piecewise Parabolic Method (PPM) for horizontal advection in air quality model (CAMx V6.10)”, Zenodo [data set], https://doi.org/10.5281/zenodo.7765218, 2023.

Colella, P. and Woodward, P. R.: The Piecewise Parabolic Method (PPM) for gas-dynamical simulations, Journal of Computational Physics, 54, 174-201, https://doi.org/10.1016/0021-9991(84)90143-8, 1984.

ENVIRON: User Guide for Comprehensive Air Quality Model with Extensions Version 6.1, https://camx-wp.azurewebsites.net/Files/CAMxUsersGuide_v6.10.pdf (last access: 20 October 2023), 2014.

ENVIRON: CAMx version 6.1, ENVIRON [code], available at: https://camx-wp.azurewebsites.net/download/source/, last access: 20 October 2023.

Huang, M., Huang, B., Mielikainen, J., Huang, H. L. A., Goldberg, M. D., and Mehta, A.: Further Improvement on GPUBased Parallel Implementation of WRF 5-Layer Thermal Diffusion Scheme, in: 2013 International Conference on Parallel and Distributed Systems, Seoul, South Korea, 15–18 December 013, https://doi.org/10.1109/icpads.2013.126, 2013.

Linford, J. C., Michalakes, J., Vachharajani, M., and Sandu, A.: Automatic Generation of Multicore Chemical Kernels, IEEE Transactions on Parallel and Distributed Systems, 22, 119-131, 10.1109/tpds.2010.106, 2011.

Mielikainen, J., Huang, B., Huang, H.-L. A., and Goldberg, M. D.: GPU Implementation of Stony Brook University 5-Class Cloud Microphysics Scheme in the WRF, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5, 625-633, 10.1109/jstars.2011.2175707, 2012.

Mielikainen, J., Huang, B., Wang, J., Allen Huang, H. L., and Goldberg, M. D.: Compute unified device architecture (CUDA)-based parallelization of WRF Kessler cloud microphysics scheme, Computers & Geosciences, 52, 292-299, 10.1016/j.cageo.2012.10.006, 2013.

NVIDIA: CUDA C++ Programming Guide Version 10.2, https://docs.nvidia.com/cuda/archive/10.2/pdf/CUDA_C_Programming_Guide.pdf (last access: 20 October 2023), 2020.

Odman, M. and Ingram, C.: Multiscale Air Quality Simulation Platform (MAQSIP): Source Code Documentation and Validation, 1996.

Sun, J., Fu, J. S., Drake, J. B., Zhu, Q., Haidar, A., Gates, M., Tomov, S., and Dongarra, J.: Computational Benefit of GPU Optimization for the Atmospheric Chemistry Modeling, Journal of Advances in Modeling Earth Systems, 10, 1952-1969, https://doi.org/10.1029/2018MS001276, 2018.

Wang, P., Jiang, J., Lin, P., Ding, M., Wei, J., Zhang, F., Zhao, L., Li, Y., Yu, Z., Zheng, W., Yu, Y., Chi, X., and Liu, H.: The GPU version of LASG/IAP Climate System Ocean Model version 3 (LICOM3) under the heterogeneous-compute interface for portability (HIP) framework and its large-scale application, GEOSCIENTIFIC MODEL DEVELOPMENT, 14, 2781-2799, 10.5194/gmd-14-2781-2021, 2021.