Reply to the comments of Referee #2 (RC2) on article gmd-2023-26

Toshiki Matsushima, Seiya Nishizawa, Shin-ichiro Shima

Thank you for taking the time to read through our manuscript and provide us with positive and helpful comments. Below are our point-by-point responses to your comments.

general comments

Most of the model improvements are of ”technical” nature. I am not computer scientist, hence I cannot completely judge these achievements.

Still, the performance achievements seem great. However, I would have liked it even better, if the comparison with traditional bulk method would be made fairer. E.g. in the SDM and bin model there is no ”cold” microphysics (yet), but the bulk method contains these tracers. Hence, it is likely that the advection cost could be reduced to 4/10. Or in the time to solution comparison there is no discussion that with the bulk methods larger microphysical timesteps are possible and might be even sufficient to reach convergence. Considering that the SDM-new128 might be more costly than the bulk method. However, I still would want to apply it, as it makes, as stated in the paper, far less assumptions.

Thank you for the compliment on the performance improvement we have made. We agree with you that the comparison between the mixed-phase microphysics bulk model and SDM is not fair, but it is appropriate in the specific case discussed in this study. We did not use warm microphysics because the original codes need to be modified. Thus, we decided to exclude cold microphysics from Seiki and Nakajima (2014) scheme, performed additional experiments, and compared the results with those of SDM. Nonetheless, we would like to retain the results for mixed-phase microphysics for discussion.

Before showing how it changed the results herein, we share our thoughts on comparing SDM and Seiki and Nakajima (2014) scheme. There are two aspects of convergence for the two-moment bulk method; spatial and temporal resolution. The two-moment bulk method imposes empirical assumptions on DSD, which may lead to less spatial variability or no dependency on spatial resolution, such as the spectral width of DSD. However, this does not mean it converges fast to increase spatial resolution; rather, it indicates that, in principle, fair comparison in terms of spatial resolution is difficult. For the temporal resolution, the time step for the two-moment bulk method could be fairly large (e.g., 5 min) in climate simulations. Based on this, one may assume that the two-moment bulk method can solve groups of droplets using large time steps by bundling the fast time evolution of individual droplets. However, as discussed in Santos et al. (2020), this is not the case. The authors performed eigenvalue analysis for a two-moment bulk scheme and found that a fast mode (<1 s) also exists in the bulk scheme. It does not considerably deviate from the time step determined by the evolution of individual droplet dynamics, such as condensation/evaporation. Thus, it is reasonable for convergence to use a sufficient time step that is determined by eigenvalues, instead of a stable time step. Alternatively, large time steps can be considered if it is stable and the mean climate does not change for earth and planetary climate simulations. For such a case, we agree that a bulk method is suitable. Based on the above discussion, we would like to maintain the same spatial and temporal resolution to compare SDM and the two-moment bulk method.

As explained later, we modified the initial condition of the mass mixing ratio of water vapor (QV) for (only) warm bubble experiments using SDM, the two-moment bulk method, and the bin method. In addition, we have modified the evaporation calculations for the two-moment bulk method. We have also added the results using the bin method with a stochastic collision–coalescence algorithm developed by Sato et al. (2009). Figures 1 and 2 in this letter correspond to Figures 3 and 4 in the original manuscript. The differences are small for warm bubble cases. The elapsed time using the warm microphysics two-moment bulk method is still comparable to that of SDM-128. The description in the revised manuscript

is modified as follows (shown in blue):

L675: The maximum throughputs of SDM-new are 61.3 and 20.1 times that of SDM-orig and two-moment bulk method, respectively.

L684: The elapsed time obtained using the two-moment bulk method was 14.0 s, and that obtained using SDM-new with 128 SDs per cell on average is 13.9 s. The maximum throughput of SDM-new is 31.6 times that of the two-moment bulk method.

Figure 3 in this letter shows the results of BOMEX and SCMS case experiments, corresponding to Figure 7 in the original manuscript. In this case, the elapsed time and energy consumption obtained using warm microphysics two-moment bulk scheme are comparable to that of the results between SDM-new32 and SDM-new64. However, as discussed in the revised manuscript, as the elapsed time for SDM increased more gradually than linearly with the number of SDs when 8-128 SDs/cell were used, the main conclusion did not change considerably. Based on the above discussion, we have changed the conclusion that the elapsed time for SDM-new32 to SDM-new128 is comparable to that of the (warm microphysics) two-moment bulk method.

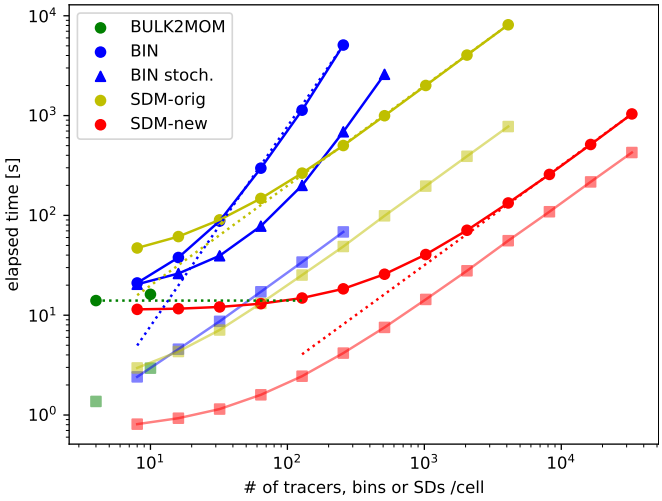

Figure 1. Elapsed times of the total (circles) and tracer advection and SD tracking (squares) using the two-moment bulk method (green), bin method (blue), SDM-orig (yellow), and SDM-new (red) with different numbers of tracers or mean SDs per cell. Elapsed times (triangles) of the total using the bin method with stochastic collision–coalescence algorithms are also shown. Here, SD tracking includes SD movement and sorting with a block as a key. The blue dotted line is proportional to N^2, the red- and yellow-dotted lines are proportional to N, and the green-dotted line indicates a constant determined by N.

It is quite a long paper. Consider whether it can be shortened. Personally, I tend to read and cite paper more often, if they are briefly and succinctly. To provide examples, I would not miss the following parts in main body of the paper:

We thank you for kindly reading through our long paper. Below are our responses to the specific examples you provided for us. For other parts, we would like to simplify them as much as possible based on the comments from two referees.

- discussion about the flow solver, i.e. focus on the microphysics (e.g. microphysics - radiation coupling is also not included only mentioned in L914. It is an interesting question how to couple SDM with radiation but clearly out of the scope of this paper). Of course there are a few details that would need to kept, like that a C grid is used and the effective resolution.

We agree with your comment. In the revised manuscript, we have simplified the description of anelastic equations in L111, flux-corrected transport in L122, model framework in L234, and other components in L914.

- the very details about Fugaku, e.g. L751ff (At that point in the paper I had forgotten what the eco-boost mode does and had to reread the paper to find it on L201. The portion of readers that have access to Fugaku and hence for which these information are relevant, are probably low).

We agree that we provided extensive detail of Fugaku for application scientists. We have simplified Section 2.3 and deleted L431 in the revised manuscript. Alternatively, we would like to retain the discussion on power performance and its depen-

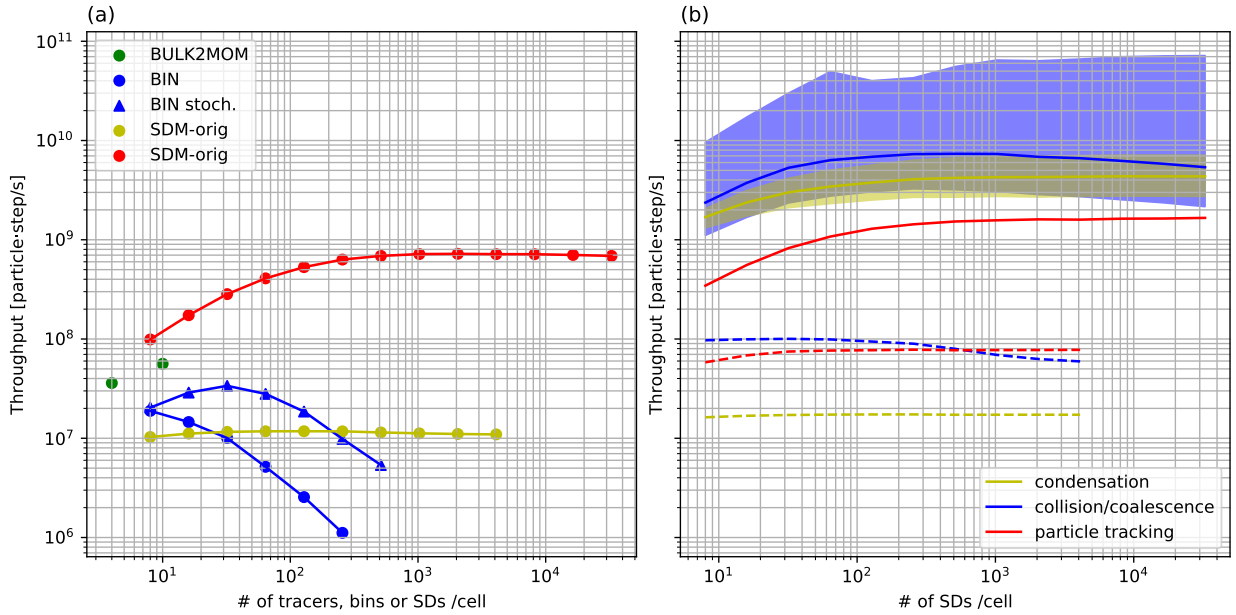

Figure 2. (a) Data throughput of microphysics for the two-moment bulk method (green), bin method (blue circles), bin method with stochastic collision–coalescence algorithms (blue triangles), SDM-orig (yellow), and SDM-new (red) with different numbers of tracers or mean SDs per cell. (b) The mean data throughput of SD tracking (SD movement and sorting with a block as a key), condensation process, and collision–coalescence using SDM-orig and SDM-new with different mean SDs per cell. The dotted and solid lines show the mean data throughputs for SDM-orig and SDM-new, respectively. The range between the minimum and maximum throughputs of condensation and collision–coalescence for SDM-new is indicated by the colors because load imbalance is crucial only for SDM-new.

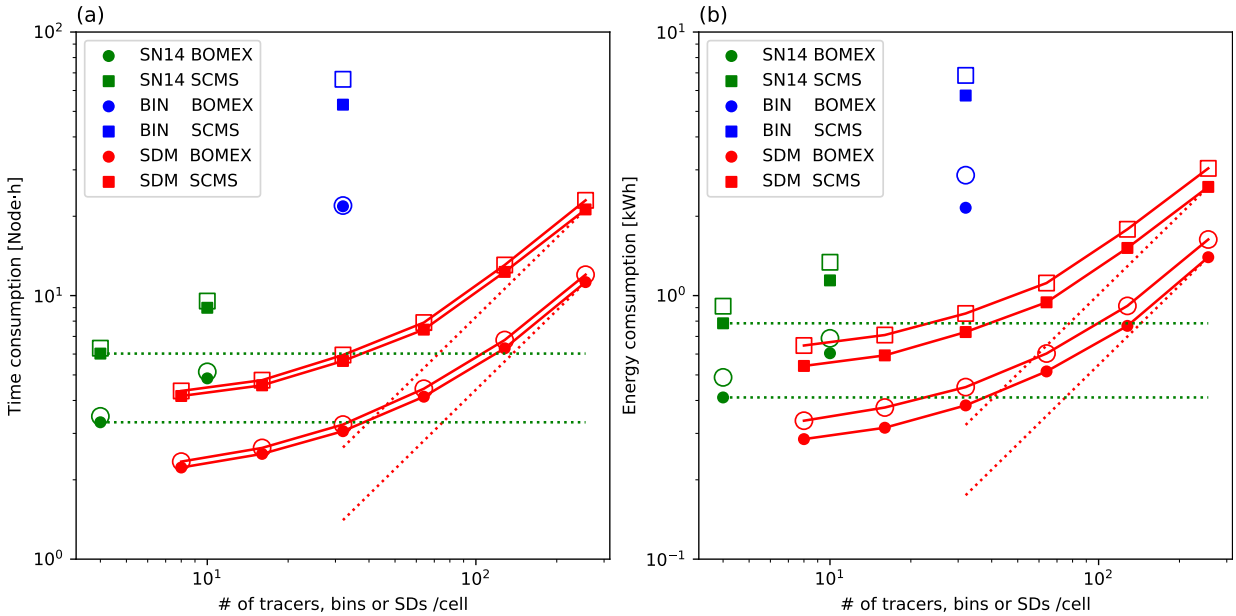

Figure 3. Computational resources for BOMEX and SCMS experiments for various cloud microphysics schemes and different numbers of tracers or mean SDs per cell: (a) node-hours using normal and boost modes; (b) energy consumption using boost and boost eco modes. In (a), the results for the boost and normal modes are shown by the filled and open markers, respectively. In (b), the results for boost eco and boost modes are shown by filled and open markers, respectively. The red-dotted lines show the lines proportional to N.

dence on modes while we provide a guide for those not interested in it because we believe it will become more important in the future.

I read section 3 as if the improvements of the SDM are of technical nature and should not deteriorate the physics. Then I was surprised that in only physical comparison between the SDM-orig and SDM-new in Fig. 6 the difference seem quite substantial around t=1200. You argue with different versions of the flow solver. What change prevents you from using the same version? Then also the discussion in L565ff could be removed. To strengthen the confidence in the SDM-new it would be worth to investigate the difference, their cause and relevance a bit more.

We investigated the difference in detail and found that the main difference is the contribution of the slightly different initial value for QV. The initial value of QV used in SDM-new is larger than expected from the given relative humidity compared with that of SDM-orig, resulting in stronger convection. After using the same initial condition for QV, the difference in the results decreased (Figure 4). The BOMEX and SCMS experiments are not affected by the modification.

Next, to investigate the contribution of the model version and improvement, we performed warm bubble experiments by taking as many SDs as possible (SD32768). Figure 5 shows the LWC for different model versions and settings. The upper right panel in Figure 5 shows the LWC obtained with SDM-new, using almost the same setup as for SDM-orig, except for some minor modifications (numerical representation precision, random number seeds, order of calculations, etc.). Then, we considered model improvements (lower left panel in Figure 5), including the exclusion of the monotone FCT, the use of time-averaged variables during each long time step for the calculations of microphysics, the use of second-order CVI, the use SDs sampling from PDF, which is proportional to total density, and the use of Sobol sequence for initialization. Lastly, we simplified the model setup by reducing collision–coalescence calculations in noncloudy cells and ignoring pressure dependence on coefficient A for condensation calculations (the lower right panel in Figure 5). The difference in the upper left and right panels is very small, indicating consistency between SDM-orig and SDM-new. The model improvement in this study resulted in slightly different results for precipitation timing and behavior of clouds in the upper layer. The simplification of the calculations resulted in a small difference in LWC.

Based on the above discussion, we conclude that the model version is unimportant and the difference in the results between SDM-orig and SDM-new becomes smaller for many SDs, indicating that sampling and randomness caused the differences. In the revised manuscript, the above discussion is provided in L705. Note that we have not provided Figure 5 in the revised manuscript but have provided it in the supplemental Data sets. According to the correction to the initial value of QV, we also performed an additional experiment and moved the corresponding figures (Figure 6, markers for particles are slightly larger) to Figure 5 in the revised manuscript.

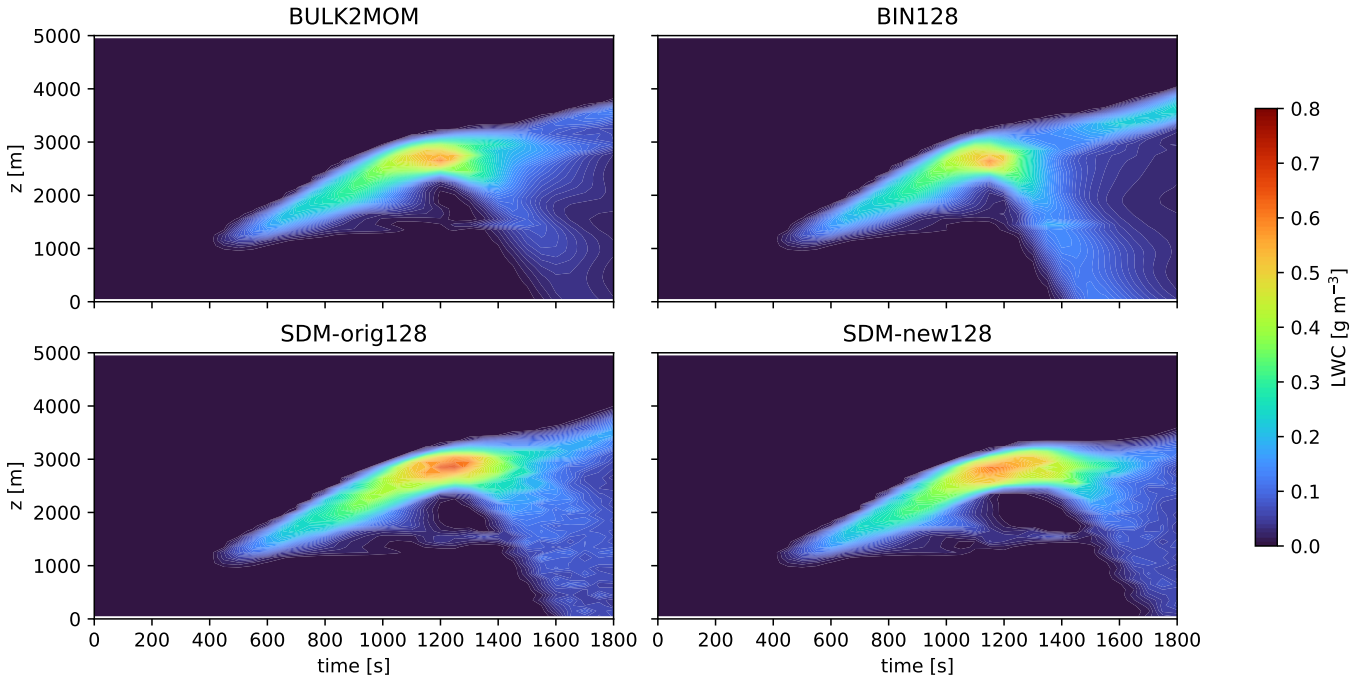

Figure 4. Horizontally averaged time–height cross-section of the liquid water content (LWC) for different cloud microphysics schemes.

specific comments

2.1 Governing equations:

L109-119: The discussion about anelastic vs fully-compressible equation is a bit confusing and maybe misleading. I understand that global communications can be the performance bottleneck. However, as the authors acknowledge, other groups have success-

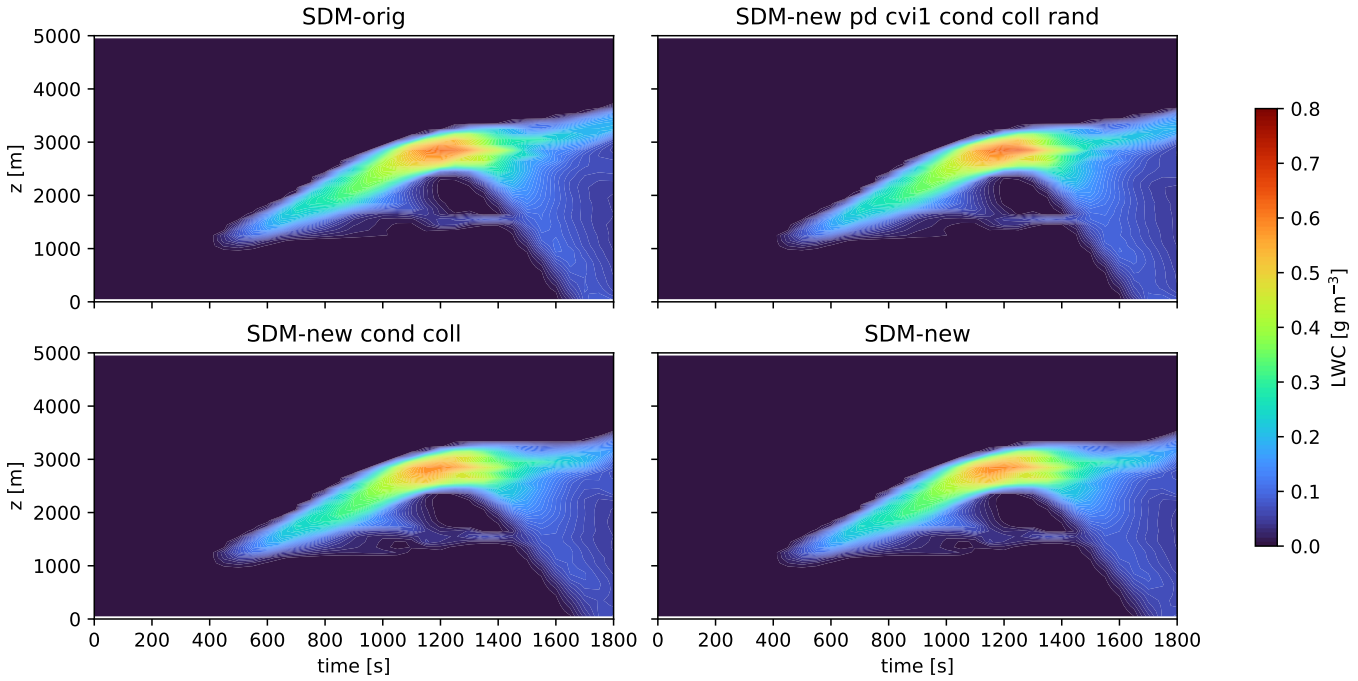

Figure 5. Same as Figure 4 but for SDM-orig32768 and SDM-new32768 using different settings

fully managed to gain an overall benefit from the larger timestep possible with the anelatic approximation. The SDM timestep seems comparable to Mellado et al. 2018 dynamics time step, while in this paper the dynamics timestep is 16 times lower. L891ff states another disadvantage.

> We agree with you that the description of anelastic equations in Section 2 is redundant and confusing. In the revised manuscript, we have specified the differences between fully-compressible and anelastic equations.

L132: Mention, that this is not a complete list of the warm phase microphysical processes. For example spontaneous and collisional break-up are not implemented, but are relevant for other meteorological situations with high rain rates. Actually, collisional break-up might be most complicated warm phase process, even if can be more naturally implemented in the SDM framework.

> In the revised manuscript, we have stated the process ignored and assumptions made for microphysics to be solved explicitly. We added the phrases in blue, as follows:
> L132: In this study, only the following warm cloud processes were considered: movement, activation/deactivation and condensation/evaporation, and collision–coalescence. Spontaneous and collisional breakup process was not considered in this study.
> L136: The ith SD moves according to the wind and falls with terminal velocity assuming that the velocity of each SD reaches the terminal velocity instantaneously.
> L146: The terms a/R and b/R^3 represent the curvature and solute effects, respectively. The ventilation effect is ignored in Eq. (2).

L150: Maybe I missed it, has the size of the collision grid ever been discussed in the context of SDM? If it is kept equal to size of the fluid cell, the interaction distance of SDs decreases with mesh resolution. I could even imagine that it could be a reason why the SD statistics in Sato et al. 2017 did not converge even at 5 m resolution. If you think that is a valid discussion point, it would fit in section 6 at L1077.

> We agree with your comment. Extending the collision grid to 2Δ–8Δ, which may be a resolved-scale length for mixing, may be reasonable for determining the interaction between fluid dynamics and microphysics. Analyzing low-pass filtered variables below filter length may be valid for investigating grid convergence. Though it is still debatable, we performed a similar analysis in our previous report (Matsushima et al. 2021). Thus, in this study, we used the same grid for the

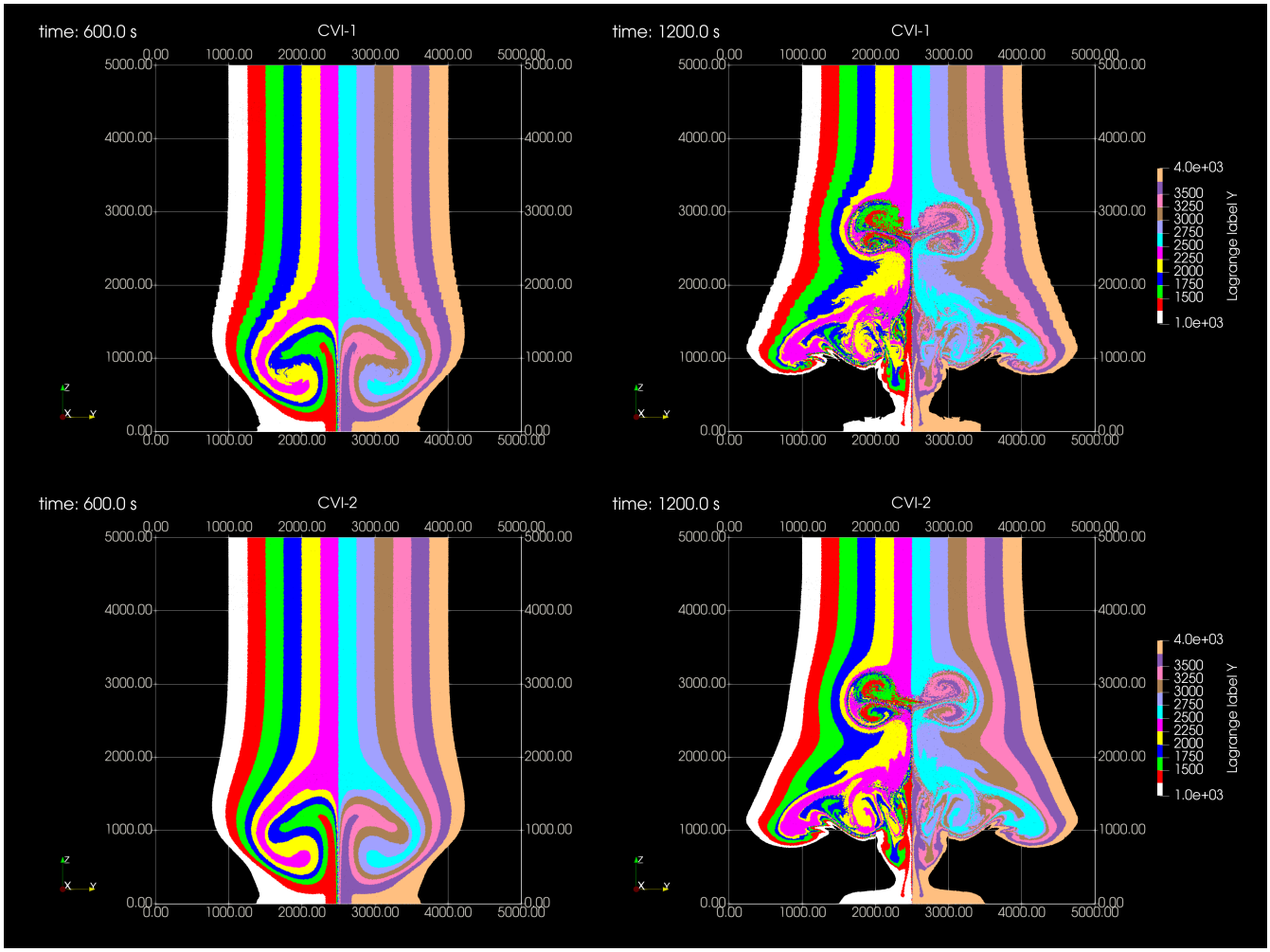

Figure 6. Distributions of SD positions at (left) $t = 600\,\mathrm{s}$ and (right) $t = 1{,}200\,\mathrm{s}$ colored by the initial y coordinate (Y) when CVI of the first-order (CVI-1) and second-order (CVI-2) spatial accuracy are used for SD movement. The range of $0 \leq y \leq 5{,}000$ and $1{,}000 \leq Y \leq 4{,}000$ are shown in each panel.

collision and model grids. In the revised manuscript, we have stated the collision grid explicitly. In the computational aspect, however, we are unsure if such a method is effective in optimization for SDM. Unfortunately, we cannot exclude sorting with cell index as a key, which is a bottleneck for collision–coalescence calculations. If we separate the collision grid from the model grid, we should subtract the noncloudy volume from the collision grid volume and avoid registering SDs within a noncloudy cell as a candidate for collision, which requires sorting with cell index anyway.

L155: A philosophical question: Do you view a super-particle as a sample from the real particles or as a statistical average over the multiplicity of real particles. I would prefer the former as the SDM could converge to the full problem for multiplicity -> 1 (even so that is not true for the O(N) collision algorithm). However, in my view turbulent (subgrid-scale) diffusion on the position space should then be included, otherwise initially very close SDs will not separate. This will prohibit the connection to DNS where two-point statistics are of interest. It seems even worse with low or fixed precision arithmetic (L441).

Thank you for the insightful question. We suppose that the philosophical difference becomes clearer if we consider how to take the limit to the real problem. We can assume two possible limits; $\xi \to 1$ with fixed Δx and $\xi \to 1$ with $\Delta x \to 0$. The first limit does not solve the real problem unless subgrid diffusion is not considered. In this case, it might not be possible to distinguish SD's position for low-precision representation. Note that the representation precision is $\sim 1\,\mu$m for 2 m grid. This is sufficient to distinguish SD's position because the mean distance between SDs ignoring inertia is $0.92\,\mathrm{cm} \gg 1\,\mu\mathrm{m}$ for CDNC of $100\,\mathrm{cm}^{-3}$. We can also use stochastic rounding to distinguish SD particles so that initially very close SDs can separate for the uncertainty of SD position. Alternatively, the second limit leads to the real problem because subgrid motion becomes relatively small. In addition, the representation precision for SD position increases with the limit to the

real problem.

Mellado et al. (2018) considered the first limit (using the original Navier–Stokes equation but changing only the kinematic viscosity among the atmospheric parameters). We implicitly assumed the second limit in the original manuscript, although we were motivated by the Mellado et al. (2018) approach. We have provided the following discussion in the revised manuscript: In the SDM, we did not consider the effect of turbulent fluctuations on movement, activation/deactivation and condensation/evaporation, and collision–coalescence owing to the high additional computational cost and amount of memory space required. However, we should include the effect of subgrid motion (or Brownian motion by kinematic viscosity) to ensure the convergence to DNS with $\xi \to 1$ while fixing the spatial grid length (Mellado et al. 2018), which should be addressed in future work.

3.1 Model framework of SCALE

L238: 3D/2D decomposition leads to a lot of data transfer via network (argument against the anelastic approximation in section 2.1). SDM physics contributes only to 50% runtime attributed to SDM. How much this MPI Wait times compared to other technical work as SD sorting?

The measured time for 3D/2D decomposition is listed in Table 1, which is provided at the bottom of this letter. The important point is that the number of MPI processes involved in communication in 3D/2D decomposition is small compared to that of all MPI processes. For example, in Section 5.2, the number of nodes involved in 3D/2D decomposition is only 16, which is much smaller than 36864. Therefore, we expect the cost to be considerably lower than that of the all-to-all communication among all nodes. We have stated this in the revised manuscript.

L245: Another argument for a 2D decomp for the SDM is that particle sedimentation in z direction does not lead to MPI communications.

We have revised L247 as follows: In addition, the amount of computation and data movement varies, depending on whether clouds and precipitation shaft are within the domain.

3.2 Initialization of super-droplets

L281: Could you make it clearer, why constant multiplicity is more natural? I can only understand that in the limit of multiplicity -> 1. It is important as it serves as motivation to develop the new init scheme.

When the number of SDs is smaller than the number of actual droplets but does not considerably deviate from it, the multiplicity of some droplets could become less than 1, depending on the PDF for sampling. In this case, it is reasonable to impose a constraint on the number so that the dynamic range of multiplicity is smaller (i.e., more similar to constant multiplicity) and multiplicity for all SDs is larger than 1.

In the revised manuscript, we have stated the following: When we sample a vast number of SDs, and the number of samples becomes close to the actual number of droplets, it is reasonable to impose a constraint on the number so that the dynamic range of multiplicity is smaller (i.e., more similar to constant multiplicity) and multiplicity for all SDs are larger than 1.

L285: This question is not completely answered in L825 ff. Add in section 6?

Please refer to L1103 (original manuscript): the sensitivity of the microphysical variability and precipitation to initialization parameter α should be further explored by high-resolution simulations.

L362-363: cache blocking and moving branches out of loops was already the way to go for vector computers, maybe citations would be appropriate.

We have added the reference Lam et al. (1991) for cache blocking.

3.3.2

L385 Important point, could you motivate the consistency of the densities a bit further. L155 states a turbulence scheme is used for tracers but not for the SDM. see comment to L155

We assume that SD number density is initially proportional to the total density. Then, the relation always holds if the eddy diffusivity of the total density and SD number density are identical, and the sedimentation of water substances is ignored. Even if the SD number density decreases with an updraft in the Lagrangian view, it is consistent with the environmental SD number density. Thus, we retained the number of SDs in convective clouds. In addition, we have reduced the non-physical variance of SD number density caused by the inconsistency of interpolated divergence at the SD position and grid divergence. In this study, we considered only the changes in SD number density due to grid-scale motion and precipitation and ignored the effect of eddy viscosity (or subgrid Lagrangian motion) because it is relatively small. Changes in the SD number density due to subgrid turbulence will be considered in future studies. In the revised manuscript, we have briefly stated the above in Introduction and Section 3.3.2.

3.3.5

L520 citations seem to be appropriated

We have added the reference Decyk and Singh (2014) (L520) for PIC sorting, although their method is slightly different from ours.

L527-534 This is not obvious to me, maybe a sketch would help?

Data hierarchy within each MPI process and an example of one-dimensional SD sorting are shown in Figure 7. In the revised manuscript, we have added Figure 7 and cited it in L527-534.

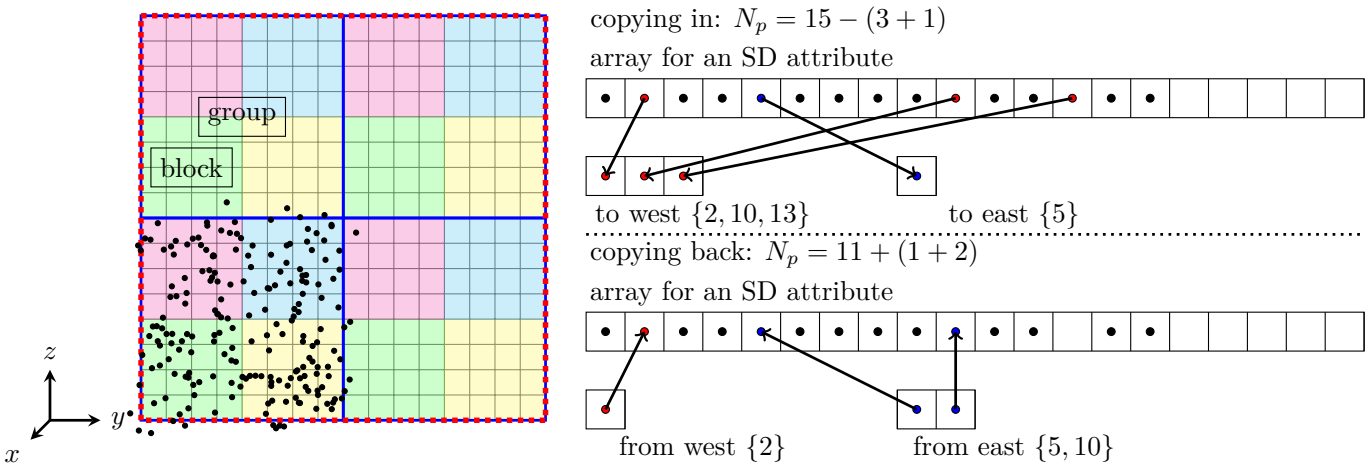

Figure 7. Data hierarchy (particle & cell, block, group) in each MPI process and algorithm for SD sorting toward x-direction in each block. In the example, an MPI process has 4 groups, a group has 4 blocks, and a block has 4×4 cells. Using a list ($\{\}$) that stores the indices, SD sorting completes by copying in the SDs moving to adjacent blocks and copying back the SDs moving into the block. The number of total SDs within a block is monitored by counting only the moving SDs.

4.1

eq. 10 this metric seems to measure the usage of the available computer power, but ignores convergence, i.e. how many tracers/bins/SDs are actually needed.

Thank you for highlighting this. We agree that it is important to include some convergence aspects in the discussion. However, it would be difficult to incorporate such into the metric. First, convergence properties can depend on the variable to be checked, for example, liquid water content, cloud droplet number concentration, and precipitation. Second, they can depend on the cloud form, amount of CCN, PDF used for initialization, and random number properties (pseudorandom number or low-discrepancy sequence). Thus, in this study, we did not mention the convergence aspect, except that 128/cell SDs are necessary for SCMS case, which was confirmed in our previous study (Matsushima et al. 2021) and other studies (Xue et al. 2022). In the revised manuscript, we have stated the limitations of the metric and cautions to be taken.

L582 A short discussion on how much optimization might be possible might be appropriate and link to the discussion in section 5 L953ff.

General optimization has been applied to Seiki and Nakajima (2014) scheme. In this scheme, the innermost loop for the vertical grid index, which performs complex calculations on each water substance, is vectorized by SIMD instructions. The innermost calculations are divided by separate loops by hand and compiler to improve computational performance by using cache. However, there may still be room for optimization to search for optimal loop fission and reordering calculations to hide operation latency. Optimization in terms of computational cost is applied to Suzuki et al. (2010) scheme. However, the innermost loops for bins are not vectorized for a small number of iterations. We have added this to the revised manuscript.

L583 This comparison seems indeed unfair, especially as you emphasize in several places that SD-new128 has similar performance (see general comment above)

We have compared the physical performance of SDM with that of the two-moment bulk scheme for only warm microphysics in the revised manuscript and deleted L583. Nonetheless, we would like to retain the results of the two-moment bulk scheme with mixed-phase microphysics for discussion in Section 6.

4.2

L625 Why has BIN a O(N**2) scaling? That implies that advection not dominant. If the reason is collisions, a linear sampling might also be possible. Hence, degree of optimization is hard to judge.

Thank you for highlighting this. For the bin method, the computational complexity is $O(N^2)$ for N bins when all possible combinations of collision are considered (${}_N\mathrm{C}_2$). A collision method similar to SDM may be applied. Sato et al. (2009) proposed reducing collision combinations by applying Monte Carlo integration inspired by SDM. As the method is implemented in the bin method (Suzuki et al. 2010), which we have used in our study, we performed additional warm bubble experiments with the option, and the results are shown in Figures 1 and 2 (BIN stoch.). We used the number of combinations of collision (M), $M = 16, 16, 16, 32, 128, 1024, 4096$ for the number of bins $N = 8, 16, 32, 64, 128, 256, 512$, respectively. The elapsed time of the bin method was reduced by using this option. To reduce the order of the computational complexity to $O(N)$, $M \sim N/2$ can be used. However, when we set the number of bins $N \geq 128$ and used $M \propto N$, the computations were terminated by large negative values of liquid water that cannot be compensated by filters. This is consistent with the previous study (Sato et al. 2009), which stated that $M \geq 0.056_N\mathrm{C}_2$ should be used to avoid large negative values of liquid water. Thus, computational complexity has not been successfully reduced. When this option can be stably used in the future, the elapsed time of the bin method becomes comparable to or faster than that of SDM-orig. In such situations, however, SDM would have advantages for considering multidimensional attributes to study collision–coalescence, as discussed by Shima et al. (2020). The above discussion has been provided in the revised manuscript.

4.3

L761 It do not understand the last sentence.

For clarity, we have revised the sentence as follows: When the number of available nodes is limited, simulations using the two-moment bulk method with more grids can be performed.

L783 what do you mean by "only SDs in one cell"?

We have revised the sentence as follows: To enable the intercomparison of models for the readers, each microphysical variable at a cell was calculated by taking statistics for SDs (only) within the cell.

L814 it is unfortunate that the bulk scheme seems to have a bug or inappropriate threshold. That weakens the conclusion in L823 and L1165 considerably.

Thank you for allowing us to reconsider this problem. The strange pattern is attributed to the fact that the bulk method cannot effectively solve evaporation by entrainment.

In Seiki and Nakajima (2014) scheme, the change of cloud water number density by evaporation was modeled such that it does not decrease unless the mean mass of cloud water (xc $\equiv 4/3\pi\rho_w R^3$, here ρ_w is water density, and R is droplet radius)

falls below the threshold xc_{ccn}. If it occurs, the cloud droplet number density completely evaporates (although smoothing is introduced). Mixing by entrainment always occurs homogeneously in this model. In the default setting, the threshold was very large (10^{-12}). This corresponds to a droplet radius of approximately 6 μm (CFADs are shown in the upper panel of Figure 8). If we reduce the threshold to a reasonably small value (4.2×10^{-15}, approximately 1 μm), the strange pattern becomes less noticeable (middle panel in Figure 8).

However, as we can assume that the real mixing can be homogeneous and inhomogeneous, the assumption of homogeneous mixing may be too simple. To incorporate the mixing scenario, we introduced the subgrid evaporation model (Morrison and Grabowski 2008, Jarecka et al. 2013), which considers a decrease in the cloud droplet number and water content, depending on the entrainment-mixing scenario. Jarecka et al. (2013) predicted a local mixing scenario calculated from the characteristic scale of cloud filaments and the fraction of a grid box occupied by cloudy air, which requires two new tracers. In this study, for simplicity and to avoid additional computational costs, we further assumed that the mixing scenario is fixed everywhere by a parameter α ($0 \leq \alpha \leq 1$, $\alpha = 0$ for homogeneous and $\alpha = 1$ for extreme inhomogeneous mixing) and evaporation is not delayed by subgrid-scale mixing. Based on the report of Jarecka et al. (2013), the parameters were set as $\alpha = 0.5$ for the pristine case (warm bubble, BOMEX) and $\alpha = 0.75$ for the polluted case (SCMS). The lower panel of Figure 8 shows the CFAD obtained using the Seiki and Nakajima (2014) scheme with subgrid-scale evaporation. CFAD slightly decreased from the original setup. The strange pattern at $R \sim 1\,\mu$m observed in the original setup completely disappeared and became consistent with the CFAD obtained using SDM. In our future studies, we will investigate CFAD when predicting local mixing scenarios with additional tracers and a more reasonable assumption for delayed subgrid-scale mixing before evaporation.

5.2

Table 3. tracking + condensation + coalescence only adds up to 141 min of 274 in microphysics. How much is due to the 3D/2D conversion + MPI wait steeming from the flow solver decomposition and how much from the technical workload needed for the microphyiscs itself like the SD sorting?

The measured times for 3D/2D decomposition and SD sorting are listed in Table 1. However, it should be noted that these times may not be the time needed to process 3D/2D decomposition and SD sorting as load imbalances from other processes will likely affect them. Also, as precise measurements are unimportant for practical cases, such as BOMEX and SCMS, we prefer to avoid measuring them.

Table 1. Elapsed time, FLOPS (peak ratio of the FLOPS [%]), Peta instructions per second, memory throughput (peak ratio of the memory throughput [%]), and particle throughput (# of floating point operations per SD)

	Time [min]	Speed [PFLOPS]	PIPS	Memory Throughput [PB/s]	Part. Throughput [particle·step/s]
Time integration loop	576	7.97 (7.04)	1.86	13.7 (37.2)	
Dynamics	290	8.55 (7.55)	2.03	20.5 (55.7)	
Microphysics	274	7.50 (6.62)	1.69	6.25 (16.9)	2.86×10^{13}
Short time step	238	9.50 (8.39)	2.19	21.3 (57.9)	
Tracer time step	15.0	5.85 (5.17)	1.78	21.6 (58.7)	
Tracking	87.9	15.3 (13.5)	2.14	2.89 (10.5)	8.91×10^{13} (171)
Condensation	32.6	18.2 (16.1)	5.35	5.28 (14.3)	2.40×10^{14} (75.9)
Coalescence	5.75	7.58 (6.69)	2.96	17.5 (47.3)	1.36×10^{15} (5.57)
SD sorting	79.2			12.5 (33.9)	
3D to/from 2D conversion	53.47				

L953ff bin method O(N**2), tracer advection is not even dominant (see comment to L625

The elapsed time depends on the computational algorithms and degree of optimization for the microphysics schemes. Here, we estimated the minimum elapsed time required by tracer advection and SD movement. In the revised manuscript, we have stated this assumption.

L972 0.8*15*33 /= 363

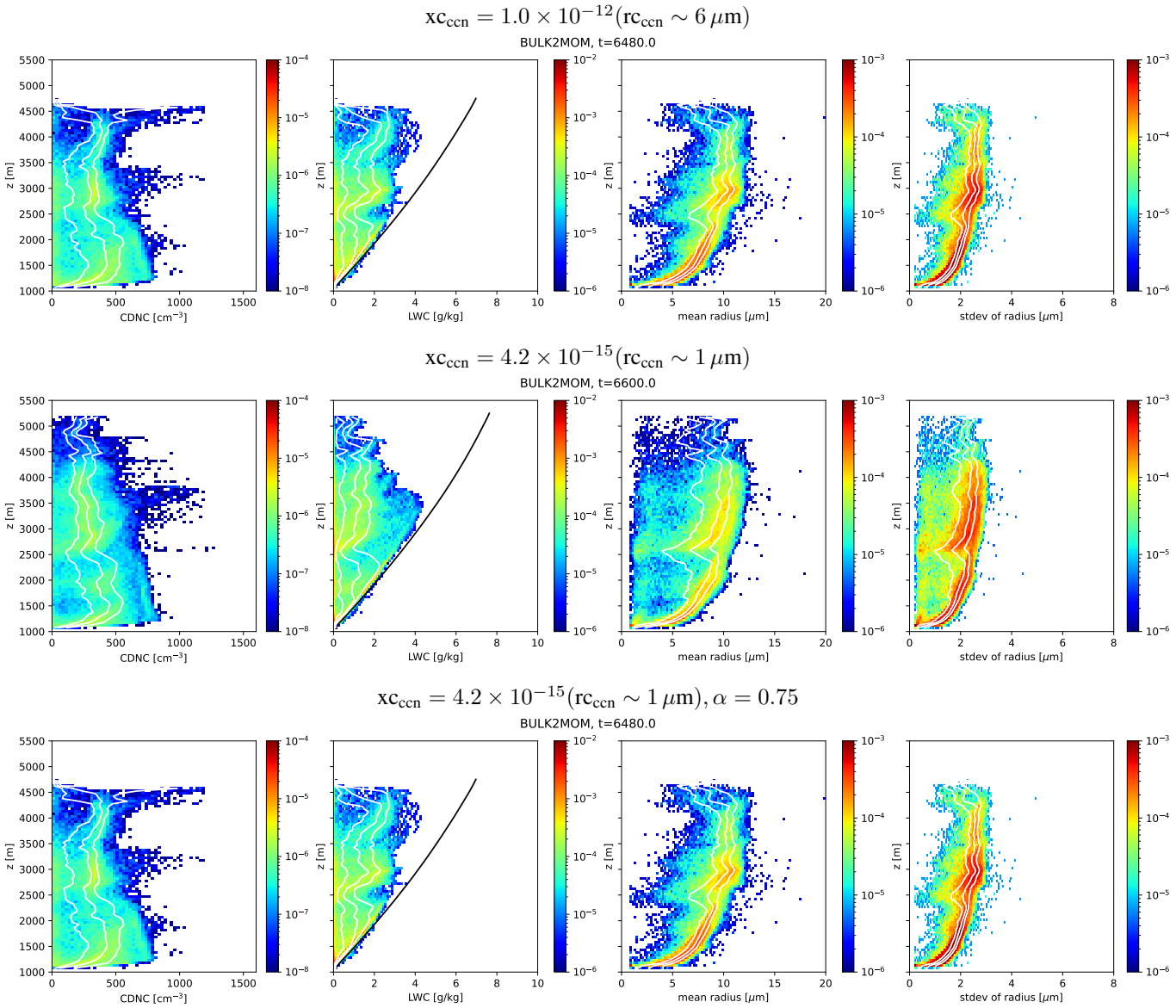

Figure 8. Contoured frequency by altitude diagrams (CFADs) of cloud droplet number concentration (CDNC), LWC, and mean and standard deviation of the radius for SCMS experiments. Snapshots of BULK2MOM with (top row) $xc_{ccn} = 1.0 \times 10^{-12}, \alpha = 0$ and (middle row) $xc_{ccn} = 4.2 \times 10^{-15}, \alpha = 0$ and (bottom row) $xc_{ccn} = 4.2 \times 10^{-15}, \alpha = 0.75$ are shown.

The estimation in L972 is based on when a memory throughput of 80% against theoretical peak performance is achieved. Thus, the elapsed time of tracer advection with 33 tracers was estimated to be 33*15*0.587/0.8 $\sim$ 363 min. We have modified L971 as follows: Then, a possible optimization could refactor the codes, and we may improve the memory throughput performance of tracer advection to achieve up to 80% of the theoretical peak performance.

6.1

L990ff are too optimistic. First, instantaneous freezing and melting assumptions are quite restrictive, e.g. leads to complete neglect of wet growth in deep convection. Mixed-phase regimes -38 to 0 are very important and frequent, e.g. riming of ice particle is a main driver for cold phase precip. Hence, I think the coexistence of water and ice particles will be needed often and extension to mixed-phase particle should be the next step after Shima 09 and Shima 20. In conclusion, I feel that the comparison to mixed-phase bulk is not justified.#f

Thank you for your comment. We have deleted the paragraph starting from L990. In L979, we refer only to the case for temperatures above 0 °C everywhere. In this paragraph, we retain the discussion on special cases in which the computational cost of mixed-phase SDM should be comparable to that of warm SDM if well implemented, whereas the computational cost of the mixed-phase two-moment bulk method increases for additional tracers. In L997 of the revised manuscript, we

have provided an example of riming for a mixed-phase regime, based on your suggestion, and described the challenges of optimization.

6.3

L1014ff Online postprocessing seems needed, for statistical analyses DSD on the scale of blocks should be sufficient. Of course the checkpoint/restart problem remains if the simulation time is longer than allowed by the compute queue. However, for such large scale applications it might be possible to discuss such rules with the computing center.

Thank you for your suggestion. Currently, statistical information about aerosols and clouds (number and water content) and activated SDs are outputted in a low-precision format. Analyses that need all information about SDs are possible only online. As you observed, we can make checkpoint/restart files in Fugaku, whose size is less than 400 TB, if we are permitted by the center. In the revised manuscript, we have mentioned that an online analysis should be performed if all information on SDs is necessary, and we may consult with the center for storing big data files. We would like to retain the discussion on data compression but in a simpler form.

technical corrections

L12: Add something like "possible" or "probably" for GPUs, as the effectiveness is not demonstrated in this paper.

Thank you for your suggestion. We have revised the sentence accordingly in the revised manuscript.

Fig 1. end of caption, dotted lines also in (d)

We have revised the statement in the revised manuscript.

L355 reformulate the sentence to avoid misinterpretation. E.g. move "except for the collision-coalescence process" to a new sentence or put it in parenthesis.

Thank you for your suggestion. We have revised the sentence accordingly.

L482 link to L375?

We have added "As discussed in Section 3.3.1" in L482 to remind readers of L375.

L661 "resists" might not be the best word

We have revised the word to "increases more gradually than linearly."

Fig 7. larger symbols in the legend

We have revised the symbols accordingly (see Figure 3 in this letter).

References

Decyk, V. K. and Singh, T. V.: Particle-in-cell algorithms for emerging computer architectures, Comput. Phys. Commun, 185, 708–719, https://doi.org/10.1016/j.cpc.2013.10.013, 2014.

Jarecka, D., Grabowski, W. W., Morrison, H., and Pawlowska, H.: Homogeneity of the subgrid-scale turbulent mixing in large-eddy simulation of shallow convection, J. Atmos. Sci., 70, 2751–2767, https://doi.org/10.1175/JAS-D-13-042.1, 2013.

Lam, M. D., Rothberg, E. E., and Wolf, M. E.: The cache performance and optimizations of blocked algorithms, Oper. Syst. Rev. (ACM), 25, 63–74, https://doi.org/10.1145/106973.106981, 1991.

Morrison, H. and Grabowski, W. W.: Modeling supersaturation and subgrid-scale mixing with two-moment bulk warm microphysics, J. Atmos. Sci., 65, 792–812, https://doi.org/10.1175/2007JAS2374.1, 2008.

Santos, S. P., Caldwell, P. M., and Bretherton, C. S.: Numerically relevant timescales in the MG2 microphysics model, J. Adv.

Model. Earth Syst., 12, e2019MS001 972, https://doi.org/10.1029/2019MS001972, 2020.

Sato, Y., Nakajima, T., Suzuki, K., and Iguchi, T.: Application of a Monte Carlo integration method to collision and coagulation growth processes of hydrometeors in a bin-type model, J. Geophys. Res., 114, https://doi.org/10.1029/2008JD011247, 2009.