Rebuttal to anonymous reviewer

The authors thank the reviewer for his thorough review. We will first address the reviewer's high-level comments, and thereafter the detailed comments point-by-point.

High-level suggested revisions:

1. *In sections 1-2, switching between anelastic and Boussinesq should be made clearer, and with what approximations. In the rest of the paper, it should be clear what "mode" each test is run in. - Claims of conservation should state the caveat that the simplified equations are in flux-conservation form, but that they are not fully mass- or energy-conservative (for example, looking at total mass, \rho_0 + \rho', in equation 10, is not conservative). p7 l18 how is it "fully energy conserving"? - A little more discussion of why this discretization was chosen, and what its benefits/limitations are. A little extra information would be a good way to flesh out the conclusion and provide more context for the reader.*
   We will improve the revised manuscript with respect to the differences between Boussinesq and anelastic in the model implementation. In short, the implementation of the governing equations is the same under both approximations, but under Boussinesq, the reference density and potential temperature are constant with height in the momentum and mass-conservation equations.

   Based on the reviewer's comments, we have not made our claims of energy conservation sufficiently clear. In Bannon (1996)'s anelastic approximation, the governing equations are energy conserving, in the sense that there is a correct transfer between kinetic and potential energy. This, however, does not mean that the discrete implementation is energy conserving. Our spatial discretization that follows Morinishi et al. (1998), conserves mass, momentum, and kinetic energy, which we demonstrate in the paper. We will make the distinction between energy conservation in the governing equations and in the implementation clear throughout the improved manuscript.

Detailed minor revisions:

2. *p1 abstract: "code reaches speedups of more than ... conventional code" running on what processors? Generally best to express it as a % of peak FLOPS and specific to the two architectures you compare in results.*
   We will explicitly mention in the abstract that it concerns single-GPU simulations and move the detailed information to the section on the scaling.

3. *p1 "approach the synoptic scales" remove the? to clarify, maybe add LES resolution (< 1km?) at "scales of 1000km or more"?*
   We will follow the reviewer's suggestion and add some explicit numbers to the statement.

4. *p2 l3, "order codes"? Older codes?*
   "Order codes" will be changed to "other codes".

5. *p2 last intro paragraph . . . it is worth mentioning Sec 5 (output), and 7 (instructions to reproduce), to encourage others to do the same, maybe mention w/ sec 13 or even move those sections to the end?*
   We agree with the reviewer that all sections need to be mentioned. We are not sure what the reviewer means by "encouraging others to do the same" Does this refer to reproducing our test cases?

6. *p2 l18 "constant with height z" maybe restate \rho_0(z) only to support eq. (2)?*
   We will write rho_0(z) instead of rho_0, to make clear that rho_0 is a function of height.

7. *p3 Derivation of eq(4) should be either referenced or add an extra step ... eq (5) should come first, for example, to introduce the potential temperature EOS that's sub- stituted into eq(4). - p3 l20 perturbational pressure form - not conservative / does not match eq (2)?*
   The reviewer is correct. We shall swap the order of the equation of state and the momentum equation and cite the paper of Bannon (1996) earlier.

8. *p4 l16, introduced N without an equation/definition?*
   We shall introduce the definition of N2 = db_0/dz in the text, and do so for all thermodynamic modes.

9. *p6-7 I appreciate the compactness of the notation and clarity in presenting it.*
   We appreciate the kind words of the reviewer. It has been a challenge to find a suitable notation.

10. *p7 eq (40), why not use a similarly compact 4th-order 5-pt wide stencil, instead of the larger 7-pt wide one?*
    The 7-pt stencil has the advantage that it is built out of the same building blocks as the other operators, and thus uses the same ghost cells.

11. *p9 DFT solver eq (45) is not clear ... assuming periodic bc's or cosine transform for Neumann bc's on pressure? Is there a reference for this approach?*
    The DFT operator is only performed in the periodic x and y directions. Based on comments of the first reviewer, we will introduce earlier in the paper that our code is periodic in the two horizontal dimensions.

12. *p9 "hat" DFT notation conflicts with "average" notation on p6.*
    We will introduce a different symbol for the Fourier transform in the revised manuscript.

13. *p9 eq(46-47) could mention "corresponding to eq(39-40) respectively" around l17-18?*
    We agree with the reviewer's suggestion and will refer to those equations.

14. *p9 l24, Ah! That's a big assumption, periodic lateral boundaries. Should be moved up and stated prominantly, along with motivation/limitations. Now I understand why p4*

l15 “periodic with slopes” was introduced.
Following both reviewers’ comments, we will introduce in the introduction that our code is doubly periodic.

15. *p11 l12-16, is the model-top pressure constant in time or modified every time step? what value is used?*
    The model top pressure is the final result of the described procedure and depends on the surface pressure and the chosen reference profiles of temperature and humidity. MicroHH has the option of a constant profile in time, as well as a reference profile that updates in time.

16. *p12 l5, is filtering actually applied in your algorithm, and if so, at what resolution? Do you do anything to prevent discrete aliasing of unresolved wavelengths?*
    We do not use explicit filtering, but rely on the grid scale as a filter, which is a common procedure with atmospheric LES. With our numerical schemes, aliasing errors are small. We will introduce a short discussion on this in the revised manuscript.

17. *p12 tilde variables conflict with tilde “intermediate velocity” in eq (41)*
    We will use a different symbol in the revised manuscript.

18. *p12 eq (67) S_{ij} subscript? and what’s the definition of S^2?*
    In the revised manuscript we will write the full expression in terms of S_{ij}. The reviewer is correct that we forgot the subscripts.

19. *p12 l21, N^2 definition here different than above p4 l16?*
    The reviewer is correct. We have failed to make clear that depending on the chosen thermodynamics, an appropriate definition of N^2 is used. We will clarify this in the improved manuscript, as mentioned in our reply to point 8.

20. *p13-15 sec 4.2 . . . is this a new atm turbulence model? The reference Wyngaard (2010) is an entire book, and it is not clear which tests warrant which boundary conditions, etc. p15 l5 is particularly confusing . . . might be worth describing Obukhov length and its use as a stability/mixing parameter, and why a look-up table is needed.*
    We will clarify the text. The lookup table is only there for performance reasons, as it outperforms a Newton-Raphson method.

21. *p15 l11, why would you not just include a background U_f and define a perturbational velocity from that? That would be compatible with periodic bc’s, guarantee mass conservation, etc.*
    By doing so, the problem will remain. If the large-scale pressure force is applied to the perturbation velocities only, it is no longer ensured that the perturbations average to zero, without applying the presented correction.

22. *p15 bottom "adveciton" should be "advection"?*
    The reviewer is correct.

23. *p17 l19. "precompiler statements"? Meaning #define of GPU CUDA code? Any thoughts or statements on maintaining the different code bases in your C++ framework?*
    The use of precompiler statements is unavoidable, as we do not want to force the non-GPU user to install CUDA and compile the GPU code as well. We have chosen for an implementation in which the GPU code based is minimized, in order to ensure maintainability. We will elaborate our description of the CUDA implementation.

24. *p18 top, MPI-IO should have a reference?*
    We will introduce a reference.

25. *p18 l9, change netCDF footnote to reference?*
    We will introduce a reference.

26. *p18, maybe sections 5-7 should be moved/merged with 13 or all in an appendix?*
    We disagree with the reviewer here. We consider the presented topics in sections 5-7 of high relevance for a model description paper. Section 13 is located at the end of the paper following the GMD guidelines.

27. *p18 l25, love the post-processing mode based on restart files!*
    We thank the reviewer for this compliment. We would be very happy if this convinces the reviewer to use our code.

28. *p19 eq (98) should "4 \pi y" be z?*
    The reviewer is correct!

29. *p19 figure 1 / p20 l1 discussion ... L1 error in 2D should asymptote to hˆ4, even with 3rd-order boundary errors (O(N) pts * O(hˆ3) boundary error vs. O(Nˆ2) pts O(hˆ4) interior error). Please explain? Also adding a 2nd set of dotted lines for 3rd- and 4th-order on the bottom set if u,v 4M fields will better show the break.*
    In the 4^{th}-order scheme, the boundary condition for vertical velocity w is set for global mass conservation rather than for 4^{th}-order accuracy. We will add a second set of dotted lines to help the reader observe the convergence of the schemes.

30. *p20 l8, "diffusion off" you mean viscosity, no source terms, etc. so that total energy should be conserved? What's your equation for "energy" in this test?*
    In this case, energy is kinetic energy. The model is run without viscosity and source terms, but with pressure solver to satisfy the continuity equation. We will clarify this in the improved manuscript.

31. *p20 l10, "its energy conservation." you mean improved? It doesn't conserve energy exactly.*
    The spatial discretization does conserve energy, but it is the time discretization that

does not. We shall clarify this in the text.

32. *p20 figure 2, maybe put top figure on log |\DeltaE| scale as well to distinguish the results better?*
    We prefer to keep our axis in its current form to show that our schemes are losing energy and therefore cannot lead to a blowup of the numerical solution. This is not possible if we plot the absolute value on a log scale. We shall clarify this in the improved manuscript.

33. *p21, line 4. Isn't there a difference in maximum CFL for each as well?*
    There is. In this experiment, however, we chose to compare the accuracy that can be achieved a fixed time step, as this allows us to estimate the convergence.

34. *p22, l6, "perfect match" . . . so perfect it's hard to see any difference at all. What do you attribute that too, since you have completely different discretizations, etc. How were the Moser 1999 results so similar? Could you quantify the differences, plot them, and explain them?*
    Both codes have fully converged results and are therefore identical if sufficient samples are averaged. Direct numerical simulation has, unlike LES, an exact solution, which makes the solution independent of the numerical schemes at sufficient resolution.

35. *p24, l12 . . . ditto for "nearly perfect match" here. Fig 6 also shows a "kink" in E_pp at higher \kappa. Is it worth explaining?*
    MOSER has spectral schemes, which introduce aliasing errors in the highest wave numbers. Even though aliasing errors are removed when the nonlinear operators are applied, the solver for the pressure introduces new ones.

36. *p26 l8, Fig 9a,d - why are the vertical velocities diverging with resolution?*
    This is often observed in LES-simulations of cumulus-topped boundary layers. Individual plumes that have a radius of only a few grid cells tend to overestimate velocity.

37. *p29 bottom p30. By putting these on a single GPU, you are avoiding communication overheads for the GPU. Did you run 1 MPI rank on the GPU? Did you run "n" MPI ranks on the CPU? For the B512 run you are getting very good (90%?) strong scaling for 1-4 CPU nodes.*
    In our view, GPUs mostly deliver a benefit if simulations can be run on a single GPU. Therefore, we have taken one GPU. Furthermore, at the moment, MicroHH is only supporting single GPU simulations. We agree with the reviewer that our comparison might be unfair in the sense that the GPU simulation does not need communication, whereas the CPU simulation does. We have, however, decided to focus on simulation of sizes that are common in atmospheric LES studies. We will improve the discussion in the paper to make this clear.

38. *p30 l4, "a parameterizations . . . has been" singular?*
    The reviewer is correct!

39. *p30 section 12 . . . could add a more comprehensive summary, call out any limitations or tradeoffs.*
    We will elaborate the conclusions and highlight MicroHH's most important features and limitations in the revised manuscript.