Reply to referee #1: Interactive comment on "A High Speed Particle Phase Discriminator (PPD-HS) for the classification of airborne particles, as tested in a continuous flow diffusion chamber"

By Fabian Mahrt et al.

Reviewer comments are reproduced in **bold** and our responses in normal typeface; extracts from the originally submitted manuscript are presented in *red italic*, and from the revised manuscript in *blue italic*.

We have numbered the reviewer's major comments for ease of cross-reference within the other reviews.

General comments:
The authors present a new method to characterise the particle phase of hydrometeors using a novel light scattering instrument and machine learning algorithm. This represents a promising advance in studying mixed phase cloud microphysics since particle phase can be determined from scattering parameters independent of the particle size. The utility of the method is demonstrated using different conditions in a CFDC to generate liquid droplets, ice crystals, and mixtures.
The manuscript is very well written and scientifically sound. I recommend that it is suitable for publication in its present form, however I have a few minor suggestions for the authors to consider, listed below. I think there is great potential for more detailed investigations of cloud microphysics with this technique, and look forward to (hopefully) future measurements of mixed phase clouds in the atmosphere. I would also like to commend the authors on the excellent documentation, data, and explanations of their methodologies and measurements provided.

We thank the reviewer for carefully reading the manuscript and the overall constructive comments on it. We hope that the responses below satisfactorily address the reviewer concerns.

Specific comments:
P5 Fig1: I would suggest clarifying that the light orange and brown shading denote *detected* light from the trigger and image laser beam.
We have changed the last sentence in the caption of Fig. 1 to now read:
"Light orange and brown shading in panel (a) and (b) correspond to light scattered by particles when passing the trigger laser beam and image laser beam, respectively, and ultimately detected by the PD and the CMOS arrays."

P6 L9: Can you provide an estimate of the azimuthal angle range of detected light?
The azimuthal angle is 9°. We now add this information in the revised manuscript (P6L12):

P6 L6-9: What is the purpose of L4-L5?
We have added the following sentence on P6L9 (initial manuscript) P6L17 (revised manuscript) to explain the use of L4 and L5:
"The lenses L4 and L5 reduce the size of the image independently in the horizontal and vertical planes, respectively, yielding the elliptical output image ultimately captured by the linear CMOS array system."

P6 L11-12: Was saturation of the CMOS arrays ever problematic.
No, (intensity) saturation of the CMOS arrays has never been observed for the particle types and size range used in the experiments of the presented study.

From Fig. 5 in the revised version one can see that the maximum particle size that would trigger the laser is approximately 70 µm. At this size, the 12-bit detector reaches a maximum value, i.e. saturates (AD = 4096), as indicated by the instrument response (AD).
We have added the following statement on P13L9 (initial manuscript) P14L7 (revised manuscript):
“It can further be seen that the a maximum AD is reached for particles of approximately 70 µm yielding an upper size limit for particles to be detected and recorded by PPD-HS (detector saturation). However, it should be noted that the maximum particle size tested here was 32 µm and that an upper size limit of PPD-HS would need to be tested in future experiments.”

P9 L1-4: Since *ΔTBC* is later determined to be a key value, perhaps it is worth mentioning it here.
We agree with the suggestion of the reviewer to mention the ΔTBC already at this point, or at least hint at the possibility of deriving further parameters from the two TBC values, which is revealed by the PCA analysis later in the manuscript. We have therefore added the following statement (page X line X in revised manuscript):
“Moreover, the TBC values of both arrays can be used to derive further sphericity parameters, e.g. the ratio of both TBCs or their absolute difference, which can improve particle classification.”

P11 L3-4: Where/how were temperature and relative humidity measured?
The specified temperatures (*T*) and relative humidities (*RH*) denote the conditions within the horizontal ice nucleation chamber (HINC), which is used to produce either cloud droplets, ice crystals or a mixture of both on the injected aerosol particles. These are calculated values derived by controlling the temperature of ice-coated chamber walls. HINC is a continuous flow diffusion chamber and etailed descriptions of its working principle including the control of *T* and *RH* conditions can be found in the indicated references P10L18: Lacher et al. (2017), Mahrt et al. (2018). This method of exposing aerosol to defined *RH* and *T* conditions has been well established for a few decades (Hussain & Saunders, 1984; Rogers, 1988).

P14 Fig5: I believe the minimum value on the x-axis should be 1µm instead of 0.1 µm.
Thank you for spotting this. We agree, the limits of the x-axis are 0.1 and 100 µm and we have corrected the figure accordingly.

P16 L22-23: Can you comment on why the machine learning algorithm would classify particles as aspherical (e.g. Fig8 39,46,74) with lower TBC and AIC values than particles classified as spherical (e.g. Fig8 29, 57, 41)?
The reviewer raises a valid point here. When considering for instance the particles #29 and #39 of Fig. 8, both show a visually symmetric scattering pattern, and should hence be classified as spherical. The visual symmetry is described in terms of 4 absolute numbers, namely TBC1, TBC2, ΔTBC and AIC, which are used by the algorithm to determine particle shape. We agree that it appears counterintuitive that the particle with overall higher TBC and AIC values (#29) is classified as spherical, whereas the particle with the lower TBC and AIC values (#39) is classified as aspherical by the random forest model, as pointed out by the reviewer.

This apparent misclassification results from imperfect training data sets, as stated on P17L7-18 (initial manuscript), and can be understood upon consideration of Fig. S13 and S14 of our SI. From Fig. S13a and b it becomes clear that there exists aspherical particles with low TBC values, comparable with those of spherical particles (overlap of blue and red curves) within the (entire) calibration data set.

During the random sampling of (200’000) aspherical particles for the training of the random forest model (see P12L17-25, initial manuscript) aspherical particles with any TBC values described by the red curves in Fig. S13a and b can be selected. Hence, there is no a priori constrain to high TBC values for aspherical particles during the training of the random forest model, which is then trained to classify any such particle with the flag “isAspherical”. In other words, the misclassification results due the overlap of the TBC distributions of spherical and aspherical particles shown in Fig. S13a and b. This overlap in turn results mainly from the VOAG produced NaCl particles that show symmetrical scattering patterns thus associated with low TBC values (Fig. S16) and to a lesser extent also, from VOAG produced PEG particles which could exhibit non-symmetrical scattering patterns and hence relatively larger TBC values (Figs. S15) despite being spherical. Here, we have simply assigned any PEG and NaCl particle produced by the VOAG, as spherical and aspherical, respectively. An improved particle classification, also for the particles questioned by the reviewer, could be achieved through visual inspection of all particles/scattering patterns within the calibration data set and manual classification as either noise, spherical or aspherical, and then only train the random forest model on this cleaned data set (see P12L17-25, initial manuscript). However, this is not feasible for the large number of particles within our data sets, as explained in the text (P17L15-18).

We note that upon inspection of Fig. S13a and b that the majority of the aspherical NaCl particles clearly show higher TBC values compared to the spherical PEG particles. This results in the overall good and correct classification of particles by the random forest algorithm (see Fig. 6), by using a sufficiently large number of particles (in our case 200’000) for training, where statistically, the majority of aspherical particles encompasses TBC values different from those of the spherical particles.

We have now addressed this limitation more explicitly through various changes throughout the main text as indicated below:

Added the statement on P13L5 (revised manuscript):
“Any PEG or NaCl particle produced as described in Sect. 3.1.1 and fulfilling these usability criteria, is defined as spherical and aspherical particle, respectively, within the calibration data set, without further visual inspection of the scattering pattern. However, it should be noted, that this approach does not filter out aspherical NaCl particles associated with rather symmetric scattering patterns and consequently low values for the symmetry parameters (see Fig. S16). Hence, this can explain the overlap in the distribution of for instance of the TBC values of both particles classes (see Fig. S13a and b) and thus potential misclassification of asymetrical NaCl as spherical particles.”

Added the statement on P13L19 (revised manuscript):
“Selection of a sufficiently large number of aspherical particles for training the random forest model ensures that a statistical particle majority will show TBC values different from those observed for spherical particles (see Fig. S13a and b).”

Added the statement on P15L15 (revised manuscript):
“In addition, some of the misclassification can result from near-spherical NaCl particles within the training data set, as discussed above.”

Changed from P17L8 (initial manuscript):
“Nevertheless, we note that there are particles classified as isAspherical, even though the scattering patterns appears symmetrical.”
To P18L12 (revised manuscript):
“Nevertheless, we note that there are particles classified as isAspherical, even though the scattering patterns appears symmetrical, for instance, particles 39 and 46 in Fig. 8, which have values for the symmetry parameters comparable to particles classified as spherical (e.g. particle 29 in Fig. 8).”

Added to P17L15 (initial manuscript), P18L22 (revised manuscript):
“The consequence is that some particles become misclassified as e.g. aspherical, despite their overall symmetric scattering patterns (see above). This error could be reduced through manual visual inspection and manual selection and definition of particle class for every particle within the calibration data set, prior to training of the random forest model.”

Added to P17L18 (initial manuscript), P18L27 (revised manuscript):
"The latter results from the majority of the spherical and aspherical particles within the calibration data set to distinctively differ in terms of their symmetry parameters (see Fig. S13)."

Added to P25L27 (revised manuscript):
"Finally, we have noted above that our random forest model is associated with a misclassification rate, resulting in some symmetrical scattering patterns to be classified as aspherical and vice versa (see Sect. 4.3 and Fig. 8). We have argued that this is a consequence of artifacts within the calibration data set (see SI Sect. S6.1) from which particles are randomly selected for the training of the classification algorithm. This error could be reduced and overall classification could be improved in future studies, upon manual cleaning of the calibration data set prior to model training."

We have further changed SI Sect. S6.1 from (SI P15L12, initial SI):
"*Thus, even though the TBC in general is a good measure for particle sphericity, an absolute threshold value above which all particles are considered aspherical cannot be applied. This can partly explain the misclassification of clearly aspherical particles as spheres. Overall, this is a shortcoming of using the particle measures within the random forest model, rather than the individual pixel information."*
To SI P15L12 (revised SI):
"Similarly, some of the NaCl particles produced by the VOAG reveal symmetrical scattering patterns, which are consequently associated with relatively low TBC values (see Fig. S16). While most of these near-spherical NaCl particles show a symmetric scattering pattern along only one of the CMOS arrays, we cannot exclude NaCl particles from our calibration data set that show symmetry comparable to spherical PEG particles, without manual inspection of these particles. Thus, even though the TBC in general is a good measure for particle sphericity, an absolute threshold value above which all particles are considered aspherical cannot be applied. Overall, this is a shortcoming of using the particle measures within the random forest model, rather than the individual pixel information, as well as defining all VOAG produced PEG and NaCl particles as spherical and aspherical, respectively, at the absence of a manual check of the individual scattering patterns."

P17 L34: Can you offer an explanation for the trimodal appearance of the AIC values for the 243K NH4NO3 case (Fig 10d)?
We do not have a direct explanation for the trimodality of the AIC values, however, given that all the values are really low, i.e. below the threshold of 0.2 (as can be inferred from our Fig. S13d), we are confident these are all cloud droplets.

SI P4 FigS4: It would be helpful to have *RTe* and *RAWe* defined in the caption (presumably real-time electronics board and raw sampling electronics?).
We have now defined both terms in the figure caption.

SI P6 FigS6: "*2 µm"* instead of "*2 mum"*
We have corrected this in the figure caption now.

SI TableS1: A minor suggestion, it would be helpful to have the data sets used in FigS12 highlighted by bold or coloured font.

We have now highlighted the datasets used in Fig. S12 in bold and noted also that in the caption of Tab. S1.

Hussain, K., & Saunders, C. P. R. (1984). Ice nucleus measurement with a continuous flow chamber. *110*(463), 75-84. doi:10.1002/qj.49711046307

Lacher, L., Lohmann, U., Boose, Y., Zipori, A., Herrmann, E., Bukowiecki, N., . . . Kanji, Z. A. (2017). The Horizontal Ice Nucleation Chamber (HINC): INP measurements at conditions relevant for mixed-phase clouds at the High Altitude Research Station Jungfraujoch. *Atmospheric Chemistry and Physics, 17*(24), 15199-15224. doi:10.5194/acp-17-15199-2017

Mahrt, F., Marcolli, C., David, R. O., Grönquist, P., Barthazy Meier, E. J., Lohmann, U., & Kanji, Z. A. (2018). Ice nucleation abilities of soot particles determined with the Horizontal Ice Nucleation Chamber. *Atmos. Chem. Phys., 18*(18), 13363-13392. doi:10.5194/acp-18-13363-2018

Rogers, D. C. (1988). Development of a continuous flow thermal gradient diffusion chamber for ice nucleation studies. *Atmospheric Research, 22*(2), 149-181. doi:http://dx.doi.org/10.1016/0169-8095(88)90005-1