Response Letter to Reviewer #2

High-Performance Geodynamics on GPUs Using the PETSc's CUDA Backend

[Ms. Ref. No.: egusphere-2026-2528]

[Kyeong-Min Lee[1], Deok-Kyu Jang[3], Cedric Thieulot[4], Byung-Dal So[1,2]*]

Text in ~~red strike-through~~: deleted from the manuscript.
Text in blue: added or revised.
Text in *italic*: comments from Reviewer #2.

The paper describes potential speed gains from combining FEniCS with a CUDA enabled solver backend. Such comparisons are always helpful to judge if it is worth the effort of combining such large frameworks.
A few points from my side:

Authors' reply:

We sincerely thank the reviewer, Christian Hüttig, for the careful evaluation of our manuscript and for the constructive comments. We agree that such comparisons are valuable for assessing whether the effort required to combine FEniCS with a CUDA-enabled solver backend is justified. We carefully considered the reviewer's suggestions and revised the manuscript accordingly. Our detailed responses are provided below.

- We are dealing with memory bandwidth limited processing, please add the theoretical memory bandwidth of the CPU system and the GPUs in use. Scaling should be somewhere in that range.

Authors' reply:

We thank the reviewer for this important comment. We agree that the performance of our calculations is largely limited by memory bandwidth. However, the original manuscript did not provide the theoretical bandwidths of the hardware in this study. In the revised manuscript, we added the theoretical memory bandwidths of both the CPU system and the GPUs to the hardware description. The Intel Xeon Gold 6338 for the CPU timing measurements has a theoretical memory bandwidth of 204.8 GB/s with eight-channel DDR4-3200 memory. The GeForce RTX 6000 Ada provides 960 GB/s, corresponding to a bandwidth ratio of approximately 4.7 relative to the CPU system. With two GPUs, the aggregate theoretical bandwidth is 1,920 GB/s, or approximately 9.4 times that of the CPU system. The measured speedups by GPU range from approximately 5 to 11 depending on the benchmark and are therefore broadly consistent with the bandwidth ratios noted by the reviewer. In some cases, the measured speedup exceeds the theoretical bandwidth ratio of 4.7 for a single GPU. This difference can arise because the CPU and GPU architectures achieve different fractions of their theoretical peak bandwidths. In addition, different BoomerAMG coarsening strategies are used for the CPU and GPU configurations, resulting in different iteration counts. We therefore use the theoretical bandwidth ratio as a first-order reference rather than a strict upper bound. The measured speedups are compared with this reference in the revised manuscript.

Lines 107-108; Section 2,
The third workstation used the same processor but was paired with two GeForce RTX 6000 Ada GPUs, each with 48 GB of memory (96 GB in total). The Xeon Gold 6338 provides eight DDR4-3200 channels, giving a theoretical memory bandwidth of 204.8 GB s^{-1}, and each GeForce RTX 6000 Ada provides 960 GB s^{-1}.

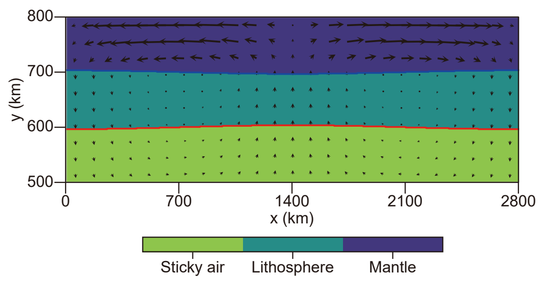
Lines 293-296; Section 4.1,

For the highest resolution (1024 × 1024, 9,447,427 DOFs), the GPU required 2.25 s, while the 32-core CPU required 17.82 s, corresponding to a speedup of approximately 7.9×. The two systems differ in theoretical memory bandwidth by 4.7×. Speedups measured across the benchmarks span 5–11×, ranging from approximately the bandwidth ratio to higher values. PCG iterations and GMRES stages are dominated by sparse matrix-vector products. Such operations exhibit high data parallelism, which leads to particularly efficient execution on a GPU equipped with large memory bandwidth. The cost of PCG and GMRES is dominated by sparse matrix-vector products, data-parallel operations that map onto the GPU, where high memory bandwidth sustains the streaming access pattern. The ratio of peak memory bandwidth between the GPU and CPU remains a poor predictor of the achievable speedup. The GPU and CPU deliver unequal fractions of peak bandwidth. The effective bandwidth ratio during the computation exceeds the nominal peak ratio. The CPU and GPU versions of BoomerAMG further adopt different coarsening strategies that shift the number of iterations. The measured speedup depends on the cost per iteration and on the number of iterations. The peak-bandwidth ratio serves as an approximate reference rather than an upper bound.

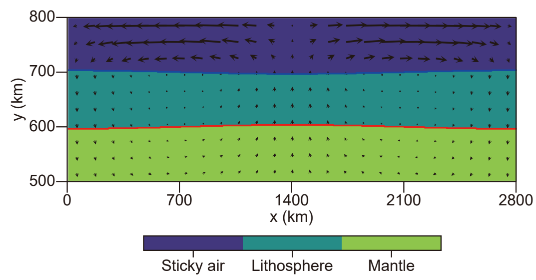
- The color descriptions in the sticky air figure are incorrect

Authors' reply:

We thank the reviewer for pointing this out. In the original manuscript, the labels for sticky air and mantle were reversed in the composition color bar of Figure 10. In the revised manuscript, we corrected the order of the labels to match the figure. Please see the revised box below.



Original Figure 10



Revised Figure 8

- Eq 4 only advects T not C_i

Authors' reply:

Equation 4 describes only the advection and diffusion of the temperature field. The equation governing the evolution of the composition field was not explicitly provided in the original manuscript. Although the level-set method was described in the text, the advection equation for the level-set function was not presented as a separate equation.

In the revised manuscript, we added the level-set advection equation and the reinitialization equation immediately after Equation 4. We also describe how the composition field is determined from the sign of the level-set function. The composition field is then used to define density and viscosity. Please see the revised text below.

Lines 127-131; Section 2.1,

The thermal evolution of the system is governed by the advection-diffusion equation (Equation 4).

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \nabla \cdot (\nabla T) \tag{4}$$

The temporal evolution of the multi-composition field depends only on advection through the velocity field.

We adopted a level-set method to reduce oscillations and numerical diffusion of the compositional field under advection-dominated problems (Hillebrand et al., 2014; Wú et al., 2023). The advection of the level-set function ϕ governs composition field's evolution (Equation 5). The level-set function is reinitialized to a signed-distance function satisfying ($|\nabla\phi| = 1$) by solving the reinitialization equation (Equation 6) using second-order ENO spatial discretization, a Godunov Hamiltonian, and RK2 pseudo-time integration.

$$\frac{\partial \phi}{\partial t} + \mathbf{u} \cdot \nabla\phi = 0 \tag{5}$$

$$\frac{\partial \phi}{\partial \tau} + S(\phi_0)\left(|\nabla\phi| - 1\right) = 0, \text{ where } S(\phi_0) = \frac{\phi_0}{\sqrt{\phi_0^2 + \epsilon^2}} \tag{6}$$

S, ϕ_0, and ϵ denote the smoothed sign function, the level-set function before reinitialization, and the smoothing parameter, respectively, with $\epsilon = 1.5 \min\left(\frac{1}{N_x}, \frac{1}{N_y}\right)$

- For the sticky air T has to be 0 and the scenario cannot be mixed with thermal convection. Eq 4 and eq 1 cannot be used in a general sense. Please state that.

Authors' reply:

We agree with the reviewer's comment. In the original manuscript, Equations 1 and 4 were presented in a general form, but we did not clearly specify which buoyancy term was activated in each benchmark. In the sticky-air benchmark, no temperature field was included, and buoyancy was controlled only by the compositional term. Therefore, (Ra = 0) and Equation 4 was not solved in this case.

Thus, the sticky-air and thermal-convection settings are not used simultaneously in the same simulation. In the revised manuscript, we now clarify that the buoyancy terms are activated selectively depending on the benchmark. We also state that Equation 4 is solved only when a temperature field is included. We also explicitly state in the sticky-air benchmark section that (Ra = 0) and that no temperature field is used. Please see the revised text below.

Lines 117-119; Section 2.1,

We incorporated the thermal (Ra) and compositional (Rb) Rayleigh numbers into the buoyancy term. Equation 1 assumes a general form, with each benchmark activating only the terms required by the physical setting. Thermal-buoyancy cases set Rb = 0, while compositional-buoyancy cases set Ra = 0 in the absence of a temperature field.

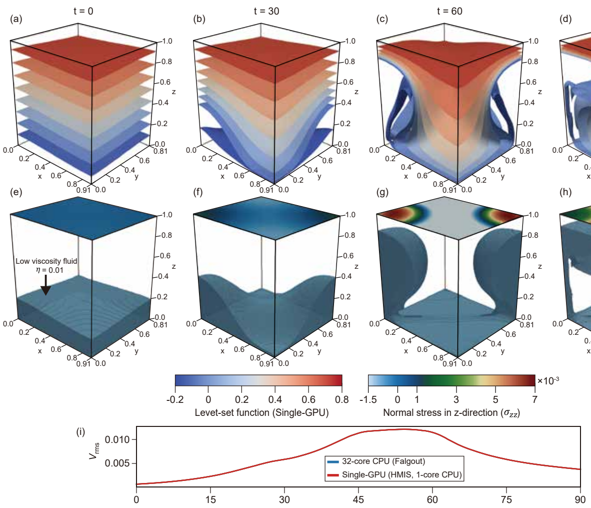
Lines 421-424; Section 4.6,

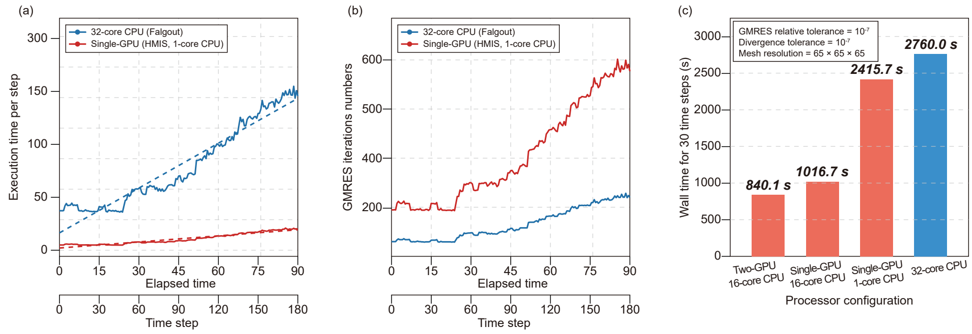
In our study, we set the viscosities of the lithospheric, mantle, and sticky-air layers (with a thickness of 100 km) to 10^{23}, 10^{22}, and 10^{18} Pa s, respectively. Buoyancy in the sticky-air benchmark is purely compositional. The isothermal configuration of Crameri et al. (2012) fixed Ra = 0 without a temperature field and thermal convection. The nondimensional parameters for the sticky-air benchmark are presented in Table S1.

- Inventing little own benchmarks (3D Taylor) is nice but since you did not compare it to anything else their only point is to show the speedups again, but they all solve the same eqns more or less, so what do we gain from that? If you want to establish it as a benchmark, you need to show the solutions of other available codes.

Authors' reply:

We agree with the reviewer's comment. The 3-D Rayleigh–Taylor case was developed specifically for this study, and no published reference solution is available for direct comparison. We agree that it is not appropriate to present this case as a benchmark without comparison against solutions from other codes. In the original manuscript, this case primarily demonstrated GPU performance for a three-dimensional problem rather than serving as a validation test. The difficulty is therefore one of placement: a case that carries no reference solution stood in the benchmark section, where verification is what the reader expects. Since the revised manuscript already includes 3D thermo-mechanical convection that demonstrates computational performance and memory usage, the 3-D Rayleigh–Taylor case is not essential for this purpose. Therefore, the 3-D Rayleigh–Taylor case was moved from the benchmark section to the Supplementary Material.





The original Figures 8 and 9 have been moved to the Supplementary Material

We sincerely thank the reviewer for the careful and constructive evaluation of our manuscript. Two of the comments were especially valuable. Asking for the theoretical memory bandwidth of the hardware gave us the right frame in which to read our speedups, and pointing out the missing equation for the composition field, the buoyancy terms active in each benchmark, and the reversed labels in the sticky-air colour bar led us to describe what we actually solve. The revision is clearer on both counts, and we are grateful for the reviewer's attention to detail.

With best regards,
Kyeong-Min Lee and Byung-Dal So

References of this response letter

F. Crameri, H. Schmeling, G. J. Golabek, T. Duretz, R. Orendt, S. J. H. Buiter, D. A. May, B. J. P. Kaus, T. V. Gerya, and P. J. Tackley. A comparison of numerical surface topography calculations in geodynamic modelling: an evaluation of the 'sticky air' method. *Geophysical Journal International*, 189(1):38–54, 2012. doi: 10.1111/j.1365-246X.2012.05388.x.

B. Hillebrand, C. Thieulot, T. Geenen, A. P. van den Berg, and W. Spakman. Using the level set method in geodynamical modeling of multi-material flows and Earth's free surface. *Solid Earth*, 5:1087–1098, 2014. doi: 10.5194/se-5-1087-2014.

Qihang Wú, Shoufa Lin, and André Unger. Modeling multi-material structural patterns in tectonic flow with a discontinuous Galerkin level set method. *Journal of Geophysical Research: Solid Earth*, 128: e2023JB026806, 2023. doi: 10.1029/2023JB026806.