Dear editor,

We are grateful to the reviewers for their constructive comments on our manuscript. The article has been revised accordingly, and point-by-point responses to the comments are provided below.

Reviewer 1:

1. Line 36: It is unclear who or what "Kinaco" refers to. Please clarify.

Response: Thanks. In the revised paper, we have added the following explanations and citations:

Kinaco is a non-hydrostatic ocean model that was developed for high-resolution numerical ocean studies.

Yamagishi, T. and Matsumura, Y.: GPU Acceleration of a Non-hydrostatic Ocean Model with a Multigrid Poisson/Helmholtz solver, Procedia Computer Science, 80, 1658–1669, https://doi.org/https://doi.org/10.1016/j.procs.2016.05.502, International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA, 2016.

Matsumura, Y., Hasumi, H. 2008. A non-hydrostatic ocean model with a scalable multigrid Poisson solver. Ocean Model. 24, 15-28. DOI= http://dx.doi.org/10.1016/j.ocemod.2008.05.001

2. Line 58: LICOM2-GPU, LICOM3-HIP, and LICOM3-CUDA are model versions, not heterogeneous supercomputers; please adjust the wording accordingly.

Response: Thanks. The clarifications of LICOM2-GPU, LICOM3-HIP, and LICOM3-CUDA are modified in the revised paper as follows.

The development of LICOM for heterogeneous supercomputers is evidenced by three key versions: LICOM2-gpu (Jiang et al., 2019), LICOM3-HIP (Wang et al., 2021), and LICOM3-CUDA (Wei et al., 2023), each specifically ported to a different computing architecture.

3. Section 2.2: The paper refers to the Sunway system as a "heterogeneous" architecture, but this is not clearly explained. Please clarify that heterogeneity arises from two distinct core types within each chip, the general-purpose MPEs and lightweight CPEs with separate memory hierarchies and instruction sets, rather than from separate CPU and GPU components. The section would also benefit from citing one or more detailed references on the SW26010 Pro system architecture.

Response: Thanks. In Section 2.2, we have clarified that heterogeneity arises from two distinct core types within each chip, the general-purpose MPEs and lightweight CPEs with separate memory hierarchies and instruction sets, rather than from separate CPU and GPU components. The reference, which contains the details of the SW26010 Pro architecture, is added to the article.

Lin, R., Yuan, X., Xue, W., Yin, W., Yao, J., Shi, J., Sun, Q., Song, C., and Wang, F.: 5 ExaFlop/s HPL-MxP Benchmark with Linear Scalability on the 40-Million-Core Sunway Supercomputer,

in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '23, Association for Computing Machinery, New York, NY, USA, ISBN 9798400701092,410 https://doi.org/10.1145/3581784.3607030, 2023.

4. Section 2.2: Please indicate the overall size of the Sunway supercomputer (e.g., total nodes, processors, or cores) to give readers a clearer sense of the system scale used for the simulations presented here

Response: Thanks. In Section 2.2, an additional clarification regarding the overall size of the new generation Sunway supercomputer has been included in the revised manuscript. The new-generation Sunway supercomputer, equipped with more than 100,000 SW26010-Pro chips, each with 390 cores, is among the fastest supercomputers globally.

5. Line 132: Please clarify what specific programming challenges are referred to, e.g., related to memory hierarchy, data communication between CPEs and MPEs, or algorithm adaptation to the Sunway architecture.

Response: Thanks. The programming challenges are added in the revised paper as follows:
The primary challenge in optimizing for the SW26010 Pro processor stems from its heterogeneous architecture. In this architecture, the Management Processing Element (MPE) handles inter-process communication and controls the overall application workflow. The main computing power, however, resides in the Core Processing Elements (CPEs). Each CPE is equipped with a manually managed Local Data Memory (LDM) that provides access speeds comparable to those of the L1 cache. CPEs can communicate directly via Remote Memory Access (RMA). To leverage the CPEs' computational capacity, code executed initially on the MPE must be ported to an Athread kernel. The MPE is responsible for launching this kernel and subsequently waiting for its completion. Consequently, effectively leveraging the unique characteristics of the CPEs is the key to achieving high performance.

6. Line 139: The term "Athread kernel" refers to the parallel programming model on Sunway, but most readers may not be familiar with it. Please provide a brief explanation of Athread and its role in parallel execution.

Response: Thanks. The parallel programming model for the Sunway architecture has been added in the revised paper as follows:

The Athread programming model is a parallel programming model for the Sunway architecture. It provides an abstraction that is closely mapped to the Sunway hardware. It offers explicit control mechanisms for managing the DMA (Direct Memory Access) controller on the CPEs. This allows programmers to efficiently move data between the main memory (controlled by the MPE) and the Local Data Memory (LDM) of each CPE, which is crucial for overcoming memory bandwidth bottlenecks. The typical execution flow involves the main program running on an MPE. The MPE calls "athread spawn" interface to create "slave" threads that execute a specified function on the CPEs. All threads in the team execute the same function, but on different portions of the data. The Athread programming model provides synchronization primitives (e.g., barriers) to coordinate the execution of these

threads. The MPE calls "athread_join" to wait for the code execution to finish.

7. Figure 1: The text and labels in Fig. 1a are too small to read clearly when printed. Please enlarge the figure or adjust the layout for better legibility.

Response: Thanks. The text and labels in Fig. 1 have been enlarged and updated in the revised paper.

8. Line 156: Suggest to place JK decomposition in quotation marks ("JK decomposition") to indicate it is a specific term introduced by the authors.

Response: Thanks. Quotation marks ("JK decomposition") have been added in the revised paper.

9. Lines 186 and Fig. 5: The discussion of IJ, IK, and WKK decomposition is confusing. Please clarify how these decomposition strategies differ and what "WKK" specifically represents.

Response: Thanks. The clarifications of IJ and IK decomposition have been added in the revised paper as follows:

In the LICOM model, space is discretized into 3-D grid points. Horizontal grid points are labeled as (I, J). Each grid point represents multiple levels, which correspond to various vertical heights in the ocean; the vertical height is denoted as K. Most data structures (arrays) in LICOM are three-dimensional arrays with the layout of (I, J, K). The elements in Fortran's arrays are stored in column-major order. Therefore, elements in dimension I are stored continuously in memory. Because there are different computational patterns in LICOM, different decomposition schemes are used for different patterns. For example, "JK decomposition" means that the computation is decomposed by assigning tasks with different J and K ranges to different CPEs. WKK is a variable name in Figure 5a.

10. Line 214: Please clarify the phrase "across tens of thousands of machines." Do you mean compute nodes?

Response: Thanks. In the revised manuscript, the term "machines" in the phrase "across tens of thousands of machines" has been replaced with "compute nodes".

11. Line 228: The term "Canuto parametrization" appears without prior introduction or reference. Please briefly explain or cite the source when first mentioning it.

Response: Thanks. In the revised paper, we have added an explanation and the corresponding reference for the term "Canuto parametrization". It refers to the schemes for vertical viscosity and diffusivity used in LICOM. The reference is as follows:

*Canuto, V. M., A. Howard, Y. Cheng, and M. S. Dubovikov, 2002: Ocean Turbulence. Part II: Vertical Diffusivities of Momentum, Heat, Salt, Mass, and Passive Scalars. J. Phys. Oceanogr., **32**, 240 – 264, https://doi.org/10.1175/1520-0485(2002)032<0240:OTPIVD>2.0.CO;2.*

12. Line 245: It appears that an equation is missing at this point in the manuscript.

Response: Thanks. We have added the equation back in Section 3.4.2.

13. Tables 1 and 3: The timestep units (presumably seconds) are missing. Please also explain why all configurations use the same timestep despite large differences in horizontal resolution. Typically, finer grids require smaller timesteps for stability.

Response: Thanks. In the revised paper, Tables 1 and 3 have been updated to specify that the timestep unit is seconds. We have added an explanation for why all configurations use the same timestep despite large differences in horizontal resolution, as follows:

To ensure a fair comparison in our scalability tests, we evaluated all data resolutions by employing the time step of the highest-resolution simulation as the universal time step. Therefore, only one factor, resolution, changes across all these experiments.

14. Sections 4.3–4.6: These sections are quite brief. Consider merging them into one cohesive section summarizing the scaling and benchmarking results to improve readability.

Response: Thanks. We have merged Sections 4.3–4.6 into one section in the revised paper.

15. Line 276: The term “super large parallel scale” likely refers to the largest simulations conducted in this study, but please state this explicitly to avoid ambiguity.

Response: Thanks. In Section 4.2, we have revised the term "super large parallel scale" to "large parallel scale" to avoid ambiguity.

16. Figures 11–13: The units of the displayed quantities (e.g., sea surface height, temperature, salinity) are missing. Please add appropriate units to the color bars or captions.

Response: Thanks. In Figures 11–13, the units of the displayed quantities have been added to the captions.

17. Code and Data Availability: The “project website” and the citation “Xu (2025)” both seem to refer to the same Zenodo record (10.5281/zenodo.15494635). Please clarify whether these are distinct (e.g., project page vs. archived version) or consolidate them to avoid redundancy.

Response: Thanks. For the Code and Data Availability section, we have consolidated the “project website” and the citation “Xu (2025)” to avoid redundancy.

18. Technical corrections

A careful proofreading or light English edit is recommended to improve readability and ensure consistent terminology.

Please follow the Copernicus manuscript composition guidelines for capitalization, abbreviations, and formatting when referring to Figures, Tables, and Sections:
https://publications.copernicus.org/for_authors/manuscript_preparation.html

Line 106: Please correct or complete the reference “Y.Q. et al.” to match the proper citation format.

Line 114: The degree symbol (°) is missing, please add.

Line 176: The sentence beginning “Inout the attribute is used···” should be revised for clarity,

e.g., “The inout attribute indicates whether the array is read-only or modified within the kernel.”

Line 221: Please fix the broken equation references (“equation ??”).

Line 265: The sentence beginning “Whenever the…” is unclear or incomplete; please revise.

Line 277: The manuscript frequently uses “mix precision,” but the correct term is “mixed precision.” Please revise throughout.

Line 336: Replace “double-only implementation” with “double-precision implementation” for accuracy.

Response: Thanks. We have corrected the errors as follows:

Line 109: We have corrected the author name in the reference from “Y.Q. et al.” to “Yu et al.”

Line 117: The degree symbol (°C) has been added.

Line 203: The sentence beginning “Inout the attribute is used…” has been replaced by “The inout attribute indicates whether the array is read-only or modified within the kernel.”

Line 250: The broken equation references (“equation ??”) have been fixed.

Line 299: The sentence beginning “Whenever the…” has been replaced by “The time consumption of both module computation and I/O processes can be accurately measured.”

Line 368: We have replaced “double-only implementation” with “double-precision implementation” for accuracy.

Reviewer 2:

1. 141 easy to get started -> easy to implement
   142 are required to be described-> need to be specified
   175 InOut the attribute is used -> InOut is an attribute that is used

Response: Thanks.
Line 162: We have replaced "easy to get started" with "easy to implement".
Line 163: We have replaced "are required to be described" with "need to be specified".
Line 203: The sentence beginning "Inout the attribute is used…" has been replaced by "The inout attribute indicates whether the array is read-only or modified within the kernel."

2. 214 "tens of thousands of machines": do you mean tens of thousands of cores? Please provide an estimate of IO performance improvemen

Response: Thanks. In the revised manuscript, the term "machines" in the phrase "across tens of thousands of machines" has been replaced with "compute nodes". In Section 4.2, we have clarified that our new split writing scheme demonstrates its advantage by achieving speedups of 2.56× and 6.9× over PNetCDF in different scenarios.

3. 221 equation numbers are missing
   245 formula is missing

Response: Thanks. The broken equation references (previously showing as "equation ??") have been fixed. We have added the equation back in Section 3.4.2.

4. L275 same -> identical
   277 Mix -> Mixed

Response: Thanks.
Line 306: We have replaced "same" with "identical".
In the revised manuscript, "mix precision" has been replaced with "mixed precision."

5. Table 1: units missing on timestep (presumably seconds)
   Fig 12: state units

Response: Thanks. In the revised paper, Tables 1 and 3 have been updated to specify that the timestep unit is seconds. The units for the quantities shown in Figures 11–13 have been added to the captions.

All changes in the revised manuscript are clearly marked in the tracked-changes version. which is submitted alongside this response letter. We thank the reviewer again for their valuable feedback, which has greatly enhanced our work

Best Regards,
Kai Xu and Hailong Liu