Response to Review 1 of "Improved Formulation of Fragmentation of Snow during Collision with Graupel/Hail based on Observations at Jungfraujoch: Cold NonDendritic Regime of Temperature"

Reviewer:

This study uses a portable chamber with two GoPro cameras to study fragmentation of snow particles through graupel-ice collisions. The authors use these results to improve the choice of parameters for a theoretical physics-based model of graupel-snow collisional breakup that could theoretically be used in microphysics models. The authors use an array of solid ice spheres to represent graupel particles whereas naturally falling atmospheric snowflakes collide and break up on collision with these spheres. The authors then manually count the numbers of fragments, their sizes, and the estimated collisional kinetic energy for each event using ImageJ software. These new parameter results with the Phillips et al. (2017) model are then compared to the original results from Phillips et al. (2017) with the new results yielding a lower root-mean-square error to the new observations compared to the original Phillips et al. (2017) parameters.

I appreciate the authors' dedication to measurements of collisional SIP. There aren't many of these experiments and it seems like the group at Lund University is at the forefront of performing such measurements. However, I found that this particular manuscript has left me with far more questions than answers regarding the optimization of the Phillips model using measurements derived from these experimental setups. There are many sources of error in estimating fragment sizes and numbers and I don't think that the authors did a good job of proving that this new formulation is indeed an improvement over previous ones. There are only approximately 100 total collisions that were evaluated in this experiment which leads to only about 10 collisions per snowflake parent size bin. The authors find that the root mean square error with the new parameterization is +/- 3 compared to +/-5 from the old set of parameters (keeping in mind that this new parameterization was optimized *using* this dataset for evaluation rather than the old set of parameter values from Phillips et al. 2017). Given that the authors estimate quite large errors (nearly 50% CKE error from observations alone), it seems quite possible that the observational errors themselves would provide uncertainties in the fragment characteristics that are larger than the differences that the authors get between the old and modified versions. There are also somewhat intangible uncertainties as well that the authors do not describe as much as they should. For example, it is not clear to me based on the provided images how obvious and easily verifiable the manual classification of fragments are using ImageJ with the GoPro videos. In fact, the only actual images that the authors have of their snowflake collisions (Figure 2) are blurry! Even missing one or two fragments from collisions at each parent size could produce profound differences in the overall statistics that the authors use in their analysis. Overall, the authors need to do a better job describing and characterizing these uncertainties. As a result, I am recommending *major revisions* before considering publication of any kind.

Response:

We are grateful for the reviewer's scrutiny of the paper. The valuable criticisms will enhance the paper.

We will provide the videos and ImageJ metadata on a website for independent checking of our analysis when the paper is published. We have added extra discussion of the errors of the counting of fragments.

Yes, the differences among the three schemes seemed tiny in Figure 5. We now realise there was an error in that plot and it is now corrected, showing more sensitivity to the choice of scheme.

In fact, there are significant differences between the three schemes as shown in Figure 6 with the modified scheme being about 20% higher in the prediction for N, as noted below. This is a systematic bias (20%), while the error in predicting N (20%) is a random error.

New text to explain this has been added (line 516-517).

Reviewer: Major Concerns:

- There is an experimental setup flaw in the authors' design that they don't mention. Even though CKE is the controlling variable for determining fragmentation characteristics, the distribution of CKE measured in this study for *parent* snowflake particles is not necessarily consistent with that of graupel-snow collisions in nature. This is because the authors hold the "graupel" particles fixed in space and allow for the snowflakes to collide with these synthetic representations. However, in nature we would generally expect the *CKE to depend on the fallspeeds and masses of both graupel and snowflake*. Now it might be the case that the experimental snowflake *happens* to produce CKE values consistent when atmospheric graupel collides with ice particles, but to me, this isn't necessarily obviously true.

Response:

We disagree that there is any problem here.

The argument of the review is that the CKE values from our experiment are artificially low, which is claimed to make re-fitting of the formulation unrealistic. There are two reasons why this argument is not true.

First, let us first consider the effect of fixing the ice sphere on the CKE in terms of masses.

The mass of a snowflake of size 3 mm is $m_s = 4 \times 10^{-7}$ kg. The mass of a natural graupel particle (bulk density of 200 kg/m3 from Gautam et al. 2024) of 5 mm diameter is $m_g = 1.3 \times 10^{-5}$kg. Thus, the graupel is so much more massive than the snow particle, that its mass has little effect on the CKE. Replacing its mass by that of the Earth (to account for the effect from fixing it) has little effect on the CKE. Since m_s/m_g = 0.01, there is only a 1% change in CKE. So there is little effect on CKE via the mass from the act of fixing the target.

Now let us consider the effect on CKE from the fixing in terms of relative velocities. A graupel particle of 5 mm falls at about 2.5 m/s in nature (Rogers and Yau 1991). The snow particles fall at about 0.7 m/s in our experiment. Thus, the act of replacing the graupel particle in free-fall by a fixed ice sphere of the same size makes the CKE lower by a factor of $((2.5 – 0.7)/0.7)^2$ = 6 (the ratio of the squares of the fall-speed differences).

So, only the fallspeed reduction is significant for the CKE, comparing the experiment with the fixed spheres against reality for free-fall collisions. Such a reduction by a factor of 6 is small compared to the range of CKEs of the experimental data, which span 2 orders of magnitude (see Figure 5: **Fig. A**). So there is no problem with the representativeness of the range of the sample of CKE values observed in our experiment, relative to those expected for natural graupel-snow collisions in free-fall.

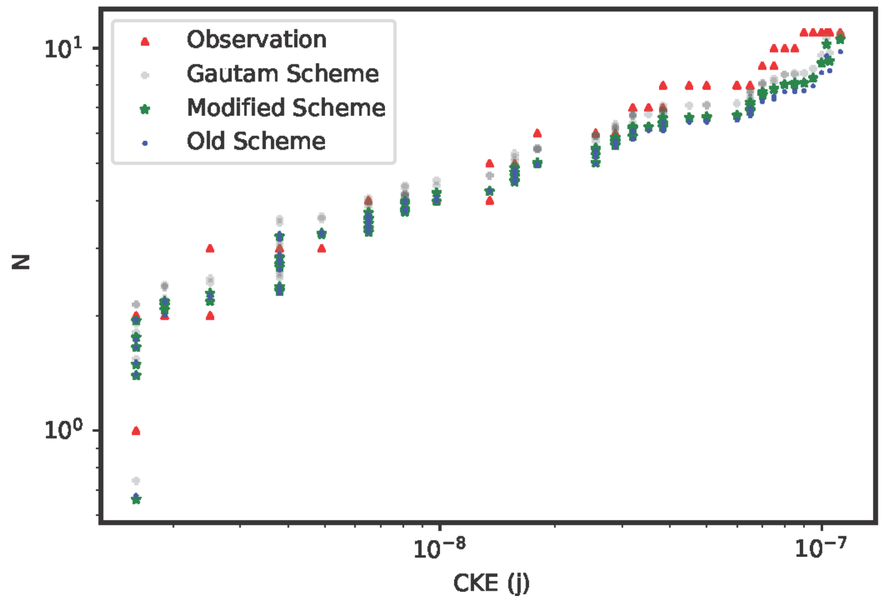

Figure A: *Comparison of number of ice fragments per collision observed (red) and modelled with the present scheme (green). From Fig 5 of the present paper.*

Second, as argued by Phillips et al. (2021, commentary in ACP on the review paper by Korolev and Leisner), the fixing of the ice spheres is immaterial for the formulation of breakup in ice-ice collisions from Phillips *et al*. (2017), since the formulation is free-standing as a theory based on classical mechanics with fundamental principles such as conservation of energy and the law of restitution. The formulation is not derived by simple curve-fitting to empirical data. Instead, fundamental constants of the theoretical formulation are defined by our lab data from experiments and these are universal to all conceivable collisions of graupel and snow, both natural and artificial.

For instance, in the experimental data, there is a correlation between CKE, incident snow size and numbers of fragments per collision. Yes, in the experimental data, any given size may be related to a lower CKE and fewer fragments than for natural snow-graupel collisions. No, that is not a problem for prediction of number of fragments in an atmospheric model because the theory applies to any permutation of sizes and CKEs, whether natural or artificial, in a universal realistic way.

If what we were doing were somehow mere curve-fitting, with numbers of fragments related to size in an empirical way, then the CKE being too low would be a problem. But that is not what we have done, since we include the true forms of the CKE dependency and size dependency in the formulation being fitted.

The commentary by Phillips et al. (2021) provides more details about this argument.

Extra text to clarify has been added at lines 165-176.

Reviewer: The authors should justify the Vardiman approach by illustrating that the CKE values that they get are *consistent* with CKE values that might be expected for *conditional CKE distributions of representative snow sizes* for typical snow and graupel collisions. That is, for a particular snow size, what would be reasonable CKE values that those snowflakes would have from different graupel collisions. Some CKE values actually show the *original* parameters performing a better job capturing the observed fragment values (Figure 5) so it's necessary to know what CKE values are more prevalent and typical than others for various snow sizes.

Response: Again, we are not doing mere curve-fitting to observational data of CKE, size and fragment numbers.

What Phillips et al. (2017) have created is a universal theory that includes explicit dependencies on CKE (with a Boltzmann type factor) and on size (contact area) derived fundamentally from classical mechanics. Yes, the CKE for a given snow collision in our field/lab experiment with the probe is lower than the corresponding natural snow-graupel collision for free-fall. No, that is not a problem because our theoretical formulation predicts the dependency on CKE in a fundamental way and applies to collisions both with artificially fixed ice spheres and with those in free-fall.

Reviewer: - It seems like the authors in this manuscript and in Gautam et al. (2024) estimate CKE errors around 30-50% or so. Gautam et al. 2024 estimate through a propagation of uncertainty analysis that this leads to approximately +/- 32% uncertainty in the predicted number of fragments. If the authors were to perform this same analysis, I suppose that this uncertainty would be even higher given that the authors estimate a 46% CKE error in the current manuscript. It seems to me that this error, if included in the analysis, would lead to a rather similar root-mean-square error with the original Phillips parameterization.

Response:

We *did* perform that analysis of propagating the errors in the reviewed paper. Indeed, that was how we found the error in the predicted number of fragments, which now we reckon is $\pm 20\%$ as stated at the end of the paper, which is indeed similar to the CKE error, as expected from the Gautam *et al*. paper.

The sample size we now realise reduces the random error of measurements among collisions. So we now find the error is much less. We did the maths of the error analysis more carefully this time.

Fresh text has been added to clarify (lines 359-365). Extra error analysis is in the caption of Table 1.

Reviewer: Given all these uncertainties, it's not clear to me why this new formulation really is worth incorporating into any scheme compared to the old version. To me, there needs to be a better empirical and statistical justification for stating that this current parameterization scheme is actually better. Can the authors also please state why these results are better or better representations than the "non-dendritic" measurements and results of Gautam et al. (2024).

Response:

The prediction of N is 20% higher in the present scheme than the old scheme from our 2017 paper. This is a systematic bias, apparent in Figure 6, as noted above. The error in the prediction (20%) is a random error, and over many uses in an atmospheric model, the random errors would likely partially cancel out.

Our idea is to use the Gautam et al. formulations for temperatures warmer than -17 degC and the new formulation for colder temperatures than that.

We have included the Gautam *et al.* formulation on the Fig. 5. It shows that the present scheme for cold dendritic temperatures (red) is consistently lower than the Gautam scheme (black) in the lower range of CKEs and vice versa in the higher range.

In summary, the present scheme is an improvement in the quest to reduce the systematic bias for the fragmentation of snow in the cold non-dendritic temperature regime. Text has been added to clarify (lines 514-518 and 563-564).

Reviewer:- I'm overall suspicious of the ability for the authors to accurately manually record the number and sizes of fragmentations from the GoPro camera videos. This is very important because the very small number of fragments recorded at each size (as shown in Figure 3, for instance). The authors should provide, as a supplement, example collision videos so that readers can better understand what the fragmentation events look like and so they can judge for themselves whether they can trust the authors' results. Because there are only about 100 total events used in this study, it doesn't seem unreasonable to upload the video clips of each collision event as supplementary material. Even a series of *high quality* stills from the videos showing an example of a fragmentation event would be helpful.

Response:

That is a good idea to upload the videos. This will be done once the paper is accepted.

Reviewer:- Figure 2: These particular photos are incredibly blurry. It even seems like the ice spheres, rather than the snowflakes, are in focus. I honestly can't even really make out the morphology of the snowflakes let alone tell that the aggregates are not composed of any dendrites. Do the authors have any better photos to show that demonstrates the snowflake morphology? Surely the authors require better quality photos/videos to determine the fragment properties, right?

Response: This the cameras yield images with this quality. They were all we could afford financially for the project. We also were unable to see the crystal habit from the images. Instead, we inferred likely crystal habit from the cloud-top temperature. The images were good enough to detect fragments > 0.2 mm.

GoPro cameras are not ideal for focusing on close-up images and this is the reason for the fuzziness. Their limit of focus is 30 cm in terms of distance from the camera. Better quality images are not possible.

Reviewer:-

Figure 3: I think it is very important here to show the percentiles or (minimum/maximum) numbers of fragments at each size. At the low and high parent sizes, there are only, on average, 2 fragments. If the authors only have about 10 or so collisions at each binned size, it seems to me that it's possible that the authors' results are not necessarily going to be statistically representative. Also, can the authors please state whether they consider collisions with no fragments in their statistics (if there were any)?

Response:

That is a good point. We have now included the error-bars in Fig. 3 and these show only a limited error.

There were practically no collisions with zero fragments.

A new Figure 3 is now included.

Reviewer:- Table 1: I'm a bit concerned with some of these parameters. The mass-dimensional power-law exponent 'b' is on the very low end of what values would be expected for unrimed or lightly rimed snowflakes.

Response:

We disagree. Our data is fine and close to what is expected from the literature.

The particles we observed were aggregates ('snowflakes'). Pruppacher and Klett (1997, their Table 2.6a) give mass-size relations based on data of Locatelli and Hobbs (1974) for four types of 'unrimed' snow crystal aggregates ('snowflakes'). These published relations have values for the exponent 'b' that are mostly slightly *lower* (1.4) than our value (1.5) at Jungfraujoch, not higher.

Ice is diverse in its morphology with innumerable species of ice particle. When one compares with other data, one needs to be careful to compare like with like.

Text has been added at lines 337-338 to state the comparison.

Reviewer: Most microphysics models will use values closer to b=2; many use empirical values from in situ aircraft measurements (e.g., Brown and Francis 1995; Baker and Lawson 2006; etc.). The authors also state in the conclusions on line 365 that the snow was partly rimed which suggests to me that a larger 'b' might be appropriate.

Response:

That is a good point. Locatelli and Hobbs (1974, their Table 1) observed b = 1.9 for 'densely rimed' aggregates and b = 1.4 for 'unrimed' aggregates. But they only used a microscope to observe the degree of riming, so these are quite fuzzy classifications.

We would probably categorise our sampled snow as lightly rimed, not from any observations of fallspeed or appearance but rather from our simulation of snow growth in a similar cloud (ACAPEX) with ψ = 0.2. Locatelli and Hobbs (1974), and perhaps other observationalists (e.g. Baker and Lawson 2006), often could not discern any rime on their sampled snow and called it "unrimed" (Pruppacher and Klett 1997, their Table 2.6a). Yet, in fact, any simulation of snow growth for in-cloud conditions would predict non-zero rime. Riming is inevitable generally to some degree.

Generally for snow to form, mixed-phase cloudy conditions are favoured (for nucleation of ice and for the humidity to support vapour growth), and those same conditions of supercooled cloud-liquid make it impossible for there to be zero rime. If rime will accumulate in the interior of the snowflake, it will not be visible and may be missed.

In summary, we agree that riming will tend to raise the exponent of the mass-size relation. However, we think our degree of riming (lightly rimed with ψ = 0.2) is likely fairly similar to snow in the literature reported (Locatelli and Hobbs 1974) as 'unrimed'. Since our value of b agrees with that for 'unrimed' snow, there is no problem.

Reviewer: These empirical values of course are simply fits; that is, there is not any true value for these parameters. With that said, I do think the authors need to at least caution the readers that the parameterizations here are limited and that there could be large inconsistencies between what the authors use in their breakup model and what is often used in bulk and bin microphysics schemes and what might be expected in regions other than Jungfraujoch.

Response:

Again, the scheme does not involve mere curve-fitting to a combination of observational data of size, fall-speed and fragmentation. If it were, then the scheme would be restricted in applicability to the characteristics of the fixed target, snow morphology and mass-size relation of the case sampled.

Rather, the scheme is based on a theory that has some universal applicability with physically justified dependencies on size and CKE. Values of the fragility coefficient, C, inferred from observations of a specific snowstorm at Jungfraujoch on a specific day are therefore not restricted to that particular day and location. Of course, from location to location, the fragility of snow and its morphology will vary with the in-cloud conditions in ways that might not be fully captured by the theoretical formulation.

Yes, we have included a caveat to this effect with two new paragraphs in the new discussion section (lines 457-469).

Minor Concerns:

Reviewer:- Figure 1: It's bewildering to me why this diagram is of such low quality when it is so simple. Surely it would take only a few minutes to produce a vectorized version of this diagram with adobe illustrator or some similar program. The figure looks like it is from the 1970s or even earlier. This is what the figure looks like on my end:

Why is the figure made up of tiny dots rather than lines? Is there any way for the authors to upload a better quality version of this diagram? I checked out the version in Gautam et al. 2024 and that version seems to be of a higher quality than what's in the current manuscript.

Response:

We agree, something went wrong with our use of the diagram. We have replaced it with a better quality one.

There is a new version of Figure 1.

Reviewer: Suggestions/Questions/Typos:

- Line 355: The authors mention the Vardiman experiments here as well as in section 2. I think that there should be a bit more discussion in the introduction regarding this and other similar experiments (see for example Appendix 2 in Kenneth Lo's 1983 dissertation (https://dspace.mit.edu/bitstream/handle/1721.1/54294/11493743-MIT.pdf?sequence=2); for example). Regarding the distribution function in Lo's thesis, it would be interesting to see how well this analytical inverse-exponential fragment (mass) distribution function (devised over 40 years ago) compares to the authors' empirical results from Figure 4 and perhaps to the results of Gautam et al. (2024).

Response:

This is a great suggestion.

Lo's thesis is consistent with a lognormal distribution of fragment mass and size. This mathematically implies only a single peak. Generally, lognormal distributions describes the size distribution of particles from fracture with nonlinear transfer of energy through the material by multiplicative processes (e.g. one breakage of a monomer leads to a 'crack' in the snowflake leading to more breakages).

That single peak is resolved by our observations, which suggests our optics is good enough and that there is no hidden peak of fragments missed at unresolved sizes.

This is clarified in new discussion section (line 479-489). We now cite Lo's thesis.